SIAM J. COMPUT.
Vol. 0, No. 0, pp. FOCS19-1–FOCS19-38
© 2021 Society for Industrial and Applied Mathematics

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

OPTIMIZATION OF THE SHERRINGTON–KIRKPATRICK HAMILTONIAN*

ANDREA MONTANARI†

Abstract. Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a symmetric random matrix with independent and identically distributed (i.i.d.) Gaussian entries above the diagonal. We consider the problem of maximizing $\langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle$ over binary vectors $\boldsymbol{\sigma} \in \{+1,-1\}^n$. 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 $\boldsymbol{\sigma}_* \in \{-1,+1\}^n$ such that $\langle \boldsymbol{\sigma}_*, \boldsymbol{A}\boldsymbol{\sigma}_*\rangle$ is at least $(1-\varepsilon)$ of the optimum value, with probability converging to one as $n \to \infty$. The algorithm's time complexity is $C(\varepsilon)\, n^2$. We generalize it to matrices with i.i.d., but not necessarily Gaussian, entries, and obtain an algorithm that computes the MAXCUT of a dense Erdős–Renyi random graph to within a factor $(1 - \varepsilon \cdot n^{-1/2})$. As a side result, we prove that, at (low) nonzero temperature, the algorithm constructs approximate solutions of the Thouless–Anderson–Palmer equations.

Key words. Sherrington–Kirkpatrick, spin glasses, replica symmetry breaking, message passing algorithms

AMS subject classifications. 68Q87, 82B44, 60K35

DOI. 10.1137/20M132016X

1. 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):

$$\begin{aligned} &\text{maximize } \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle \\ &\text{subject to } \boldsymbol{\sigma} \in \{+1,-1\}^n . \end{aligned} \tag{1.1}$$

From a worst-case perspective, this problem is NP-hard and indeed hard to approximate within a sublogarithmic factor [ABE+05]. 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 [SK75]. 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 [MPV87, Tal10, Pan13b]. Generalizations of this model have been used to understand structural glasses, random combinatorial problems, neural networks, and a number of other systems [EVdB01, MPZ02, WL12, Nis01, MM09].

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 $H_n(\boldsymbol{\sigma}_*) \ge (1-\varepsilon)\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n}$

*Received by the editors February 19, 2020; accepted for publication (in revised form) August 24, 2020; published electronically January 7, 2021. A conference version of this paper was presented at the 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS).
https://doi.org/10.1137/20M132016X
Funding: This work was partially supported by NSF grants DMS-1613091, CCF-1714305, and IIS-1741162, and by ONR grant N00014-18-1-2729.
†Electrical Engineering and Statistics, Stanford University, Stanford, CA 94305 USA (montanari@stanford.edu).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$H_n(\boldsymbol{\sigma})$. 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 (1.1). 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 $(t,x) \in [0,1] \times \mathbb{R}$:

$$\begin{aligned} &\partial_t \Phi(t,x) + \frac{1}{2}\beta^2 \partial_x^2 \Phi(t,x) + \frac{1}{2}\beta^2 \mu(t) \big(\partial_x \Phi(t,x)\big)^2 = 0\,, \\ &\Phi(1,x) = \log 2\cosh x\,. \end{aligned} \tag{1.2}$$

It is understood that this is to be solved backward in time with the given final condition at $t = 1$. Existence and uniqueness were proved in [JT16]. We will also write Φ_μ to emphasize the dependence of the solution on the measure μ. The Parisi functional is then defined as

$$\mathsf{P}_\beta(\mu) \equiv \Phi_\mu(0,0) - \frac{1}{2}\beta^2 \int_0^1 t\,\mu(t)\,\mathrm{d}t\,. \tag{1.3}$$

The relation between this functional and the original optimization problem is given by a remarkable variational principle, first proposed by Parisi [Par79] and established rigorously more than twenty-five years later by Talagrand [Tal06b] and Panchenko [Pan13a].

THEOREM 1 (Talagrand [Tal06b]). *Consider the partition function* $Z_n(\beta) = \sum_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \exp\{\beta H_n(\boldsymbol{\sigma})\}$. *Then we have, almost surely (and in* L^1*),*

$$\lim_{n\to\infty} \frac{1}{n} \log Z_n(\beta) = \min_{\mu\in\mathscr{P}([0,1])} \mathsf{P}_\beta(\mu)\,. \tag{1.4}$$

The following consequence for the optimization problem (1.1) is elementary; see, e.g., [DMS17].

COROLLARY 1.1. *We have, almost surely,*

$$\lim_{n\to\infty} \frac{1}{2n} \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle = \lim_{\beta\to\infty} \frac{1}{\beta} \min_{\mu\in\mathscr{P}([0,1])} \mathsf{P}_\beta(\mu)\,. \tag{1.5}$$

Remark 1.1. The limit $\beta \to \infty$ on the right-hand side of (1.5) can be removed by defining a new variational principle directly "at $\beta = \infty$." Namely, the right-hand side of (1.5) can be replaced by $\min_\gamma \hat{\mathsf{P}}(\gamma)$ where $\hat{\mathsf{P}}$ is a modification of P and the minimum is taken over a suitable functional space [AC17]. In this paper we use the $\beta < \infty$ formulation. Follow up work [AMS20] showed that it is also possible to work directly at $\beta = \infty$, at the expense of some additional technical work.

We also note that while we stated Theorem 1 and Corollary 1.1 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 $\mathbb{E}\{|A_{ij}|^3\} \le C/n^{3/2}$ [CH06]. (Indeed, even weaker conditions are sufficient [DMS17].)

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Existence and uniqueness of the minimizer of $\mathsf{P}_\beta(\,\cdot\,)$ were proved in [AC15] and [JT16], 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 [MPV87]. In particular, this conjecture is supported by high precision numerical solutions of the variational problem for P_β [CR02, OSS07, SO08]. Rigorous evidence was recently obtained in [ACZ17]. 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 $\beta > 1$ [Ton02]).

We are now in position to state our main result. The reader interested in a compact description of the algorithm used in this proof may consult Appendix B.

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 other words, entries are subgaussian with common subgaussian parameter C_*/n).

In other words, on average, the optimization problem (1.1) is much easier than in the 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, which is encoded in Assumption 1.

Remark 1.2 (role of the no-overlap-gap assumption). The proof of Theorem 2 proceeds in two steps. We first develop a characterization of the value achieved by a general class of algorithms, which are parametrized by two functions $g, v : \mathbb{R}\times\mathbb{R}_{\ge 0} \to \mathbb{R}$. This analysis does not really depend on the specific structure of problem (1.1), apart from making use of the data matrix $\boldsymbol{A} \sim \mathsf{GOE}(n)$. In particular, it does not use the structure of the constraint $\boldsymbol{\sigma} \in \{+1,-1\}^n$. The results of this analysis are summarized in Theorem 4.

We then specialize this analysis to problem (1.1), by choosing functions g, v as

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

to match the prediction of the Parisi formula. This step is, of course, dependent on the structure of the constraint set, and makes use of Assumption 1 in a very specific passage. Namely, we use the assumption in checking that our proposal for g, v satisfies condition (a) of Lemma 2.8. This is done at the beginning of the proof of Theorem 4.

Remark 1.3 (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 $\partial_x\Phi(t,x)$, $\partial_x^2\Phi(t,x)$ at specified points (t,x), with each query costing $O(1)$ operations.

This is a reasonable model for two reasons: (i) The PDE (1.2) 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 [Gue03] and strongly convex [AC15, JT16]. Further, the PDE solution Φ is continuous in μ and can be characterized as a fixed point of a certain contraction [JT16]. Because of these reasons we expect that an oracle to compute $\Phi(t,x)$, $\partial_x\Phi(t,x)$, $\partial_x^2\Phi(t,x)$ 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 [MPV87, Tal10].

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 [BLM15]. 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 5 contains such a generalization.

As a special case of random matrices $\boldsymbol{A}$ with independent subgaussian entries, we can consider (centered) adjacency matrices of dense Erdős–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 Erdős–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 a 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 [DMS17] that[1]

$$\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \mathsf{CUT}_G(\boldsymbol{\sigma}) = |E_n|/2 + (n^3p(1-p)/2)^{1/2}\mathsf{P}_* + o(n^{3/2})\,,$$

$$\mathsf{P}_* := \lim_{\beta\to\infty} \frac{1}{\beta} \min_{\mu\in\mathscr{P}([0,1])} \mathsf{P}_\beta(\mu)\,.$$

[1] In [DMS17], the same result is shown to hold for sparser graphs, as long as the average degree diverges: $np_n \to \infty$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

In other words, MAXCUT on dense Erdős–Renyi random graphs is nontrivial 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 1.2. *Under Assumption* 1*, for any* $\varepsilon > 0$ *there exists an algorithm (with complexity at most* $C(\varepsilon)\, n^2$*), that takes as input an Erdős–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 a balanced cut* $\boldsymbol{\sigma}_* = \boldsymbol{\sigma}_*(G) \in \{+1,-1\}^n$ *(with* $\langle \boldsymbol{\sigma}_*, \mathbf{1}\rangle | \le 1$*), such that*

$$\left(\mathsf{CUT}_G(\boldsymbol{\sigma}_*) - \frac{|E_n|}{2}\right) \ge (1-\varepsilon) \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \left(\mathsf{CUT}_G(\boldsymbol{\sigma}_*) - \frac{|E_n|}{2}\right). \tag{1.6}$$

The rest of this section provides further background. In section 2 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 3 we use this approach to prove Theorem 2. In section 4 we show that the same message passing algorithm of section 2 produces approximate solutions of the TAP equations. Finally, section 5 discusses a generalization of Theorem 2 using universality. The reader interested in a succinct description of the algorithm (with some technical bells and whistles removed), is urged to read Appendix B.

1.1. Further background. As mentioned above—under suitable complexity theory assumptions—there is no polynomial-time algorithm that approximates the quadratic program (1.1) better than within a factor $O((\log n)^c)$, for some $c > 0$ [ABE+05]. 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 [Gam18] proved that exact computation of the partition function $Z_n(\beta)$ is hard on average.

A natural approach to the quadratic program (1.1) would be to use a convex relaxation. A spectral relaxation yields $\max_{\boldsymbol{\sigma}\in\{+1,-1\}} H_n(\boldsymbol{\sigma})/n \le \lambda_1(\boldsymbol{A})/2 = 1 + o_n(1)$, and hence is not tight for large n. This can be compared to a numerical evaluation of Parisi's formula which yields $\mathsf{P}_* \approx 0.763166$ [CR02, Sch08]. Rounding the spectral solution yields $H_n(\boldsymbol{\sigma}_{\rm sp}) = 2/\pi + o_n(1) \approx 0.636619$. 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 [MS16]. After a preprint of this paper was posted, two groups [KB19, MRX19] 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 [Mon16], based on insights from statistical physics [CK94, BCKM98]. The same presentation also outlined the basic strategy followed in the present paper, which uses an iterative approximate message passing (AMP) algorithm. These types of algorithms were first proposed in the context of signal processing and compressed sensing [Kab03, DMM09]. Their rigorous analysis was developed by Bolthausen [Bol14] and subsequently generalized in several papers [BM11, JM13, BLM15, BMN19]. 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,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

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 [MPV87]. Roughly speaking, the structure of the overlap distribution should reflect the connectivity properties of the level sets $\mathcal{L}_n(\varepsilon) \equiv \{\boldsymbol{\sigma} : \ H_n(\boldsymbol{\sigma}) \ge (1-\varepsilon) \max_{\boldsymbol{\sigma}'} H_n(\boldsymbol{\sigma}')\}$. 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., [GS14].

Important progress towards clarifying this connection was achieved recently in two remarkable papers [ABM18, Sub18].

Addario-Berry and Maillard [ABM18] study an abstract optimization problem that is thought to capture some key features of 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 polynomial time in the problem dimensions. From an optimization perspective, the random energy model is somewhat unnatural, in that specifying an instance requires memory that is exponential in the problem dimensions.

Subag [Sub18] 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 unrelaxed problem. The mathematical insight of [Sub18] 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 [MV85, CPS18]) this extension is far from obvious. In particular, uniform control of the Hessian is not as straightforward as in [Sub18].

The algorithm presented here is partially inspired by [Sub18] (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 [Bol14, BM11, JM13, BMN19], its analysis does not require uniform control and is relatively simple.

Finally, a conference version of this work was presented at FOCS 2019. The present version of the manuscript analyzes a simpler version of the algorithm, at the cost of some additional technical work in the proofs. The previous version introduced some truncation steps in the algorithm, which were unnecessary but simplified the proofs.

1.2. 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. Further, we denote by $\boldsymbol{x} \odot \boldsymbol{y} \in \mathbb{R}^n$ their entrywise product. 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

$$\hat{p}_{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k} \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_{1,i}, \dots, x_{k,i})} \, . \tag{1.7}$$

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 $\boldsymbol{X}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

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 of order ℓ (and we write $f \in \mathrm{PL}(\ell)$) if $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \le C(1 + \|\boldsymbol{x}\|^{\ell-1} + \|\boldsymbol{y}\|^{\ell-1})\|\boldsymbol{x} - \boldsymbol{y}\|$. Notice that PL(1) is the class of Lipschitz functions, and $f \in \mathrm{PL}(\ell)$, $g \in \mathrm{PL}(\ell')$ implies $fg \in \mathrm{PL}(\ell + \ell')$.

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

$$W_\ell(\mu,\nu) \equiv \left\{ \inf_{\gamma \in \mathcal{C}(\mu,\nu)} \int \|\boldsymbol{x} - \boldsymbol{y}\|_2^\ell \gamma(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{y}) \right\}^{1/\ell}, \tag{1.8}$$

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 $(\mu_n)_{n\ge 1}$, and μ on $\mathbb{R}^d$, we say that μ_n converges in Wasserstein distance to μ (and write $\mu_n \xrightarrow{W_\ell} \mu$) if $\lim_{n\to\infty} W_\ell(\mu_n, \mu) = 0$. It is well known that $\mu_n \xrightarrow{W_\ell} \mu$ 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 $\psi \in \mathrm{PL}(\ell)$. In turn, this happens if the convergence holds for all bounded Lipschitz functions ψ, and for $\psi(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^\ell$ [Vil08, 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_∞.

2. A general message passing algorithm.

2.1. Approximate message passing (AMP) iteration. Our algorithm is based on the general approximate message passing (AMP) iteration. Consider a sequence of (weakly differentiable) functions $f_k : \mathbb{R}^{k+2} \to \mathbb{R}$, and an initialization $\boldsymbol{u}^0 \in \mathbb{R}^n$ and additional vector $\boldsymbol{y} \in \mathbb{R}^n$ independent of $\boldsymbol{A}$. The AMP iteration is defined by letting, for $k \ge 0$,

$$\begin{aligned} \boldsymbol{z}^{k+1} &= \boldsymbol{A}\, f_k(\boldsymbol{z}^0, \dots, \boldsymbol{z}^k; \boldsymbol{y}) - \sum_{j=1}^{k} \mathsf{b}_{k,j} f_{j-1}(\boldsymbol{z}^0, \dots, \boldsymbol{z}^{j-1}; \boldsymbol{y}), \\ \mathsf{b}_{k,j} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f_k}{\partial z_i^j}(z_i^0, \dots, z_i^k; y_i). \end{aligned} \tag{2.1}$$

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

PROPOSITION 2.1. *Consider the AMP iteration* (2.1)*, and assume $f_k : \mathbb{R}^{k+2} \to \mathbb{R}$ to be pseudo-Lipschitz functions of order m. Further assume that for each ℓ, $\hat{p}_{\boldsymbol{z}^0,\boldsymbol{y}} \xrightarrow{W_\ell} p_{Z_0,Y}$ where $p_{Z_0,Y}$ is a probability distribution on $\mathbb{R}^2$ with finite moments of all orders ($\int (|z_0| + |y|)^\ell p_{Z_0,Y}(\mathrm{d}z_0, \mathrm{d}y) < \infty$), and $p_{Z_0,Y} = p_{Z_0} \otimes p_Y$, where each of p_{Z_0}, p_Y is either deterministic or has a Lebesgue density.*

Let $(Z_j)_{j\ge 1}$ be a centered Gaussian process independent of (Z_0, Y) with covariance $\boldsymbol{Q} = (Q_{kj})_{k,j\ge 1}$ determined recursively via

$$Q_{k+1,j+1} = \mathbb{E}\big\{ f_k(Z_0, \dots, Z_k; Y) f_j(Z_0, \dots, Z_j; Y) \big\}, \qquad k, j \ge 0. \tag{2.2}$$

Assume that $\boldsymbol{Q}_{\le k} := (Q_{ij})_{1\le i,j\le k}$ is strictly positive definite for each $k \le T$.

Then for any $k, \ell \in \mathbb{N}$, $k \le T$, and any function $\psi : \mathbb{R}^{k+2} \to \mathbb{R}$ pseudo-Lipschitz

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

of order ℓ, we have

$$\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi(z_i^0,\dots,z_i^k;y_i)=\mathbb{E}\psi(Z_0,\dots,Z_k;Y)\,. \tag{2.3}$$

This proposition follows immediately from the general analysis of AMP algorithms developed in [JM13, BMN19]. The only difference with respect to earlier results is in the fact that we assume f_k to be pseudo-Lipschitz rather than Lipschitz continuous. This generalization is established in Appendix A.

2.2. Incremental approximate message passing (IAMP). We next consider a special case of the general AMP setting. Fix two parameters $\delta>0$, $s\ge 0$, and functions $\widehat{g}_k:\mathbb{R}\to\mathbb{R}$, $k\in\mathbb{N}$, for $k\in\mathbb{Z}$, $v:\mathbb{R}\times\mathbb{R}_{\ge 0}\to\mathbb{R}$. We consider the general iteration (2.1), with the following choice of functions f_k:

$$f_k(z_0,\dots,z_k;y):=\sum_{\ell=1}^{k}\widehat{g}_\ell\left(x_{\ell-1}\right)\cdot(z_\ell-z_{\ell-1})+y\,, \tag{2.4}$$

$$x_k=x_{k-1}+v(x_{k-1},k\delta)\,\delta+s\,(z_k-z_{k-1})\,,\quad x_0=0\,. \tag{2.5}$$

Recall our convention $f_j=0$ for $j<0$. We further set $\widehat{g}_j=0$ for $j<1$. We further set $p_{Z_0,Y}=\delta_0\otimes\mathsf{N}(0,\delta)$. In other words, we set the initialization $\boldsymbol{z}^0=\mathbf{0}$, and the additional randomness $\boldsymbol{y}$ with independent and identically distributed (i.i.d.) Gaussian coordinate with variance δ. We note that, by (2.5), x_k is indeed a function of $z_0,\dots,z_k$, and therefore, f_k is a function of $z_0,\dots,z_k$ as stated.

LEMMA 2.2 (state evolution for IAMP). *Consider the incremental AMP iteration, and assume $v:\mathbb{R}\times\mathbb{R}_{\ge 0}\to\mathbb{R}$ and $\widehat{g}_k:\mathbb{R}\to\mathbb{R}$ to be such that, for each k, $v(\,\cdot\,,k\delta),\widehat{g}_k:\mathbb{R}\to\mathbb{R}$ are Lipschitz continuous.*

Let $(Z_j^\delta)_{j\ge 0}$ be a Gaussian martingale with $Z_0^\delta=0$, and variance $\mathrm{Var}(Z_k^\delta)=\mathbb{E}\{(Z_k^\delta)^2\}=q_k$ defined recursively by setting $X_0^\delta=0$, $q_0=0$, $q_1=\delta$, and for $k\ge 1$,

$$\begin{aligned}q_{k+1}&=q_k+\mathbb{E}\{\widehat{g}_k(X_{k-1}^\delta)^2\}\cdot(q_k-q_{k-1})\,,\\ X_k^\delta&=X_{k-1}^\delta+v(X_{k-1}^\delta;k\delta)\,\delta+s\,(Z_k^\delta-Z_{k-1}^\delta)\,.\end{aligned} \tag{2.6}$$

Assume that, for any $k\le T$, $\mathbb{E}\{\widehat{g}_k(X_{k-1}^\delta)^2\}>0$ strictly.

Then for any $k\in\mathbb{N}$, $k\le T$, and any pseudo-Lipschitz function $\psi:\mathbb{R}^{k+2}\to\mathbb{R}$, we have

$$\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi(z_i^0,\dots,z_i^k,y_i)=\mathbb{E}\psi(Z_0^\delta,\dots,Z_k^\delta,Y)\,. \tag{2.7}$$

Before passing to the proof of this lemma, we notice that (2.6) provides a recursive definition of the joint distribution of $(X_\ell^\delta,Z_\ell^\delta)_{\ell\ge 0}$. Namely, assume the joint distribution of $(X_\ell^\delta,Z_\ell^\delta)_{0\le\ell\le k-1}$ is given, with $(Z_\ell^\delta)_{0\le\ell\le k-1}$ a (centered) Gaussian martingale with $\mathbb{E}\{(Z_\ell^\delta)^2\}=q_\ell$. In order to extend this distribution, we set $q_k=q_{k-1}+\mathbb{E}\{\widehat{g}_{k-1}(X_{k-2}^\delta)^2\}\cdot(q_{k-1}-q_{k-2})$, define Z_k^δ by letting $Z_k^\delta-Z_{k-1}^\delta\sim\mathsf{N}(0,q_k-q_{k-1})$ independent of $(X_\ell^\delta,Z_\ell^\delta)_{0\le\ell\le k-1}$, and then define X_k^δ using (2.6).

Proof of Lemma 2.2. Consider (2.4), (2.5), and note that, for any k, x_k is a Lipschitz function of $z_0,\dots,z_k$. Hence f_k defined in (2.4) is pseudo-Lipschitz continuous,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

and we can therefore apply Proposition 2.1. Note that $(Z_j^\delta)_{j\ge 1}$ is a Gaussian process with covariance $(Q_{jk})_{j,k\ge 1}$ determined by (2.2). Setting $Z^\delta = Z_0 = 0$ with probability one, $(Z_j^\delta)_{j\ge 0}$ is also a Gaussian process with covariance $\boldsymbol{Q} = (Q_{jk})_{j,k\ge 0}$ which extends the previous one with $Q_{k,0} = Q_{0,k} = 0$ for all $k \ge 0$. For any $k \in \mathbb{N}$ such that $\boldsymbol{Q}_{\le k} := (Q_{jl})_{1\le j,l\le k}$ is strictly positive definite (see Proposition 2.1),

$$\frac{1}{n}\sum_{i=1}^n \psi(z_i^0,\dots,z_i^k) \xrightarrow{p} \mathbb{E}\psi(Z_0^\delta,\dots,Z_k^\delta)\,. \tag{2.8}$$

We next set $q_k = Q_{kk}$. We will prove that for all $k \ge 0$, the following claim, denoted by $\mathcal{C}(k)$, holds: for all $0 \le j \le k$ we have $Q_{k,j} = Q_{j,k} = Q_{j,j}$. We prove this claim by induction over k. As a preliminary remark notice that $\mathcal{C}(k)$ is equivalent to the claim that, defining $U_j^\delta = Z_j^\delta - Z_{j-1}^\delta$, the $(U_j^\delta)_{1\le j\le k}$ are independent Gaussian random variables. This can be easily checked by computing the correlation coefficient (for $j < l \le k$)

$$\mathbb{E}\{U_j^\delta U_l^\delta\} = Q_{jl} - Q_{j-1,l} - Q_{j,l-1} + Q_{j-1,l-1}\,. \tag{2.9}$$

This vanishes for all $1 \le j < l \le k$ if and only if $\mathcal{C}(k)$ holds. Note that $\mathcal{C}(k)$ is also equivalent to the claim that $(Z_j^\delta)_{0\le j\le k}$ is a Gaussian martingale.

We next proceed to prove the claim by induction. For $k = 0$ there is nothing to prove. We assume next that $\mathcal{C}(k)$ holds, and prove $\mathcal{C}(k+1)$. We need to prove that $Q_{l,k+1} = Q_{ll}$ for all $0 \le l \le k$. For $l = 0$ this is immediate since $Q_{k+1,0} = 0$ by Proposition 2.1.

Next consider $l = j+1$, $0 \le j \le k$. By (2.2) we have (recalling that $U_j^\delta := Z_j^\delta - Z_{j-1}^\delta$)

$$\begin{aligned} Q_{j+1,k+1} &= \mathbb{E}\left\{\left(Y + \sum_{i=1}^j \widehat{g}_i(X_{i-1}^\delta)U_i^\delta\right)\left(Y + \sum_{m=1}^k \widehat{g}_m(X_{m-1}^\delta)U_m^\delta\right)\right\} \\ &= \delta + \sum_{i=1}^j\sum_{m=1}^k \mathbb{E}\big\{\widehat{g}_i(X_{i-1}^\delta)U_i^\delta\,\widehat{g}_m(X_{m-1}^\delta)U_m^\delta\big\}\,. \end{aligned}$$

Recall that by the induction hypothesis $U_1^\delta,\dots,U_k^\delta$ are independent, and further that X_q^δ is a function of $U_1^\delta,\dots,U_q^\delta$. It follows that the last expectation is nonvanishing only if $i = m$, and therefore, we get

$$\begin{aligned} Q_{j+1,k+1} &= \delta + \sum_{l=1}^j \mathbb{E}\big\{\widehat{g}_l(X_{l-1}^\delta)^2(U_l^\delta)^2\big\} \\ &= \delta + \sum_{l=1}^j \mathbb{E}\big\{\widehat{g}_l(X_{l-1}^\delta)^2\big\}\,\mathbb{E}\{(U_l^\delta)^2\} = Q_{j+1,j+1}\,. \end{aligned}$$

This concludes the proof of claim $\mathcal{C}(k+1)$, and therefore the induction.

Also note that the last equation, applied to $j = k$, yields, for $k \ge 1$

$$\begin{aligned} q_{k+1} &= \delta + \sum_{l=1}^k \mathbb{E}\big\{\widehat{g}_l(X_{l-1}^\delta)^2\big\}(q_l - q_{l-1}) \\ &= q_k + \mathbb{E}\big\{\widehat{g}_k(X_{k-1}^\delta)^2\big\}(q_k - q_{k-1})\,. \end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Further, the first equality here holds for $k = 0$ as well, implying $q_1 = \delta$.

Finally notice that, under the assumption $\mathbb{E}\{\widehat{g}_k(X^\delta_{k-1})^2\} > 0$, $(q_k)_{\geq 0}$ is a strictly increasing sequence, and therefore, $\boldsymbol{Q}_{\leq k} := (Q_{jl})_{1\leq j,l\leq k}$ is strictly positive definite for all $k \leq T$, thus checking the assumptions of Proposition 2.1. This concludes the proof. ∎

We are now in position to define the output of the message passing algorithm (this will be our candidate for a near optimum of problem (1.1), after suitable rounding). We fix $q > 0$ and define (recalling the definition of f_k in (2.4), (2.5))

$$\boldsymbol{m}^k := f_k(\boldsymbol{z}^0, \dots, \boldsymbol{z}^k; \boldsymbol{0}) = \sum_{j=1}^{k} \widehat{g}_j\big(\boldsymbol{x}^{j-1}\big) \odot (\boldsymbol{z}^j - \boldsymbol{z}^{j-1}), \tag{2.10}$$

$$\boldsymbol{m} := \boldsymbol{m}^{\lfloor q/\delta \rfloor}. \tag{2.11}$$

Note that the vector $\boldsymbol{m} \in \mathbb{R}^n$ depends on parameters δ, q, s, and on the functions $\widehat{g}_k, v$, for $k \geq 0$. Parameter δ will be taken small enough but independent of n. The next section will be devoted to choosing q and the functions $\widehat{g}_k, v$. In this section we will establish some general properties of $\boldsymbol{m}$.

LEMMA 2.3. *Consider the IAMP iteration, and assume $v : \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ and $\widehat{g}_k : \mathbb{R} \to \mathbb{R}$ to satisfy the assumptions of Lemma 2.2. Further, assume $\partial_x \widehat{g}_k(x)$ to exist and be Lipschitz continuous.*

Define the random variables

$$M^\delta_k := \sum_{j=1}^{k} \widehat{g}_j(X^\delta_{j-1})\,(Z^\delta_j - Z^\delta_{j-1}), \qquad M^\delta := M^\delta_{\lfloor q/\delta \rfloor}. \tag{2.12}$$

Then we have, for any ℓ, and any pseudo-Lipschitz function of order ℓ, $\psi : \mathbb{R} \to \mathbb{R}$, and any $k \leq \lfloor q/\delta \rfloor$,

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \psi(m^k_i) = \mathbb{E}\{\psi(M^\delta_k)\}, \tag{2.13}$$

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{2n} \langle \boldsymbol{m}, \boldsymbol{A}\boldsymbol{m}\rangle = \sum_{k=1}^{\lfloor q/\delta \rfloor - 1} \mathbb{E}\{(Z^\delta_k - Z^\delta_{k-1})^2\}\, \mathbb{E}\{\widehat{g}_{k+1}(X^\delta_k)\}\, \mathbb{E}\{\widehat{g}_k(X^\delta_{k-1})^2\}. \tag{2.14}$$

Proof. Equation (2.13) follows immediately from Lemma 2.2 upon noticing that m^k_i is a pseudo-Lipschitz function of z^0_i, …, z^k_i (because $\widehat{g}_j(x^{j-1}_i)$ and $z^j_i - z^{j-1}_i$ are both Lipschitz). Hence $\psi(m_i)$ is a pseudo-Lipschitz function of $z_{0,i}$, …, $z_{k,i}$.

In order to prove (2.14), we will write $\boldsymbol{g}^k = \widehat{g}_k\left(\boldsymbol{x}^{k-1}\right) \odot (\boldsymbol{z}^k - \boldsymbol{z}^{k-1})$ and $K = \lfloor q/\delta \rfloor$. With these notations, we have

$$\boldsymbol{m} = \sum_{j=1}^{K} \boldsymbol{g}^j, \qquad \boldsymbol{m}^\ell = \sum_{j=1}^{\ell} \boldsymbol{g}^j, \tag{2.15}$$

$$\boldsymbol{z}^{k+1} = \boldsymbol{A}(\boldsymbol{m}^k + \boldsymbol{y}) - \sum_{j=1}^{k} \mathsf{b}_{kj} \boldsymbol{m}^{j-1}. \tag{2.16}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

We further notice that, for $j \le k$

$$\mathsf{b}_{k,j} = \frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{k}\left\{\frac{\partial \widehat{g}_\ell}{\partial z_j}(x_i^{\ell-1})(z_i^\ell - z_i^{\ell-1}) + \widehat{g}_\ell(x_i^{\ell-1})(\mathbf{1}_{\ell=j} - \mathbf{1}_{\ell=j+1})\right\}.$$

In what follows we will define $\mathsf{b}_{k,j} = 0$ for $j > k$, and recall that the random variables X_k^δ, U_k^δ are defined via (2.6) and $U_k^\delta := Z_k^\delta - Z_{k-1}^\delta$. We thus have

$$\begin{aligned}
&\operatorname*{p-lim}_{n\to\infty}(\mathsf{b}_{k,j} - \mathsf{b}_{k-1,j})\\
&= \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial \widehat{g}_k}{\partial z_j}(x_i^{k-1})(z_i^k - z_i^{k-1}) + \widehat{g}_k(x_i^{k-1})(\mathbf{1}_{k=j} - \mathbf{1}_{k=j+1})\right\}\\
&\overset{P(\mathrm{a})}{=} \mathbb{E}\left\{\frac{\partial \widehat{g}_k}{\partial z_j}(X_{k-1}^\delta)(Z_k^\delta - Z_{k-1}^\delta) + \widehat{g}_k(X_{k-1}^\delta)(\mathbf{1}_{k=j} - \mathbf{1}_{k=j+1})\right\}\\
&\overset{(\mathrm{b})}{=} \mathbb{E}\left\{\widehat{g}_k(X_{k-1}^\delta)\right\}(\mathbf{1}_{k=j} - \mathbf{1}_{k=j+1}) := \tilde{\mathsf{b}}_k(\mathbf{1}_{k=j} - \mathbf{1}_{k=j+1}),
\end{aligned} \tag{2.17}$$

where in (a) we used the fact that both $\partial_x \widehat{g}_k$ and $\widehat{g}_k$ are Lipschitz, and x_i^{k-1} is a Lipschitz function of $z_i^0, \dots, z_i^{k-1}$. In (b) we used the martingale property of $(Z_j^\delta)_{j\ge 0}$. Further, the random variables U_k^δ, X_k^δ are defined as in the proof of Lemma 2.2.

Next, notice that, for $j < k$, since $(U_i^\delta)_{i\ge 0}$ are martingale differences, and X_{k-1}^δ, X_{j-1}^δ are functions of $(U_i^\delta)_{0\le i\le k-1}$,

$$\begin{aligned}
\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle \boldsymbol{g}^j, \boldsymbol{g}^k\rangle &= \mathbb{E}\big\{\widehat{g}_j(X_{j-1}^\delta)\, U_j^\delta \widehat{g}_k(X_{k-1}^\delta)\, U_k^\delta\big\}\\
&= \mathbb{E}\big\{\widehat{g}_j(X_{j-1}^\delta)\, U_j^\delta \widehat{g}_k(X_{k-1}^\delta)\big\}\mathbb{E}\big\{U_k^\delta\big\} = 0.
\end{aligned} \tag{2.18}$$

By a similar argument, for $j \le k$,

$$\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle \boldsymbol{g}^j, (\boldsymbol{z}^{k+1} - \boldsymbol{z}^k)\rangle = 0. \tag{2.19}$$

On the other hand,

$$\begin{aligned}
\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\|\boldsymbol{g}^k\|^2 &= \mathbb{E}\big\{\widehat{g}_k(X_{k-1}^\delta)^2\,(U_k^\delta)^2\big\}\\
&= \mathbb{E}\big\{\widehat{g}_k(X_{k-1}^\delta)^2\big\}\,\mathbb{E}\big\{(U_k^\delta)^2\big\}.
\end{aligned} \tag{2.20}$$

By taking the difference of the AMP iterations (2.16) at two subsequent times, we get

$$\boldsymbol{A}\boldsymbol{g}^k = \boldsymbol{z}^{k+1} - \boldsymbol{z}^k + \sum_{j=1}^{k}(\mathsf{b}_{k,j} - \mathsf{b}_{k-1,j})\boldsymbol{m}^{j-1} \tag{2.21}$$

$$= \boldsymbol{z}^{k+1} - \boldsymbol{z}^k + \tilde{\mathsf{b}}_k \boldsymbol{g}^{k-1} + \mathbf{err}^k. \tag{2.22}$$

Using (2.17) and (2.20), we get $\|\mathbf{err}^k\|^2/n \xrightarrow{p} 0$, and therefore, $\langle \boldsymbol{g}^j, \mathbf{err}^k\rangle/n \xrightarrow{p} 0$. Hence, for $j \le k$,

$$\begin{aligned}
\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle \boldsymbol{g}^j, \boldsymbol{A}\boldsymbol{g}^k\rangle &= \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle \boldsymbol{g}^j, \boldsymbol{z}^{k+1} - \boldsymbol{z}^k\rangle + \tilde{\mathsf{b}}_k \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle \boldsymbol{g}^j, \boldsymbol{g}^{k-1}\rangle + \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle \boldsymbol{g}^j, \mathbf{err}^k\rangle\\
&\overset{(\mathrm{a})}{=} \tilde{\mathsf{b}}_k \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\langle \boldsymbol{g}^j, \boldsymbol{g}^{k-1}\rangle\\
&\overset{(\mathrm{b})}{=} \mathbf{1}_{\{k=j+1\}}\mathbb{E}\{\widehat{g}_{j+1}(X_j^\delta)\}\mathbb{E}\big\{\widehat{g}_j(X_{j-1}^\delta)^2\big\}\,\mathbb{E}\big\{(U_j^\delta)^2\big\},
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

where (a) follows from (2.19) and (b) from (2.18) and (2.20). We finally can compute (using the fact that $\boldsymbol{A}$ is symmetric)

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

∎

2.3. Small step size limit. In the case of models with no overlap gap, it is natural to consider the limit of small step size $\delta\to 0$. In order to identify this limit, it is useful to summarize the equations that characterize the state evolution process at $\delta>0$ fixed:

- $(Z_k^\delta)_{k\ge 0}$ is a centered Gaussian martingale, with variance $\mathbb{E}\{(Z_k^\delta)^2\}=q_k$ determined by letting $q_0=0$, $q_1=\delta$ and, for $k\ge 1$ (cf. (2.6)),

$$q_{k+1}=q_k+\mathbb{E}\{\widehat{g}_k(X_{k-1}^\delta)^2\}\cdot(q_k-q_{k-1})\,. \tag{2.23}$$

- The process $\{(X_k^\delta,M_k^\delta)\}_{k\ge 0}$ is defined by $X_0^\delta=0$ and (cf. (2.6) and (2.12))

$$X_k^\delta=X_{k-1}^\delta+v(X_{k-1}^\delta;k\delta)\,\delta+s\,(Z_k^\delta-Z_{k-1}^\delta)\,, \tag{2.24}$$

$$M_k^\delta=\sum_{j=1}^{k}\widehat{g}_j(X_{j-1}^\delta)\,(Z_j^\delta-Z_{j-1}^\delta)\,. \tag{2.25}$$

We will choose functions $\widehat{g}_k$ so that $\mathbb{E}\{\widehat{g}_k(X_{k-1}^\delta)^2\}=1$. It is therefore natural to imagine that $(Z_k^\delta)_{k\ge 0}$ converges to Brownian motion. Motivated by this heuristics, we will introduce a stochastic differential equation (SDE) description.

Definition 2.4. *We say that the functions $g,v:\mathbb{R}\times\mathbb{R}_{\ge 0}\to\mathbb{R}$ are* acceptable *if the following conditions hold, for some constant C_0,C_1:*

$$|v(x,t)|\vee|g(x,t)|\le C_0(1+|x|)\,, \tag{2.26}$$

$$|v(x_1,t)-v(x_2,t)|\vee|g(x_1,t)-g(x_2,t)|\le C_1|x_1-x_2|\,. \tag{2.27}$$

Given acceptable functions $g,v:\mathbb{R}\times\mathbb{R}_{\ge 0}\to\mathbb{R}$, let $(B_t)_{t\ge 0}$ be a standard Brownian motion. We define the process $(X_t,M_t)_{t\ge 0}$ via

$$\mathrm{d}X_t=v(X_t,t)\,\mathrm{d}t+s\,\mathrm{d}B_t\,,\qquad \mathrm{d}M_t=g(X_t,t)\,\mathrm{d}B_t\,, \tag{2.28}$$

with initial condition $X_0=M_0=0$. Equivalently,

$$X_t=\int_0^t v(X_r,r)\,\mathrm{d}r+s\,B_t\,,\qquad M_t=\int_0^t g(X_r,r)\,\mathrm{d}B_r\,, \tag{2.29}$$

where the last integral is understood in Ito's sense. Existence and uniqueness of strong solutions of this SDE are proven—for instance—in [Øks03, Theorem 5.2.1].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

We will prove that indeed this SDE provides a good approximation on the state evolution process at $\delta > 0$. We begin by stating a few properties of the continuous time process (X_t, M_t). While these properties are standard, we provide proofs for the reader's convenience.

LEMMA 2.5. *Let $s \in \mathbb{R}$, and $g, v : \mathbb{R} \times \mathbb{R}_{\ge 0} \to \mathbb{R}$ be acceptable functions. Then there exists a constant C depending only on constants C_0 C_1 of Definition* 2.4*, such that the following hold for all $t, r \ge 0$:*

$$\mathbb{E}(X_t^2) \le e^{Ct} - 1\,, \tag{2.30}$$

$$\mathbb{E}(|X_t - X_r|^2) \le e^{C(t\vee r)}|t - r|\,. \tag{2.31}$$

Proof. Throughout the proof, C will denote a constant that depends on the bounds on g, v, and on s but can change from line to line. Consider (2.30). By Ito's formula

$$\mathrm{d}(X_t^2) = 2X_t v(X_t, t)\mathrm{d}t + 2sX_t \mathrm{d}B_t + s^2\mathrm{d}t\,, \tag{2.32}$$

whence, using $|xv(x,t)| \le C(1 + x^2)$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(X_t^2) = 2\mathbb{E}\{X_t v(X_t, t)\} + s^2 \le C\big\{1 + \mathbb{E}(X_t^2)\big\}\,. \tag{2.33}$$

Equation (2.30) follows by Gronwall's lemma. Equation (2.31) follows by essentially the same argument (fixing r and differentiating with respect to $t > r$). ∎

Before stating our approximation result, we recall the definition of functions of bounded total variation.

DEFINITION 2.6. *A function $f : [a, b] \to \mathbb{R}$ has* total variation *bounded by B if, for any $m \in \mathbb{N}$ and any $a = s_0 \le s_1 \le \cdots \le s_m = b$, we have*

$$\sum_{\ell=1}^{m} \big|f(s_\ell) - f(s_{\ell-1})\big| \le B\,. \tag{2.34}$$

The total variation of f (denoted by $\|f\|_{\mathrm{TV}}$) is the least constant B such that this bound holds for all m and all partitions $(s_j)_{j\le m}$.

LEMMA 2.7. *Given $s \ge 0$ and acceptable functions $g, v : \mathbb{R} \times \mathbb{R}_{\ge 0} \to \mathbb{R}$, let (X_t, M_t) be the process defined by* (2.28), (2.29)*. Let $f : \mathbb{R} \times \mathbb{R}_{\ge 0} \to \mathbb{R}$ be acceptable and further assume that $f(x,t) = f_0(t)\, f_1(x,t)$ where f_0 is bounded and has bounded total variation on $[0, T]$, and $f_1 : \mathbb{R} \times [0, T] \to \mathbb{R}$ is Lipschitz continuous. Define $t_j = j\delta$ for $j \in \mathbb{N}$.*

Then there exists a constant C (depending on the constants C_0, C_1 of Definition 2.4 *on T, on $\|f_0\|_{\mathrm{TV}}$, $\|f_0\|_\infty$, and on the Lipschitz constant of f_1) such that*

$$\sum_{j=0}^{\lfloor T/\delta\rfloor - 1} \int_{t_j}^{t_j+\delta} \mathbb{E}\big\{\big[f(X_t, t) - f(X_{t_j}, t_{j+1})\big]^2\big\}\,\mathrm{d}t \le C\delta\,. \tag{2.35}$$

Proof. Note that, for $t \in [0, T]$, $|f(x_1,t) - f(x_2,t)| \le \|f_0\|_\infty |f_1(x_1,t) - f_1(x_2,t)| \le$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

$C|x_1 - x_2|$. We have

$$\begin{aligned}
&\mathbb{E}\big\{\big[f(X_t,t) - f(X_{t_j},t_{j+1})\big]^2\big\}\\
&\le 2\,\mathbb{E}\big\{\big[f(X_t,t) - f(X_{t_j},t)\big]^2\big\} + 2\,\mathbb{E}\big\{\big[f(X_{t_j},t) - f(X_{t_j},t_{j+1})\big]^2\big\}\\
&\le C\,\mathbb{E}\{(X_t - X_{t_j})^2\} + 4\,\mathbb{E}\big\{\big[f_0(t)f_1(X_{t_j},t) - f_0(t)f_1(X_{t_j},t_{j+1})\big]^2\big\}\\
&\quad + 4\,\mathbb{E}\big\{\big[f_0(t)f_1(X_{t_j},t_{j+1}) - f_0(t_{j+1})f_1(X_{t_j},t_{j+1})\big]^2\big\}\,.
\end{aligned}$$

By Lemma 2.5, we have $\mathbb{E}\{|X_t - X_s|^2\} \le C|t-s|$ for all $t, s \le T$, whence, for $t \in [t_j, t_{j+1}]$,

$$\begin{aligned}
\mathbb{E}\big\{\big[f(X_t,t) - f(X_{t_j},t_{j+1})\big]^2\big\} &\le C\delta + Cf_0(t)^2\delta^2 + C[f_0(t) - f_0(t_{j+1})]^2\mathbb{E}\{1 + X_{t_j}^2\}\\
&\le C\delta + C\|f_0\|_\infty[f_0(t) - f_0(t_{j+1})]^2\\
&\le C\delta + C|f_0(t) - f_0(t_{j+1})|\,.
\end{aligned}$$

We thus have, letting $K := \lfloor T/\delta \rfloor$,

$$\sum_{j=0}^{K-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{\big[f(X_t,t) - f(X_{t_j},t_{j+1})\big]^2\big\}\,\mathrm{d}t \tag{2.36}$$

$$\le C\sum_{j=0}^{K-1}\int_{t_j}^{t_j+\delta}\big[\delta + |f_0(t) - f_0(t_{j+1})|\big]\mathrm{d}t \tag{2.37}$$

$$\le C\delta + C\delta\sup_{(t'_j)_{j<K}}\sum_{j=0}^{K-1}|f_0(t'_j) - f_0(t_{j+1})|\,, \tag{2.38}$$

where in the last expression, the supremum is over sequences of points $t'_j \in [t_j, t_{j+1}]$. It is immediate to see that $\sum_{j=0}^{K-1}|f_0(t'_j) - f_0(t_{j+1})| \le \|f_0\|_{\mathrm{TV}}$, whence the claim follows. ∎

LEMMA 2.8. *Given $s \ge 0$ and acceptable functions $g, v : \mathbb{R}\times\mathbb{R}_{\ge 0} \to \mathbb{R}$, let (X_t, M_t) be the process defined by (2.28), (2.29). Further, assume the following:*

(a) $\mathbb{E}\{g(X_t,t)^2\} = 1$ *for all* $t \ge 0$.
(b) *For each of $f \in \{g, v\}$, we have $f(x,t) = f_0(t)\,f_1(t,x)$, where f_0 is bounded and has bounded total variation on $[0,q]$, and $f_1 : \mathbb{R}\times[0,q] \to \mathbb{R}$ is Lipschitz continuous.*

Consider the state evolution iteration of (2.6), whereby $q_0 = 0$, $q_1 = \delta$, and $\widehat{g}_k$ is defined recursively via

$$\widehat{g}_k(x) \equiv \frac{g(x, k\delta)}{\mathbb{E}\{g(X^\delta_{k-1}, k\delta)^2\}^{1/2}}\,. \tag{2.39}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

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

$$\max_{k\le \lfloor q/\delta\rfloor} \mathbb{E}\big(|X_k^\delta - X_{k\delta}|^2\big) \le C\delta\,, \tag{2.40}$$

$$\max_{k\le \lfloor q/\delta\rfloor} \mathbb{E}\big(|M_k^\delta - M_{k\delta}|^2\big) \le C\sqrt{\delta}\,, \tag{2.41}$$

$$\mathbb{E}\big(|M_{\lfloor q/\delta\rfloor}^\delta - M_q|^2\big) \le C\sqrt{\delta}\,, \tag{2.42}$$

$$\sum_{k=1}^{\lfloor q/\delta\rfloor -1} \mathbb{E}\{(U_k^\delta)^2\}\,\mathbb{E}\{\widehat{g}_{k+1}(X_k^\delta)\}\,\mathbb{E}\{\widehat{g}_k(X_{k-1}^\delta)^2\} = \int_0^q \mathbb{E}\{g(X_t,t)\}\,\mathrm{d}t + O(\delta^{1/4})\,. \tag{2.43}$$

(Here C *is a constant depending only on the constant appearing in the assumptions on* g, v *and on* q, s. *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 q and on the constants appearing in the assumptions on g, v, and can change from line to line. Note that, by construction, $q_j = j\delta$ for all j. Hence we can construct the discrete and continuous processes on the same probability space by letting $Z_j^\delta = B_{t_j}$ for all $j \ge 0$. Recalling that $U_j^\delta := Z_j^\delta - Z_{j-1}^\delta$, we also have $(U_j^\delta)_{j\ge 1} \sim_{iid} \mathsf{N}(0,\delta)$.

We then state the difference between the two processes as

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

By taking the second moment, and using the Cauchy–Schwarz inequality, we get

$$\mathbb{E}\big\{\big[X_{k\delta} - X_k^\delta\big]^2\big\} \le k\sum_{j=0}^{k-1} \delta \int_{t_j}^{t_j+\delta} \mathbb{E}\big\{\big[v(X_t,t) - v(X_j^\delta, t_{j+1})\big]^2\big\}\,\mathrm{d}t\,. \tag{2.44}$$

By Lemma 2.7, and using $k\delta \le q \le C$,

$$\mathbb{E}\big\{\big[X_{k\delta} - X_k^\delta\big]^2\big\} \le 2k\delta \sum_{j=0}^{k-1} \int_{t_j}^{t_j+\delta} \mathbb{E}\big\{\big[v(X_{t_j}, t_{j+1}) - v(X_j^\delta, t_{j+1})\big]^2\big\}\,\mathrm{d}t + C\delta \tag{2.45}$$

$$\le C\sum_{j=0}^{k-1} \int_{t_j}^{t_j+\delta} \mathbb{E}\big\{\big(X_{t_j} - X_j^\delta\big)^2\big\}\,\mathrm{d}t + C\delta\,. \tag{2.46}$$

Letting $\Delta_k \equiv \mathbb{E}\{[X_{t_k} - X_k^\delta]^2\}$, we get

$$\Delta_k \le C\delta \sum_{j=0}^{k-1} \Delta_j + C\delta\,. \tag{2.47}$$

By Gronwall's inequality, this implies the bound $\mathbb{E}(|X_{t_k} - X_k^\delta|^2) \le C\delta$ as stated in (2.40).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

In order to prove (2.41), note that

$$\begin{aligned}\Big|\mathbb{E}\{g(X^\delta_{k-1},t_k)^2\}-\mathbb{E}\{g(X_{t_k},t_k)^2\}\Big| &\\ &\overset{(a)}{\le} \mathbb{E}\big\{[g(X^\delta_{k-1},t_k)-g(X_{t_k},t_k)]^2\big\}^{1/2}\mathbb{E}\{[g(X^\delta_{k-1},t_k)+g(X_{t_k},t_k)]^2\}^{1/2}\\ &\overset{(b)}{\le} C\,\mathbb{E}\big\{|X^\delta_{k-1}-X_{t_k}|^2\big\}^{1/2}\mathbb{E}\{1+(X^\delta_{k-1})^2+X^2_{t_k}\}^{1/2}\\ &\overset{(c)}{\le} C\sqrt{\delta}\,,\end{aligned}$$

where (a) is Cauchy–Schwarz, (b) follows since $|g(x_1,t)-g(x_2,t)|\le\|g_0\|_\infty|g_1(x_1,t)-g(x_2,t)|\le C|x_1-x_2|$, and therefore, $|g(x,t)|\le|g(0,t)|+C|x|\le C(1+|x|)$, and (c) follows by Lemma 2.5 and (2.40). On the other hand,

$$\mathbb{E}\big\{|X^\delta_{k-1}-X_{t_k}|^2\big\}\le 2\,\mathbb{E}\big\{|X^\delta_{k-1}-X_{t_{k-1}}|^2\big\}+2\,\mathbb{E}\big\{|X_{t_{k-1}}-X_{t_k}|^2\big\}\le C\delta\,,$$

where we bounded the first term by (2.40), and the second by Lemma 2.5.

Since by assumption $\mathbb{E}\{g(X_{t_k},t_k)^2\}=1$, the last two displays imply $1-C\sqrt{\delta}\le\mathbb{E}\{g(X^\delta_{k-1},t_k)^2\}\le 1+C\sqrt{\delta}$. We thus obtain

$$\mathbb{E}\big\{[\widehat{g}_k(X^\delta_{k-1})-g(X^\delta_{k-1},t_k)]^2\big\}\le C\sqrt{\delta}\,. \tag{2.48}$$

We decompose the difference of M_t and M^δ_k as

$$M_{k\delta}-M^\delta_k=\sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta}\big[g(X_t,t)-\widehat{g}_{j+1}(X^\delta_j)\big]\,\mathrm{d}B_t\,.$$

Using the fact that X_t is measurable on $(B_s)_{s\le t}$ and X^δ_j is measurable on $(B_s)_{s\le t_j}$, we get, by Ito's isometry,

$$\mathbb{E}\big\{\big[M_{k\delta}-M^\delta_k\big]^2\big\}=\sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{[g(X_t,t)-\widehat{g}_{j+1}(X^\delta_j)]^2\big\}\,\mathrm{d}t\,.$$

Therefore,

$$\begin{aligned}\mathbb{E}\big(|M_{k\delta}-M^\delta_k|^2\big) &\le 2\sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{[g(X_t,t)-g(X^\delta_j,t_{j+1})]^2\big\}\,\mathrm{d}t\\ &\quad+2\sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{[\widehat{g}_{j+1}(X^\delta_j)-g(X^\delta_j,t_{j+1})]^2\big\}\,\mathrm{d}t\\ &\overset{(a)}{\le} 4\sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{[g(X_{t_j},t_{j+1})-g(X^\delta_j,t_{j+1})]^2\big\}\,\mathrm{d}t\\ &\quad+4\sum_{j=0}^{k-1}\int_{t_j}^{t_j+\delta}\mathbb{E}\big\{[g(X_t,t)-g(X_{t_j},t_{j+1})]^2\big\}\,\mathrm{d}t+C\sqrt{\delta}\\ &\overset{(b)}{\le} C\delta\sum_{j=0}^{k-1}\mathbb{E}\big\{(X_{t_j}-X^\delta_j)^2\big\}\,\mathrm{d}t+C\sqrt{\delta}+C\delta\\ &\overset{(c)}{\le}\sqrt{\delta}\,,\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

where (a) follows from (2.48), (b) from Lemma 2.7 and the Lipschitz assumption on g, and (c) from (2.40). This proves (2.41). The bound of (2.42) follows since, setting $K := \lfloor q/\delta \rfloor$, we have

$$\mathbb{E}\big(|M_q - M_{K\delta}|^2\big) = \int_{K\delta}^{q} \mathbb{E}\big\{g(X_t,t)^2\big\}\,\mathrm{d}t \le C\int_{K\delta}^{q} \mathbb{E}\big\{1+X_t^2\big\}\,\mathrm{d}t \le C\delta\,. \tag{2.49}$$

Finally, (2.43) 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 $s, q \ge 0$ and $g, v : \mathbb{R}\times[0,q] \to \mathbb{R}$ be acceptable functions. Define the process (X_t, M_t) using the SDE (2.28) with initial condition $X_0 = Z_0 = 0$. Assume the following:*

(a) $\mathbb{E}\{g(X_t,t)^2\} = 1$ *for all* $t \ge 0$.
(b) *For each of* $f \in \{g, v\}$, *we have* $f(x,t) = f_0(t)\,f_1(t,x)$, *where* f_0 *is bounded and has bounded total variation on* $[0,q]$, *and* $f_1 : \mathbb{R}\times[0,q] \to \mathbb{R}$ *is Lipschitz continuous.*
(c) $\partial_x g(x,t)$ *exists and is Lipschitz continuous.*

Define the incremental AMP iteration $(\boldsymbol{z}^k)_{k\ge 0}$, and let $\boldsymbol{m}$ be given by (2.10). Finally, let $\psi : \mathbb{R} \to \mathbb{R}$ be a pseudo-Lipschitz function of order ℓ. Then, for any $\varepsilon > 0$ there exist $\delta_*(\varepsilon) > 0$ such that if $\delta \le \delta_*(\varepsilon)$, we have

$$\left|\operatorname*{p-lim}_{n\to\infty} \frac{1}{2n}\langle \boldsymbol{m}, \boldsymbol{A}\boldsymbol{m}\rangle - \int_0^q \mathbb{E}\{g(X_t,t)\}\,\mathrm{d}t\right| \le \varepsilon\,, \tag{2.50}$$

$$\left|\operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(m_i) - \mathbb{E}\{\psi(M_q)\}\right| \le \varepsilon\,. \tag{2.51}$$

(Further, the above limits in probability are nonrandom quantities.)

Proof. Notice that the condition $\mathbb{E}\{\widehat{g}_k(X_{k-1}^{\delta})^2\} > 0$ of Lemmas 2.2 and 2.3 holds by construction (indeed, $\mathbb{E}\{\widehat{g}_k(X_{k-1}^{\delta})^2\} = 1$ by (2.39)). The conclusion follows immediately from Lemmas 2.3 and 2.8. ∎

Remark 2.1. Let us emphasize that this theorem does not use in any way Assumption 1. Indeed, it does not even refer to the original optimization problem (1.1); in particular, the constraint $\boldsymbol{\sigma} \in \{+1,-1\}^n$ does not play any role here. The theorem simply describes the behavior of a certain class of algorithms parametrized by functions v, g, and by the coefficient s. Depending on the choices of these functions, the algorithm will produce a good approximate solution of the original problem, or not.

3. Proof of the main theorem.

3.1. Choosing the nonlinearities. In view of Theorem 3, we need to choose the coefficients g, s, v in the SDE (2.28) to solve the following stochastic optimal

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

control problem:

$$\begin{aligned} &\text{maximize} && \int_0^q \mathbb{E}\{g(X_t,t)\}\,\mathrm{d}t \\ &\text{subject to} && \mathbb{P}(M_q \in [-1,1]) = 1\,, \qquad (3.1)\\ & && \mathrm{d}X_t = v(X_t,t)\,\mathrm{d}t + s\,\mathrm{d}B_t\,, \quad M_t = \int_0^t g(X_r,r)\,\mathrm{d}B_r\,, \qquad (3.2)\\ & && g, v \in \mathscr{A}, s \in \mathbb{R}_{\ge 0}\,, \end{aligned}$$

where $\mathscr{A}$ is the class of functions that satisfies the assumptions of Theorem 3. By that theorem, the value of this problem is the asymptotic optimal value achieved by the IAMP algoritm for problem (1.1).

For additional context, recall that Parisi formula (1.3) admits a dual formulation in terms of an optimal control problem. This problem was studied, among others, in [AC15, JT16], to establish uniqueness of the minimizer μ_β. The problem (3.1) is related but not equivalent to the one studied in [AC15, JT16]. Follow-up work [AMS20] studied solutions of the problem (3.1) in a broader context, and its relation with the Parisi formula.

Here we will not attempt to solve directly the problem (3.1), and instead we will compare it with the structure of the Parisi formula. This will motivate a guess for the two functions g, v and the constant s, 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).

Unless stated otherwise, in this section we set $\beta > \beta_0$ as per Assumption 1, and set $q = q_* = q_*(\beta)$ and $\mu = \mu_\beta$ the unique minimizer of the Parisi functional. We also fix Φ to be the solution of the PDE (1.2) with $\mu = \mu_\beta$.

There is a natural SDE associated with the Parisi variational principle that was first introduced in physics [Par80, SD84, MPV87], and recently studied in the probability theory literature [AC15, JT16]:

$$\mathrm{d}X_t = \beta^2\mu(t)\partial_x\Phi(t,X_t)\,\mathrm{d}t + \beta\,\mathrm{d}B_t\,, \qquad X_0 = 0\,. \tag{3.3}$$

Motivated by the comparison of this equation with (2.28), we set the coefficients g, s, v as follows:

$$v(x,t) = \beta^2\mu(t)\partial_x\Phi(t,x)\,, \quad s = \beta\,, \quad g(x,t) = \beta\partial_x^2\Phi(t,x)\,. \tag{3.4}$$

3.2. Analytical properties of the function Φ. In this section we collect a few useful properties of the function Φ that solves the Parisi PDE (1.2). Most of these properties are reproduced from earlier papers, and we only provide bibliographic references. For some of the identities, we explain the basic argument, for the reader's convenience.

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

LEMMA 3.1. *For any probability distribution $\mu \in \mathscr{P}([0,1])$, let $\Phi = \Phi_\mu$ be the corresponding solution of the Parisi PDE* (1.2)*. Then the following hold:*

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

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

$$\left|\partial_x \Phi(t,x)\right| \le 1\,, \quad 0 < \partial_x^2 \Phi(t,x) \le 1\,, \quad \left|\partial_x^3 \Phi(t,x)\right| \le 4\,. \tag{3.5}$$

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

(iv) $\partial_x^i \Phi(t,x)$ *is Lipschitz continuous on* $[0,1] \times \mathbb{R}$ *for all* $j \ge 0$.

Proof. Points (i) and (iii) are Theorem 4 in [JT16]. Point (ii) is Proposition 2(ii) in [AC15]. Finally, point (iv) follows immediately from points (i), (ii), and (iii). ∎

Lemma 3.1 will be used to prove that the choice (3.4) satisfies the regularity assumptions in Theorem 3. We next have to check the normalization condition and compute the value achieved by the algorithm.

LEMMA 3.2. *For* $\mu \in \mathscr{P}([0,1])$, *let* $\Phi = \Phi_\mu$ *be the corresponding solution of the Parisi PDE* (1.2). *Then we have*

$$M_t = \int_0^t \beta \partial_x^2 \Phi(t, X_s)\, \mathrm{d}B_s = \partial_x \Phi(t, X_t)\,. \tag{3.6}$$

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

Proof. This identity follows, e.g., from Lemma 2 in [AC15]. The basic argument is as follows: by differentiating the PDE (1.2) with respect to x, we obtain the following equation for $\Phi_x = \partial_x \Phi$:

$$\partial_t \Phi_x(t,x) + \frac{1}{2}\beta^2 \partial_x^2 \Phi_x(t,x) + \beta^2 \mu(t) \Phi_x(t,c) \partial_x \Phi_x(t,x) = 0\,. \tag{3.7}$$

Define $\overline{M}_t = \partial_x \Phi(t, X_t)$: our objective is to prove that $\overline{M}_t = M_t$. Notice that $x \mapsto \Phi(t,x)$ is an even function for all t, and therefore, $\overline{M}_0 = \partial_x \Phi(0,0) = 0 = M_0$. Further, by Ito's formula

$$\begin{aligned}
\mathrm{d}\overline{M}_t &= \partial_t \Phi_x(t, X_t)\, \mathrm{d}t + \partial_x \Phi_x(t, X_t)\, \mathrm{d}X_t + \frac{1}{2}\beta^2 \partial_x^2 \Phi_x(t, X_t)\, \mathrm{d}t \\
&= \Big\{ \partial_t \Phi_x(t, X_t) + \beta^2 \mu(t) \Phi_x(t,c) \partial_x \Phi_x(t,x) + \frac{1}{2}\beta^2 \partial_x^2 \Phi_x(t,x) \Big\} + \beta \partial_x \Phi_x(t, X_t)\, \mathrm{d}B_t \\
&= \beta \partial_x \Phi_x(t, X_t)\, \mathrm{d}B_t\,.
\end{aligned}$$

Therefore, we have $\mathrm{d}\overline{M}_t = \mathrm{d}M_t$, which proves our claim. Lemma 3.1(ii) implies $|M_t| \le 1$ almost surely. ∎

LEMMA 3.3. *Let* $\mu = \mu_\beta$ *and* $\Phi = \Phi_\mu$ *be the solution of the Parisi PDE* (1.2). *Then, for all* $t \in \mathrm{supp}(\mu_\beta)$, *we have*

$$\mathbb{E}\big\{ \big(\partial_x \Phi(t, X_t)\big)^2 \big\} = t\,, \tag{3.8}$$

$$\mathbb{E}\big\{ \big(\beta \partial_x^2 \Phi(t, X_t)\big)^2 \big\} \le 1\,. \tag{3.9}$$

Further, under Assumption 1, *for all* $0 \le t \le q_*$, *we have*

$$\mathbb{E}\big\{ \big(\beta \partial_x^2 \Phi(t, X_t)\big)^2 \big\} = 1\,. \tag{3.10}$$

Proof. Equations (3.8) and (3.9) are Proposition 1 in [Che17]. For (3.10) note that by (39) in the same paper, we have, for any $t_1 < t_2 \le q_*$,

$$\mathbb{E}\{(\partial_x \Phi(t_2, X_{t_2}))^2\} - \mathbb{E}\{(\partial_x \Phi(t_1, X_{t_1}))^2\} = \int_{t_1}^{t_2} \mathbb{E}\big\{ \big(\beta \partial_x^2 \Phi(t, X_t)\big)^2 \big\}\, \mathrm{d}t\,, \tag{3.11}$$

and therefore, the claim follows from (3.8). ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

LEMMA 3.4. *Under Assumption* 1, *let* $\mu = \mu_\beta$ *and let* $\Phi = \Phi_\mu$ *be the solution of the Parisi PDE* (1.2). *Then, for all* $0 \le t \le q_*$, *we have*

$$\mathbb{E}\{\partial_x^2 \Phi(t, X_t)\} = \int_t^1 \mu(s)\, \mathrm{d}s\,. \tag{3.12}$$

Proof. Consider $t \in [0, q_*]$ a continuity point of μ. Then the proof of Lemma 16 in [JT16] yields

$$\partial_x^2 \Phi(t, X_t) = 1 - \mu(t)\big(\partial_x \Phi(t, X_t)\big)^2 - \mathbb{E}\left\{\int_t^1 \big(\partial_x \Phi(s, X_s)\big)^2 \mu(\mathrm{d}s)\right\}. \tag{3.13}$$

Taking expectation and using Fubini's theorem alongside (3.8), we get

$$\mathbb{E}\{\partial_x^2 \Phi(t, X_t)\} = 1 - \mu(t)t - \int_t^1 s\, \mu(\mathrm{d}s) = \int_t^1 \mu(s)\, \mathrm{d}s. \tag{3.14}$$

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 $\partial_x^2 \Phi$ is Lipschitz (cf. Lemma 3.1) 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* $s \ge 0$, *and* $g, v : \mathbb{R} \times \mathbb{R}_{\ge 0} \to \mathbb{R}$ *be defined as per* (3.4), *and set* $q = q_*(\beta)$ *for* $\beta > \beta_0$. *Further, let*

$$\mathcal{E}(\beta) \equiv \frac{\beta}{2}[1 - (1 - q_*(\beta))^2] - \frac{\beta}{2}\int_0^1 s^2\, \mu_\beta(\mathrm{d}s)\,. \tag{3.15}$$

Define the IAMP iteration $(\boldsymbol{z}^k)_{k \ge 0}$ *via* (2.1), (2.4), (2.5), *with* $\widehat{g}_k$ *given by* (2.39), *and let* $\boldsymbol{m}$ *be given by* (2.10). *Then, for any* $\varepsilon > 0$ *there exist* $\delta_*(\varepsilon) > 0$, *such that for any* $\delta \le \delta_*(\varepsilon)$ *we have*

$$\left|\operatorname*{p-lim}_{n\to\infty} \frac{1}{2n}\langle \boldsymbol{m}, \boldsymbol{A}\boldsymbol{m}\rangle - \mathcal{E}(\beta)\right| \le \varepsilon\,, \tag{3.16}$$

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n} d(\boldsymbol{m}, [-1, 1]^n)^2 \le \varepsilon\,. \tag{3.17}$$

(Further the above limits in probability are nonrandom quantities.)

Proof. We claim that the choice (3.4) satisfies the regularity assumptions in Theorem 3. In particular, we have the following:

- $v, g : \mathbb{R} \times [0, 1] \to \mathbb{R}$ are Lipschitz continuous in x, uniformly in t, because $\partial_x \Phi, \partial_x^2 \Phi : [0, 1] \times \mathbb{R} \to \mathbb{R}$ are Lipschitz continuous (see Lemma 3.1), and $0 \le \mu(t) \le 1$ for all $t \in [0, 1]$. Further, $\partial_x \Phi, \partial_x^2 \Phi : [0, 1] \times \mathbb{R} \to \mathbb{R}$ are bounded. Together, these facts imply that g, v are acceptable.
- Condition (a) holds for $t \in [0, q_*(\beta)]$ by (3.10) in Lemma 3.3. Here, we crucially use the no-overlap gap assumption.
- Condition (b) holds because g is itself Lipschitz continuous, and $v(x, t) = v_0(t)v_1(x, t)$, where $v_0(t) = \mu(t)$ has total variation bounded by one, and $v_1(x, t) = \beta^2 \partial_x \Phi(t, x)$ is Lipschitz, again by Lemma 3.1.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

- Condition (c) is satisfied because $\partial_x^3\Phi$ is Lipschitz by Lemma 3.1.

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

$$\mathcal{E}(\beta) = \beta \int_0^{q_*} \int_t^1 \mu(s)\,\mathrm{d}s\,\mathrm{d}t\,. \tag{3.18}$$

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

3.3. Sequential rounding and putting everything together. Theorem 4 constructs a vector $\boldsymbol{m} \in \mathbb{R}^n$. It is not difficult to round this to a vector with entries in $\{+1,-1\}$, as detailed in the next lemma.

LEMMA 3.5. *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 $\boldsymbol{z} \in \mathbb{R}^n$ such that $d(\boldsymbol{z},[-1,1]^n)^2 \le n\,\varepsilon_0$, the algorithm returns a vector $\boldsymbol{\sigma}_* \in \{+1,-1\}^n$ such that*

$$\frac{1}{2n}\langle \boldsymbol{\sigma}_*, \boldsymbol{A}\boldsymbol{\sigma}_*\rangle \ge \frac{1}{2n}\langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{z}\rangle - 20\Big(\sqrt{\varepsilon_0} + \frac{1}{\sqrt{n}}\Big)\,. \tag{3.19}$$

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 $\tilde{H}_n(\boldsymbol{x}) = H_n(\boldsymbol{x}) - \sum_{i=1}^n A_{ii}x_i^2/2 = \sum_{i<j\le n} A_{ij}x_ix_j$.

We construct $\boldsymbol{\sigma}_*$ in two steps. First, we let $\tilde{\boldsymbol{z}}$ be the projection of $\boldsymbol{z}$ onto the hypercube $[-1,+1]^n$ (i.e., $\tilde{\boldsymbol{z}} \in [-1,+1]^n$ is such that $\|\tilde{\boldsymbol{z}} - \boldsymbol{z}\|^2 = d(\boldsymbol{z},[-1,+1]^n)^2 \le n\,\varepsilon_0$). 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 $1 \le \ell \le n$,

$$\tilde{\boldsymbol{z}}(\ell)_i = \begin{cases} \tilde{\boldsymbol{z}}(\ell-1)_i & \text{if } i \neq \ell,\\ \operatorname{sign}\big(h_{1,\ell}(\tilde{\boldsymbol{z}}(\ell-1)_{\sim\ell};\boldsymbol{A})\big) & \text{if } i = \ell. \end{cases} \tag{3.20}$$

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) $\tilde{H}_n(\boldsymbol{\sigma}_*) \ge \tilde{H}_n(\tilde{\boldsymbol{z}})$.
(ii) $|\tilde{H}_n(\boldsymbol{\sigma}_*) - H_n(\boldsymbol{\sigma}_*)| \le 10\sqrt{n}$, $|\tilde{H}_n(\tilde{\boldsymbol{z}}) - H_n(\tilde{\boldsymbol{z}})| \le 10\sqrt{n}$ with probability at least $1 - e^{-2n}$.
(iii) $|H_n(\boldsymbol{z}) - H_n(\tilde{\boldsymbol{z}})| \le 20n\sqrt{\varepsilon_0}$ with probability at least $1 - e^{-2n}$.

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

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

$$|\tilde{H}_n(\boldsymbol{x}) - H_n(\boldsymbol{x})| \le \frac{1}{2}\sum_{i=1}^n |A_{ii}| \equiv \tau(\boldsymbol{A})\,. \tag{3.21}$$

Now we have $\mathbb{E}\tau(\boldsymbol{A}) = \sqrt{2n/\pi}$, and τ is a Lipschitz function of the Gaussian vector $(A_{ii})_{i\le n}$. Hence the desired bounds follow by Gaussian concentration.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

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

$$\big|H_n(\boldsymbol{z}) - H_n(\tilde{\boldsymbol{z}})\big| \le \frac{1}{2}|\langle \boldsymbol{v}, \boldsymbol{A}\boldsymbol{v}\rangle| + |\langle \boldsymbol{v}, \boldsymbol{A}\tilde{\boldsymbol{z}}\rangle| \tag{3.22}$$

$$\le \frac{1}{2}\lambda_{\max}(\boldsymbol{A})\|\boldsymbol{v}\|^2 + \lambda_{\max}(\boldsymbol{A})\|\boldsymbol{v}\|\,\|\tilde{\boldsymbol{z}}\| \tag{3.23}$$

$$\le n\lambda_{\max}(\boldsymbol{A})\Big[\frac{1}{2}\varepsilon_0 + \sqrt{\varepsilon_0}\Big] \le 2n\lambda_{\max}(\boldsymbol{A})\sqrt{\varepsilon_0}\,. \tag{3.24}$$

The desired probability bound follows by concentration of the largest eigenvalue of GOE matrices [AGZ09]. ∎

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

$$\mathcal{E}_0(\beta) \le \operatorname*{p\text{-}lim}_{n\to\infty} \frac{1}{2n} \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle \le \mathcal{E}_0(\beta) + \frac{\log 2}{\beta}\,. \tag{3.25}$$

Proof. First, consider the expectation

$$E_n := \frac{1}{n}\mathbb{E} \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} H_n(\boldsymbol{\sigma})$$

(recall that $H_n(\boldsymbol{\sigma}) = \langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle/2$). Recall the definitions of partition function $Z_n(\beta)$, Gibbs measure ν_β, and free energy density $F_n(\beta)$:

$$Z_n(\beta) = \sum_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \exp(\beta H_n(\boldsymbol{\sigma}))\,, \qquad \nu_\beta(\boldsymbol{\sigma}) := \frac{1}{Z_n(\beta)}\exp(\beta H_n(\boldsymbol{\sigma}))\,, \tag{3.26}$$

$$F_n(\beta) := \frac{1}{n\beta}\mathbb{E}\log Z_n(\beta)\,. \tag{3.27}$$

A standard thermodynamic identity [MM09, Chapter 2] yields $F_n(\beta) = \mathbb{E}\nu_\beta(H_n(\boldsymbol{\sigma})) + \beta^{-1}S(\nu_\beta)$, where $\nu_\beta(H_n(\boldsymbol{\sigma}))$ denotes the average of the Hamiltonian $H_n(\boldsymbol{\sigma})$ with respect to the measure ν_β, and $S(q)$ is the Shannon entropy of the probability distribution q. Further, $F_n'(\beta) = -\beta^{-2}S(\nu_\beta) \le 0$ and $F_n(\beta) \to E_n$ as $\beta\to\infty$. Hence

$$\mathbb{E}\nu_\beta(H_n(\boldsymbol{\sigma})) \overset{\text{(a)}}{\le} E_n \overset{\text{(b)}}{\le} F_n(\beta) \overset{\text{(c)}}{\le} \mathbb{E}\nu_\beta(H_n(\boldsymbol{\sigma})) + \frac{\log 2}{\beta}\,. \tag{3.28}$$

(Here (a) follows since $\nu_\beta(H_n(\boldsymbol{\sigma})) \le \max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} H_n(\boldsymbol{\sigma})$, (b) since $\beta \mapsto F_n(\beta)$ is nonincreasing as shown above, and (c) since $S(q) \le S(q_{\text{unif}}) = \log 2$, with q_{unif} the uniform measure over the hypercube.)

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 [Tal06a, Theorem 1.2]; 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 [Tal06a], with a difference in normalization of β.)

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

This proves

$$\mathcal{E}_0(\beta) \le \liminf_{n\to\infty} E_n \le \limsup_{n\to\infty} E_n \le \mathcal{E}_0(\beta) + \frac{\log 2}{\beta}\,.$$

By Gaussian concentration $\mathbb{P}(|H_n(\boldsymbol{\sigma})/n - E_n| \ge \varepsilon) \le 2\exp(-n\varepsilon^2/C)$ for some constant C, and hence the claim follows. ∎

LEMMA 3.7. *Let* $q_*(\beta) \equiv \sup\{q \in [0,1] : q \in \mathrm{supp}(\mu_\beta)\}$. *Then, for any* $\beta > 0$,

$$\beta^2(1 - q_*(\beta))^2 \le 1\,. \tag{3.29}$$

Proof. The PDE (1.2) can be solved for $t \in (q_*, 1]$ using the Cole–Hopf transformation $\Phi = \log u$. This yields $\Phi(q_*, x) = (\beta^2(1-q_*)/2) + \log 2\cosh x$, whence $\partial_x\Phi(q_*,x) = \tanh(x)$ and $\partial_x^2\Phi(q_*,x) = 1 - \tanh(x)^2$. Substituting in (3.8), (3.9), we get

$$\mathbb{E}\big\{\tanh(X_{q_*})^2\big\} = q_*\,, \tag{3.30}$$

$$\beta^2\mathbb{E}\big\{\big(1-\tanh(X_{q_*})^2\big)^2\big\} \le 1\,. \tag{3.31}$$

Hence

$$\beta^2(1-q_*)^2 = \beta^2\mathbb{E}\big\{1-\tanh(X_{q_*})^2\big\}^2 \le \beta^2\mathbb{E}\big\{\big(1-\tanh(X_{q_*})^2\big)^2\big\} \le 1\,. \tag{3.32}$$ ∎

The proof of 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 1.1, and we further have $E_* \ge 1/2$ (this can be proved by the same thermodynamic argument as in the proof of Lemma 3.6, noting that $(1/n)\log_n Z_n(\beta) \to \log 2 + (\beta^2/4)$ for $\beta \le 1$ [Pan13b]). It is, therefore, sufficient to output $\boldsymbol{\sigma}_*$ such that, with high probability, $H_n(\boldsymbol{\sigma}_*)/n \ge E_* - (\varepsilon/3)$.

Let $\beta = 10/\varepsilon$. By Lemmas 3.6 and 3.7, we have $\mathcal{E}(\beta) \ge E_* - (\varepsilon/5)$. Applying the algorithm of Theorem 4 for δ small enough, we obtain, with high probability, a vector $\boldsymbol{m} \in \mathbb{R}^n$ such that $H_n(\boldsymbol{m})/n \ge E_* - \varepsilon/4$ and $d(\boldsymbol{m}, [-1,1]^n)^2 \le n\varepsilon^2/10^6$. The proof is completed by using the rounding procedure of Lemma 3.5. ∎

4. Relation with the TAP equations. The *TAP equations* were introduced by Thouless, Anderson, and Palmer [TAP77] as a tool to study the Gibbs measure:

$$\nu_{\beta,h}(\boldsymbol{\sigma}) := \frac{1}{Z_n(\beta)}\exp\left\{\beta H_n(\boldsymbol{\sigma}) + h\langle \mathbf{1}, \boldsymbol{\sigma}\rangle\right\}\,. \tag{4.1}$$

(This generalizes (3.26) by the introduction of the linear term $h\langle\mathbf{1},\boldsymbol{\sigma}\rangle$, with $\mathbf{1}$ the all-ones vector.) The TAP equations are a set of n nonlinear equations in the n unknowns $\boldsymbol{m} = (m_1, \dots, m_n)$:

$$h\mathbf{1} + \beta\boldsymbol{A}\boldsymbol{m} - \mathrm{atanh}(\boldsymbol{m}) - \beta^2(1-q_{\boldsymbol{m}})\boldsymbol{m} = \mathbf{err}(n)\,, \quad q_{\boldsymbol{m}} \equiv \frac{1}{n}\|\boldsymbol{m}\|_2^2\,, \tag{4.2}$$

where $\mathbf{err}$ is a small error term. Exact solutions correspond to $\mathbf{err}(n) = 0$ while for approximate solutions $\mathbf{err}(n)$ is small in a suitable sense. In their seminal work, Thouless, Anderson, and Palmer [TAP77] argued heuristically that the mean of the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Gibbs measure $\boldsymbol{m}_{\beta,h} \equiv \sum_{\boldsymbol{\sigma}\in\{+1,-1\}} \nu_{\beta,h}(\boldsymbol{\sigma})\boldsymbol{\sigma}$ approximately solves (4.2). Subsequent physics research clarified that this claim only holds at high temperature, namely if $\beta \le \beta_{\rm AT}(h)$, where the critical value $\beta_{\rm AT}(h)$ is known as the Almeida–Thouless line [MPV87]. However, the TAP equations are believed to hold for $\beta > \beta_{\rm AT}(h)$ as well if the mean $\boldsymbol{m}_{\beta,h}$ is replaced by the mean over a "pure state" $\boldsymbol{m}^{\alpha}_{\beta,h} \equiv \sum_{\boldsymbol{\sigma}\in\mathcal{S}_\alpha} \nu_{\beta,h}(\boldsymbol{\sigma})\boldsymbol{\sigma}$. We refer to [CPS18, CPS19] for recent mathematical progress on this topic.

Here we will not be concerned with the physical interpretation of TAP equations, but with the computational problem of finding approximate solutions. Bolthausen [Bol14] gave an iterative algorithm that converges to approximate solutions for $\beta \le \beta_{\rm AT}(h)$. This is an algorithm of AMP type, and essentially amounts to iterating the TAP equations themselves. The algorithm does not converge for $\beta > \beta_{\rm AT}(h)$.

In this section we prove that the algorithm described in section 2, when used in conjunction with the specific choice of functions g_k, s, v in section 3, actually constructs an approximate solution of the TAP equations for for $\beta > \beta_0$, under Assumption 1. For coherence with the rest of the paper, we focus on the case $h = 0$: the case $h \neq 0$ can be treated by a generalization of the present approach that is deferred to future work. Notice that Assumption 1 is believed to hold for all $\beta > \beta_{\rm AT}(0) = 1$, hence covering the full range of parameters for which the approach of [Bol14] fails.

As in the previous section, we set $q = q_*$, $v(x,t) = \beta^2\mu(t)\partial_x\Phi(t,x)$, $s(x,t) = \beta$, $g(x,t) = \beta\partial_x^2\Phi(t,x)$, and

$$\widehat{g}_k(x) \equiv \frac{g(x,k\delta)}{\mathbb{E}\{g(X^\delta_{k-1},k\delta)^2\}^{1/2}}\,. \tag{4.3}$$

Throughout, we set $K_* = \lfloor q_*/\delta\rfloor$ and recall that $\boldsymbol{z}^k$, $\boldsymbol{x}^k$, and $\boldsymbol{m}$ are given by

$$\boldsymbol{z}^{k+1} = \boldsymbol{A}(\boldsymbol{m}^k + \boldsymbol{y}) - \sum_{j=1}^{k} \mathsf{b}_{kj}\boldsymbol{m}^{j-1}\,, \tag{4.4}$$

$$\boldsymbol{x}^j = \boldsymbol{x}^{j-1} + v(\boldsymbol{x}^{j-1}, j\delta)\,\delta + \beta(\boldsymbol{z}^j - \boldsymbol{z}^{j-1})\,, \tag{4.5}$$

$$\boldsymbol{m}^k = \sum_{j=1}^{k} \widehat{g}_j(\boldsymbol{x}^{j-1}) \odot (\boldsymbol{z}^j - \boldsymbol{z}^{j-1}) := \sum_{j=1}^{k} \boldsymbol{g}^j\,, \tag{4.6}$$

with $\boldsymbol{m} = \boldsymbol{m}^{K_*}$. Finally, we will repeatedly use the fact that the PDE (1.2) can be solved on $(q_*, 1]$ using the Cole–Hopf transformation, which yields $\Phi(q_*, x) = \log 2\cosh(x) + \beta^2(1-q_*)/2$.

LEMMA 4.1. *Setting $K_* = \lfloor q_*/\delta\rfloor$, we have*

$$\lim_{\delta\to 0} \text{p-lim}_{n\to\infty} \frac{1}{n}\left\|\boldsymbol{m} - \tanh(\boldsymbol{x}^{K_*})\right\|^2 = 0\,. \tag{4.7}$$

Proof. By Lemma 2.2, we have

$$\text{p-lim}_{n\to\infty} \frac{1}{n}\left\|\boldsymbol{m} - \tanh(\boldsymbol{x}^{K_*})\right\|^2 = \mathbb{E}\left\{\left[M^\delta - \partial_x\Phi(q_*, X^\delta_{K_*})\right]^2\right\}\,. \tag{4.8}$$

On the other hand, using Lemma 2.8, we obtain

$$\lim_{\delta\to 0} \mathbb{E}\left\{\left[M^\delta - \partial_x\Phi(q_*, X^\delta_{K_*})\right]^2\right\} = \mathbb{E}\left\{\left[M_{q_*} - \partial_x\Phi(q_*, X_{q_*})\right]^2\right\} = 0\,, \tag{4.9}$$

where the last identity follows from Lemma 3.2. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

LEMMA 4.2. *Setting $K_* = \lfloor q_*/\delta \rfloor$, we have*

$$\lim_{\delta\to 0} \text{p-lim}_{n\to\infty} \frac{1}{n} \left\| \beta \boldsymbol{A}\boldsymbol{m} - \boldsymbol{x}^{K_*} - \beta^2(1-q_*)\tanh(\boldsymbol{x}^{K_*}) \right\|^2 = 0\,. \tag{4.10}$$

Proof. By (4.4), we have

$$\begin{aligned} \boldsymbol{A}\boldsymbol{m} &= \boldsymbol{z}^{K_*+1} - \boldsymbol{z}^1 + \sum_{j=1}^{K_*} \mathsf{b}_{K_*,j} \boldsymbol{m}^{j-1} \\ &= \boldsymbol{z}^{K_*+1} - \boldsymbol{z}^1 + \sum_{j=1}^{K_*} \tilde{\mathsf{b}}_{K_*,j} \boldsymbol{g}^{j-1} + \mathsf{err}\,, \end{aligned}$$

where, by (2.17), and using the fact that $\|\boldsymbol{m}^j\|^2/n$ is bounded by (2.13), we have $\|\mathsf{err}\|^2/n \xrightarrow{p} 0$. Using (2.17), together with the fact that $\|\boldsymbol{f}_k\|^2/n$, $\|\boldsymbol{u}^k\|^2/n$ are bounded by Lemma 2.2, we get

$$\begin{aligned} &\text{p-lim}_{n\to\infty} \frac{1}{n} \left\| \beta \boldsymbol{A}\boldsymbol{m} - \boldsymbol{x}^{K_*} - \beta^2(1-q_*)\tanh(\boldsymbol{x}^{K_*}) \right\|^2 \\ &= \text{p-lim}_{n\to\infty} \frac{1}{n} \left\| \beta \boldsymbol{z}^{K_*+1} - \beta \boldsymbol{z}^1 + \beta \sum_{j=1}^{K_*} \tilde{\mathsf{b}}_{K_*,j} \boldsymbol{g}^{j-1} - \boldsymbol{x}^{K_*} - \beta^2(1-q_*)\tanh(\boldsymbol{x}^{K_*}) \right\|^2 \end{aligned} \tag{4.11}$$

$$\begin{aligned} &= \mathbb{E}\Bigg\{ \Bigg[\beta Z^\delta_{K_*+1} - \beta Z^\delta_1 + \beta \sum_{j=1}^{K_*} \mathbb{E}\{\widehat{g}_j(X^\delta_{j-1})\} \widehat{g}_{j-1}(X^\delta_{j-2}) U^\delta_{j-1} \\ &\qquad - X^\delta_{K_*} - \beta^2(1-q_*)\tanh(X^\delta_{K_*}) \Bigg]^2 \Bigg\}\,. \end{aligned} \tag{4.12}$$

Next, using again Lemma 2.8, we have $\sum_{k=1}^{K_*} U^\delta_{k+1} \xrightarrow{L_2} B_{q_*}$, $X^\delta_{K_*} \xrightarrow{L_2} X_{q_*}$ as $\delta \to 0$, and

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

where in the last step we used Lemma 3.4. By Fubini's theorem

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where in the last step we used (3.6). Substituting these limits in (17), we get

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

where we used the fact that X_t solves the SDE (3.3), and $\Phi(q_*,x)=\log 2\cosh(x)+\beta^2(1-q_*)/2$. ∎

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 $s\ge 0$ and $g,v:\mathbb{R}\times\mathbb{R}_{\ge 0}\to\mathbb{R}$ be defined as per (3.4), and set $q=q_*(\beta)$ for $\beta>\beta_0$. Define the incremental AMP iteration $(\boldsymbol{z}^k)_{k\ge 0}$ via (2.1), (2.4), (2.5), with $\widehat{g}_k$ given by (2.39), and let $\boldsymbol{m}$ be given by (2.10). (The same iteration is given explicitly in (4.5), (4.6).)*

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

$$\frac{1}{n}\left\|\beta\boldsymbol{A}\tanh(\boldsymbol{x}^{K_*})-\boldsymbol{x}^{K_*}-\beta^2(1-q_*)\tanh(\boldsymbol{x}^{K_*})\right\|\le\varepsilon\,. \tag{4.13}$$

Proof. The theorem follows immediately from Lemmas 4.1 and 4.2, using the fact that, with high probability, $\boldsymbol{A}$ has an operator norm bounded by $2+\varepsilon$ [AGZ09]. ∎

5. Universality. In this section we use the universality results of [BLM15] 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 [BLM15, Theorem 4], and applying the same reduction as in the first part of the proof of Proposition 2.1, we obtain the following.

PROPOSITION 5.1. *Consider the AMP iteration (2.1), 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 $k,\ell\in\mathbb{N}$, and any pseudo-Lipschitz function of order ℓ, $\psi:\mathbb{R}^{k+2}\to\mathbb{R}$, we have*

$$\frac{1}{n}\sum_{i=1}^n\psi(z_i^0,\dots,z_i^k;y_i)\xrightarrow{p}\mathbb{E}\psi(Z_0,\dots,Z_k;Y)\,. \tag{5.1}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Here $(Z_j)_{j\ge 1}$ is a centered Gaussian process independent of (Z_0, Y) with covariance $\boldsymbol{Q} = (Q_{kj})_{k,j\ge 1}$ determined recursively via

$$Q_{k+1,j+1} = \mathbb{E}\big\{ f_k(Z_0,\dots,Z_k;Y) f_j(Z_0,\dots,Z_j;Y)\big\}\,, \qquad k,j\ge 0\,. \tag{5.2}$$

Notice an important difference with respect to Proposition 2.1: instead of Lipschitz functions, we require the functions f_k to be polynomials. However, this result is strong enough to allow us to 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 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 (2.4), (2.5),

$$f_k(z_0,\dots,z_k;y) := \sum_{\ell=1}^{k} \hat{p}_{\ell,M}\left(z_1,\dots,z_{\ell-1}\right)\cdot(z_\ell - z_{\ell-1}) + y\,, \tag{5.3}$$

and set $p_{Z_0,Y} = \delta_0\otimes \mathsf{N}(0,\delta)$. 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

$$\lim_{M\to\infty}\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi(z_i^0,\dots,z_i^k) = \mathbb{E}\psi(Z_0^\delta,\dots,Z_k^\delta)\,, \tag{5.4}$$

where the Gaussian martingale $(Z_\ell^\delta)_{\ell\ge 0}$ is defined as in Lemma 2.2. Given this claim, the rest of the proof of Theorem 2 can be applied verbatim to this—slightly different—algorithm.

In order to prove the claim (5.4), we proceed as in the proof of Lemma 2.2. Namely, by applying Proposition 5.1, we get

$$\operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi(z_i^0,\dots,z_i^k) = \mathbb{E}\psi(Z_0,\dots,Z_k^{\delta,M})\,, \tag{5.5}$$

where $(Z_\ell^{\delta,M})_{\ell\ge 0}$ is a centered Gaussian process. Using the same argument as in Lemma 2.2, we obtain that $(Z_\ell^{\delta,M})_{\ell\ge 0}$ is a martingale. Further, letting $q_\ell^M \equiv \mathbb{E}\{(Z_\ell^{\delta,M})^2\}$, Proposition 5.1 yields the following recursion for $k\ge 1$ (with initial condition $q_0^M = 0$, $q_1^M = \delta$):

$$q_{k+1}^M = q_k^M + \mathbb{E}\{\hat{p}_{k,M}(Z_1^{\delta,M},\dots,Z_{k-1}^{\delta,M})^2\}\cdot(q_k^M - q_{k-1}^M)\,. \tag{5.6}$$

The claim (5.4) follows by showing that we can choose polynomials $(\hat{p}_{k,M})_{k\ge 0}$ so that $\lim_{M\to\infty} q_k^M = q_k$ for each $k\le 1/\delta$. This can be done by induction over k. As a preliminary remark, notice that the sequence of constants q_k defined recursively via

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

(2.6) are all finite. Indeed if $q_k < \infty$, then $\mathbb{E}\{\widehat{g}_k(X_{k-1}^\delta)^2\} < \infty$ (because X_{k-1}^δ is a Lipschitz function of $(Z_\ell^\delta)_{\ell \le k-1}$) and, therefore, $q_{k+1} < \infty$ as well. We denote by $C_0 := q_{\lfloor 1/\delta \rfloor} = \max(q_k : k \le 1/\delta)$ an upper bound on these variances.

The basis of the induction is trivial since $q_0^M = 0 = q_0$ for all M. We assume, therefore, $\lim_{M\to\infty} q_\ell^M = q_\ell$ for all $\ell \le k$. Without loss of generality we can consider that for any $M \ge 1$ we have $q_1^M, \dots, q_k^M \le 2C_0$. Indeed, by the induction hypothesis this holds for all M large enough, and we can always redefine 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 (2.6) can be written as $X_k^\delta = h_k(Z_1^\delta, \dots, Z_{k-1}^\delta)$ for a certain function h_k that is bounded by a polynomial (of degree depending on k). We then choose the polynomial $\hat{p}_{k,M}(\,\cdot\,)$ so that

$$\mathbb{E}\Big\{\big|\widehat{g}_k(h_k(Z_1^{\delta,M}, \dots, Z_{k-1}^{\delta,M})) - \hat{p}_{k,M}(Z_1^{\delta,M}, \dots, Z_{k-1}^{\delta,M})\big|^2\Big\} \le \frac{1}{M}\,. \tag{5.7}$$

Such a polynomial 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 in $L^2(\mathbb{R}^{k-1}, \mu_k)$, where μ_k is the joint distribution of $Z_1^{\delta,M}, \dots, Z_{k-1}^{\delta,M}$). The variance bound $q_1^M, \dots, q_k^M \le 2C_0$ is used in controlling the error term.

The induction claim then follows by

$$\begin{aligned}\lim_{M\to\infty} \mathbb{E}\{\hat{p}_{k,M}(Z_1^{\delta,M}, \dots, Z_{k-1}^{\delta,M})^2\} &= \lim_{M\to\infty} \mathbb{E}\{\widehat{g}_k(h_k(Z_1^{\delta,M}, \dots, Z_{k-1}^{\delta,M}))^2\} \\ &= \mathbb{E}\{\widehat{g}_k(h_k(Z_1^{\delta}, \dots, Z_{k-1}^{\delta}))^2\}\,,\end{aligned} \tag{5.8}$$

where the last equality holds by dominated convergence. ∎

Corollary 1.2 follows by applying Theorem 6 with $\boldsymbol{A}$, a suitably centered and normalized adjacency matrix.

Proof of Corollary 1.2. 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 $i \ne j$,

$$A_{ij} = \begin{cases} -\sqrt{\frac{1-p}{np}} & \text{if } (i,j) \in E, \\ \sqrt{\frac{p}{n(1-p)}} & \text{if } (i,j) \notin E, \end{cases} \tag{5.9}$$

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

$$\mathsf{CUT}_G(\boldsymbol{\sigma}) = \frac{1}{2}|E_n| - \frac{p}{4}\langle \boldsymbol{\sigma}, \mathbf{1}\rangle^2 + \frac{1}{4}\sqrt{np(1-p)}\,\langle \boldsymbol{\sigma}, \boldsymbol{A}\boldsymbol{\sigma}\rangle\,. \tag{5.10}$$

Recall that we know from [DMS17] $\max_{\boldsymbol{\sigma}\in\{+1,-1\}^n} \mathsf{CUT}_G(\boldsymbol{\sigma}) = |E_n|/2 + (n^3p(1-p)/2)^{1/2}\mathsf{P}_* + o(n^{3/2})$. Let $\boldsymbol{\sigma}_1$ denote the output of the algorithm of Theorem 6, on input $\boldsymbol{A}$. Applying this theorem and Proposition 5.1, we get

$$\text{p-lim}\inf_{n\to\infty} \frac{1}{2n}\langle \boldsymbol{\sigma}_1, \boldsymbol{A}\boldsymbol{\sigma}_1\rangle \ge (1-\varepsilon)\mathsf{P}_*\,, \tag{5.11}$$

$$\text{p-lim}_{n\to\infty} \frac{1}{n}\langle \boldsymbol{\sigma}_1, \mathbf{1}\rangle = 0\,. \tag{5.12}$$

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$\lfloor \ell/2 \rfloor$ entries of $\boldsymbol{\sigma}_1$ so that $|\langle \boldsymbol{\sigma}_*, \mathbf{1}\rangle| \le 1$. We then have, with high probability,

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

(Here $\|\boldsymbol{A}\|_{\mathrm{op}}$ denotes the operator norm of matrix $\boldsymbol{A}$.) Therefore, since $|\langle \boldsymbol{\sigma}_1, \mathbf{1}\rangle|/n = \ell/n \xrightarrow{p} 0$, and $\|\boldsymbol{A}\|_{\mathrm{op}} \le 2.01$ with high probability [AGZ09], we get

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

which completes the proof. ∎

Appendix A. Proof of Proposition 2.1. As mentioned in the main text, Proposition 2.1 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 [JM13, Theorem 1]. Let us briefly recall the class of algorithms considered in [JM13], adapting the notations to the present ones. (We limit ourselves to considering the "one-block" case in the language of [JM13]).

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

$$\begin{aligned}
F_t\ &: \mathbb{R}^T \times \mathbb{R}^2 \to \mathbb{R}^T\,, \\
&(x_1, \dots, x_T, w_1, w_2) \mapsto F_t(x_1, \dots, x_T, w_1, w_2)\,.
\end{aligned}$$

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

Then [JM13] analyzes the following AMP iteration, which produces a sequence of iterates $\boldsymbol{x}^t \in \mathbb{R}^{n\times T}$ for $t \ge 0$ (whereby $F_{-1}(\cdots) \equiv 0$ by definition):

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

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

$$(\mathsf{B}_t)_{ij} = \frac{1}{n}\sum_{\ell=1}^n (D_{\boldsymbol{x}}F_t(\boldsymbol{x}_\ell^t; \boldsymbol{w}_\ell))_{ij} = \frac{1}{n}\sum_{\ell=1}^n \frac{\partial F_{t,i}}{\partial x_{\ell,j}^t}(\boldsymbol{x}_\ell^t; \boldsymbol{w}_\ell)\,. \tag{A.2}$$

Under the assumption that $\boldsymbol{x}^0, \boldsymbol{w}$ are independent of $\boldsymbol{A}$, and $\hat{p}_{\boldsymbol{x}^0,\boldsymbol{w}} \equiv n^{-1}\sum_{i=1}^n \delta_{\boldsymbol{x}_i^0,\boldsymbol{w}_i}$ which converges in W_ℓ, [JM13, Theorem 1] determines the W_ℓ asymptotics of the empirical distribution of $\boldsymbol{x}^t, \boldsymbol{w}$.

First, consider the case in which the functions $f_j : \mathbb{R}^{j+1} \to \mathbb{R}$ are Lipschitz continuous. Proposition 2.1 can be recast as a special case of [JM13, Theorem 1]. First, notice that we can always choose an n-independent T such that the time horizon k in (2.3) satisfies $k \le T$. We then consider the iteration (A.1) with initialization $\boldsymbol{x}^0 = \mathbf{0}$, data vectors $\boldsymbol{w} = (\boldsymbol{z}^0, \boldsymbol{y})$, and update functions given by

$$F_t(x_1, x_2, \dots, x_T, w_1, w_2)_t = f_t(w_1, x_1, \dots, x_t; w_2)\,. \tag{A.3}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

With this setting, the vector $(x^t_{i,j})_{i\le n} \in \mathbb{R}^n$ coincides with $\boldsymbol{z}^t$ as given in (2.1) for all $j \le t \le T$. The recursion of (2.2) follows from the analogous recursion in [JM13, Theorem 1].

Next, we have to consider the case of functions $f_j : \mathbb{R}^{j+1} \to \mathbb{R}$ that are not necessarily Lipschitz, but pseudo-Lipschitz of order $m \ge 1$. For a large constant $M \ge 1$, and for each $j \in \mathbb{N}$, we let $f^M_j : \mathbb{R}^{j+1} \to \mathbb{R}$ be a Lipschitz function such that $f^M_j(\boldsymbol{s}) = f_j(\boldsymbol{s})$ for $\|\boldsymbol{s}\| \le M$. Such a function exists by Kirszbraun's extension theorem and has the same Lipschitz constant as f_j on $\{\boldsymbol{s} \in \mathbb{R}^{j+1} : \|\boldsymbol{s}\| \le M\}$, $\mathrm{Lip}(f^M_j) \equiv L_M \le CM^m$. Notice further that

$$\big|f_j(\boldsymbol{s}) - f^M_j(\boldsymbol{s})\big| \le C(1+\|\boldsymbol{s}\|)^m \mathbf{1}_{\|\boldsymbol{s}\|\ge M} \le \frac{C}{M}(1+\|\boldsymbol{s}\|)^{m+1}\,. \tag{A.4}$$

Denote by $(\boldsymbol{z}^{M,j})_{j\ge 0}$ the AMP iterates obtained replacing f_j by f^M_j for each j:

$$\begin{aligned} \boldsymbol{z}^{M,k+1} &= \boldsymbol{A}\, f^M_k(\boldsymbol{z}^{M,0},\dots,\boldsymbol{z}^{M,k};\boldsymbol{y}) - \sum_{j=1}^{k} \mathsf{b}^M_{k,j} f^M_{j-1}(\boldsymbol{z}^{M,0},\dots,\boldsymbol{z}^{M,j-1};\boldsymbol{y})\,,\\ \mathsf{b}^M_{k,j} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial f^M_k}{\partial z^j_i}(z^{M,0}_i,\dots,z^{M,k}_i;y_i)\,. \end{aligned} \tag{A.5}$$

We denote by $\boldsymbol{Q} = (Q_{ij})_{i,j\ge 1}$ the state evolution covariance associated to functions $(f_j)_{j\ge 0}$ and by $\boldsymbol{Q}^M = (Q^M_{ij})_{i,j\ge 1}$ the covariance associated to functions $(f^M_j)_{j\ge 0}$. Using (A.4) in the state evolution recursion, we get, by an induction argument,

$$\lim_{M\to\infty} Q^M_{ij} = Q_{ij} \qquad \forall i,j \ge 1\,. \tag{A.6}$$

In what follows, given a sequence of random variables X_n and a constant C, we write $\text{p-limsup}_{n\to\infty} X_n < C$ if

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \ge C\right) = 0\,.$$

If $X_{n,M}$ also depends on the parameter M, we write $\lim_{M\to\infty} \text{p-limsup}_{n\to\infty} X_{n,M} \le C$ if there exists a sequence of nonrandom constants C_M such that $\text{p-limsup}_{n\to\infty} X_n < C_M$ and $\lim_{M\to\infty} C_M = C$. Finally, we write $X_n = O(1)$ if $\text{p-limsup}_{n\to\infty} X_n < C$ $\text{p-limsup}_{n\to\infty}(-X_n) < C$ for some constant C: let us emphasize that this is different from "big Oh in probability." Namely, if $X_n = O(1)$, then there exists a finite constant C such that $\mathbb{P}(|X_n| > C) \to 0$.

We will prove by induction that the following claims—denoted by $\mathcal{A}(k) = (\mathcal{A}_1(k), \mathcal{A}_2(k), \mathcal{A}_3(k))$—hold for all k:

$\mathcal{A}_1(k)$. The empirical distributions of $\boldsymbol{z}^{M,k}, \boldsymbol{z}^k$ have bounded moments of all orders, uniformly in M, n, in probability. Namely, for each $k, \ell \in \mathbb{Z}$, there exist constants $C_{k,\ell} < \infty$ such that

$$\text{p-limsup}_{n\to\infty} \frac{1}{n}\|\boldsymbol{z}^k\|^\ell_\ell < C_{k,\ell}\,, \qquad \text{p-limsup}_{n\to\infty} \frac{1}{n}\|\boldsymbol{z}^{M,k}_i\|^\ell_\ell < C_{k,\ell}\,. \tag{A.7}$$

$\mathcal{A}_2(k)$. The iterates $\boldsymbol{z}^{M,k}$ approximate well $\boldsymbol{z}^k$ in ℓ_2. Namely, for each $k \ge 0$, we have

$$\lim_{M\to\infty} \text{p-limsup}_{n\to\infty} \frac{1}{n}\|\boldsymbol{z}^k - \boldsymbol{z}^{M,k}\|^2 = 0\,. \tag{A.8}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

$\mathcal{A}_3(k)$. For any ℓ and any function $\psi : \mathbb{R}^{k+2} \to \mathbb{R}$, $\psi \in \mathrm{PL}(\ell)$,

(A.9)
$$\lim_{M\to\infty} \text{p-limsup}_{n\to\infty} \left| \frac{1}{n}\sum_{i=1}^{n} \psi(z_i^0, \dots, z_i^k; y_i) - \frac{1}{n}\sum_{i=1}^{n} \psi(z_i^{M,0}, \dots, z_i^{M,k}; y_i) \right| = 0\,.$$

Together with (A.6), and with the results above (i.e., the claim (2.3) for the Lipschitz functions f_j^M), this implies the desired claim

$$\text{p-lim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(z_i^0, \dots, z_i^k; y_i) = \mathbb{E}\psi(Z_0, \dots, Z_k; Y)\,. \tag{A.10}$$

Recall that, by assumption, $\boldsymbol{Q}_{\le k} := (Q_{ij})_{1\le i,j\le k}$ is strictly positive definite for each k. Equivalently, there is no k, and there exist nonvanishing coefficients $(c_j)_{j\le k}$ such that $\sum_{j=0}^{k} c_k f_j(Z_0, \dots, Z_j; Y) = 0$ almost surely.

The base case $k = 0$ trivially holds. We next assume that $\mathcal{A}(k)$ holds and prove $\mathcal{A}(k+1)$.

Proof of $\mathcal{A}_1(k+1)$. The claim for $\boldsymbol{z}^{M,k}$ holds because the functions f_j^M are Lipschitz, and hence we can use the result above (namely, (2.3) for Lipschitz functions) to get

$$\frac{1}{n}\|\boldsymbol{z}^{M,k+1}\|_\ell^\ell \xrightarrow{p} \mathbb{E}\{(Z_{k+1}^M)^\ell\}\,, \tag{A.11}$$

where $Z_{k+1}^M \sim \mathsf{N}(0, Q_{k+1,k+1}^M)$. Using (A.6), the claim follows.

In order to prove the claim for $\boldsymbol{z}^k$, we will use the shorthand $\boldsymbol{f}^j = f_j(\boldsymbol{z}^0, \dots, \boldsymbol{z}^j; \boldsymbol{y})$, and can therefore rewrite (2.1) as

$$\boldsymbol{z}^{k+1} = \boldsymbol{A}\,\boldsymbol{f}^k - \boldsymbol{r}^k\,, \quad \boldsymbol{r}^k := \sum_{j=1}^{k} \mathsf{b}_{k,j}\boldsymbol{f}^{j-1}\,. \tag{A.12}$$

Let $\boldsymbol{F}_j \in \mathbb{R}^{n\times(j+1)}$ be the matrix with columns $\boldsymbol{f}^0, \dots, \boldsymbol{f}^j$, $\boldsymbol{P}_j$ the projector onto the column space of $\boldsymbol{F}_j$, and $\boldsymbol{P}_j^\perp \equiv \boldsymbol{I}_n - \boldsymbol{P}_j$. By the induction hypothesis $\mathcal{A}_3(k)$, we have

$$\text{p-lim}_{n\to\infty} \frac{1}{n}\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{F}_{k-1} = \boldsymbol{Q}_{\le k-1} \succeq \varepsilon \boldsymbol{I}_k\,, \tag{A.13}$$

where the last inequality follows from the assumption about the functions f_j not being linearly degenerate (we write $\boldsymbol{A} \succeq \boldsymbol{B}$ if the matrix $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite). This, in particular, implies that we can write $\boldsymbol{P}_{k-1} = \boldsymbol{F}_{k-1}(\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{F}_{k-1})^{-1}\boldsymbol{F}_{k-1}^{\mathsf{T}}$.

Following earlier work [BM11, JM13, BMN19], we then rewrite (A.12) as

$$\begin{aligned}
\boldsymbol{z}^{k+1} &= \boldsymbol{P}_{k-1}^\perp \boldsymbol{A}\boldsymbol{P}_{k-1}^\perp \boldsymbol{f}^k + \boldsymbol{P}_{k-1}^\perp \boldsymbol{A}\boldsymbol{P}_{k-1}\boldsymbol{f}^k + \boldsymbol{P}_{k-1}\boldsymbol{A}\boldsymbol{f}^k - \boldsymbol{r}^k \\
&\overset{\mathrm{d}}{=} \boldsymbol{P}_{k-1}^\perp \tilde{\boldsymbol{A}}\boldsymbol{P}_{k-1}^\perp \boldsymbol{f}^k + \boldsymbol{P}_{k-1}^\perp \boldsymbol{A}\boldsymbol{P}_{k-1}\boldsymbol{f}^k + \boldsymbol{P}_{k-1}\boldsymbol{A}\boldsymbol{f}^k - \boldsymbol{r}^k \\
&= \tilde{\boldsymbol{A}}\boldsymbol{P}_{k-1}^\perp \boldsymbol{f}^k - \boldsymbol{P}_{k-1}\tilde{\boldsymbol{A}}\boldsymbol{P}_{k-1}^\perp \boldsymbol{f}^k - \boldsymbol{r}^k + \boldsymbol{P}_{k-1}^\perp \boldsymbol{A}\boldsymbol{P}_{k-1}\boldsymbol{f}^k + \boldsymbol{P}_{k-1}\boldsymbol{A}\boldsymbol{f}^k \\
&:= \boldsymbol{v}_1 - \boldsymbol{v}_2 - \boldsymbol{v}_3 + \boldsymbol{v}_4 + \boldsymbol{v}_5\,,
\end{aligned}$$

where $\tilde{\boldsymbol{A}}$ is distributed as $\boldsymbol{A}$ but independent of the σ-algebra $\mathcal{F}_k \equiv \sigma(\{\boldsymbol{f}^j, \boldsymbol{z}^j\}_{j\le k})$. We next bound the ℓth norm of each of the terms in the last expression:

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

1. Consider $\boldsymbol{v}_1$. Since $\tilde{\boldsymbol{A}}$ is independent of $\boldsymbol{u} := \boldsymbol{P}_{k-1}^{\perp}\boldsymbol{f}^k$, we have $\boldsymbol{v}_1 \stackrel{\mathrm{d}}{=} (\|\boldsymbol{u}\|_2/\sqrt{n})\boldsymbol{g} + g_0(\boldsymbol{u}/\sqrt{n})$, where $(g_0, \boldsymbol{g}) \sim \mathsf{N}(0, \boldsymbol{I}_{n+1})$. Also, notice that
$$\frac{1}{\sqrt{n}}\|\boldsymbol{u}\|_2 \le \frac{1}{\sqrt{n}}\|\boldsymbol{f}^k\|_2 = O(1)\,,$$
where in the last step we used the induction hypothesis $\mathcal{A}_3(k)$. Further, notice that
$$\boldsymbol{P}_{k-1}\boldsymbol{f}^k = \sum_{j=0}^{k-1} a_{kj}\boldsymbol{f}^j\,, \qquad a_{kj} := \sum_{l=0}^{k-1}(\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{F}_{k-1})^{-1}_{jl}\langle \boldsymbol{f}^l, \boldsymbol{f}^k\rangle\,.$$
Using (A.13) and the induction hypothesis $\mathcal{A}_3(k)$, we obtain that $|a_{kj}| = O(1)$. Therefore, since $\boldsymbol{u} := \boldsymbol{f}^k - \boldsymbol{P}_{k-1}\boldsymbol{f}^k$, and using again the induction hypothesis $\mathcal{A}_3(k)$,
$$\begin{aligned}\frac{1}{n^{1/\ell}}\|\boldsymbol{u}\|_\ell &\le \frac{1}{n^{1/\ell}}\|\boldsymbol{f}^k\|_\ell + \frac{1}{n^{1/\ell}}\|\boldsymbol{P}_{k-1}\boldsymbol{f}^k\|_\ell\\ &\le O(1) + \sum_{j=0}^{k-1}\operatorname*{p-limsup}_{n\to\infty}\left(|a_{kj}|\,\frac{1}{n^{1/\ell}}\|\boldsymbol{f}^j\|_\ell\right) = O(1)\,.\end{aligned}$$
We conclude that
$$\begin{aligned}\operatorname*{p-limsup}_{n\to\infty}\frac{1}{n^{1/\ell}}\|\boldsymbol{v}_1\|_\ell &\le \operatorname*{p-limsup}_{n\to\infty}\frac{1}{\sqrt{n}}\|\boldsymbol{u}\|_2\cdot\frac{1}{n^{1/\ell}}\|\boldsymbol{g}\|_\ell\\ &\quad + \operatorname*{p-limsup}_{n\to\infty}\frac{|g_0|}{n^{1/2-1/\ell}}\cdot\frac{1}{n^{1/\ell}}\|\boldsymbol{u}\|_\ell\\ &\le C\operatorname*{p-limsup}_{n\to\infty}\frac{1}{n^{1/\ell}}\|\boldsymbol{g}\|_\ell + C\operatorname*{p-limsup}_{n\to\infty}\frac{|g_0|}{n^{1/2-1/\ell}} \le C\,,\end{aligned}$$
where the last inequality follows from the law of large numbers.
2. Term $\boldsymbol{v}_2$ is bounded by a similar argument by noting that (for g_0, $\boldsymbol{g}$ as above)
$$\begin{aligned}\boldsymbol{v}_2 &\stackrel{\mathrm{d}}{=} \boldsymbol{P}_{k-1}\left\{\frac{\|\boldsymbol{P}_{k-1}^{\perp}\boldsymbol{f}^k\|_2}{\sqrt{n}}\,\boldsymbol{g} + \frac{g_0}{\sqrt{n}}\boldsymbol{P}_{k-1}^{\perp}\boldsymbol{f}^k\right\}\\ &= \frac{\|\boldsymbol{P}_{k-1}^{\perp}\boldsymbol{f}^k\|_2}{\sqrt{n}}\,\boldsymbol{P}_{k-1}\boldsymbol{g}\,.\end{aligned}$$
As shown above, $\|\boldsymbol{P}_{k-1}^{\perp}\boldsymbol{f}^k\|_2/\sqrt{n} = O(1)$. Further,
$$\begin{aligned}&\boldsymbol{P}_{k-1}\boldsymbol{g} = \frac{1}{\sqrt{n}}\sum_{j=0}^{k-1}\tilde{g}_j\boldsymbol{f}^j\,,\\ &(\tilde{g}_0,\dots,\tilde{g}_{k-1})\big|_{\boldsymbol{F}_{k-1}} \sim \mathsf{N}(\boldsymbol{0}, (\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{F}_{k-1}/n)^{-1})\,.\end{aligned}$$
By the induction hypothesis $\mathcal{A}_3(k)$, we have $\|\boldsymbol{f}^j\|_\ell/n^{1/\ell} = O(1)$, and therefore,
$$\begin{aligned}\operatorname*{p-limsup}_{n\to\infty}\frac{1}{n^{1/\ell}}\,\|\boldsymbol{P}_{k-1}\boldsymbol{g}\|_\ell &\le C\operatorname*{p-limsup}_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{j=0}^{k-1}|\tilde{g}_j|\\ &\le C\operatorname*{p-limsup}_{n\to\infty}\frac{1}{\sqrt{n}}\|\tilde{\boldsymbol{g}}\|_2 = 0\,,\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

where the last identity follows by applying the Markov inequality conditionally on $\boldsymbol{F}_{k-1}$, since $\mathbb{E}\{\|\tilde{\boldsymbol{g}}\|_2^2|\boldsymbol{F}_{k-1}\} = \mathsf{Tr}((\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{F}_{k-1}/n)^{-1}) \xrightarrow{p} \mathsf{Tr}(\boldsymbol{Q}_{\le k}^{-1})$, which is bounded by (A.13).

3. Next, consider $\boldsymbol{v}_3 = \boldsymbol{r}^k = \sum_{j=1}^k \mathsf{b}_{k,j}\boldsymbol{f}^j$. Notice that

$$\mathsf{b}_{k,j} := \frac{1}{n}\sum_{i=1}^n \psi_{k,j}(z_i^0,\dots,z_i^k;y_i)\,, \qquad \psi_{k,j}(z_i^0,\dots,z_i^k;y_i) := \frac{\partial f_k}{\partial z_i^j}(z_i^0,\dots,z_i^k;y_i)\,.$$

Since $f_k \in \mathrm{PL}(m)$, we have $|\psi_{k,j}(\boldsymbol{s})| \le C(1+\|\boldsymbol{s}\|)^{m-1}$. Let $\boldsymbol{s}_i = (z_i^0,\dots,z_i^k;y_i)$, and fix $R \ge 1$ a large constant. Then

$$\begin{aligned}\left|\mathsf{b}_{k,j} - \frac{1}{n}\sum_{i=1}^n \psi_{k,j}(\boldsymbol{s}_i)\,\mathbf{1}_{\|\boldsymbol{s}_i\|\le R}\right| &\le \frac{C}{n}\sum_{i=1}^n (1+\|\boldsymbol{s}_i\|)^{m-1}\,\mathbf{1}_{\|\boldsymbol{s}_i\|\ge R}\\ &\le \frac{C}{nR}\sum_{i=1}^n (1+\|\boldsymbol{s}_i\|)^m \le \frac{C'}{R}\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\operatorname*{p\text{-}lim}_{n\to\infty}\mathsf{b}_{k,j} &= \lim_{R\to\infty}\operatorname*{p\text{-}lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \psi_{k,j}(\boldsymbol{s}_i)\mathbf{1}_{\|\boldsymbol{s}_i\|\le R}\\ &\overset{\text{(a)}}{=} \lim_{R\to\infty}\mathbb{E}\big\{\psi_{k,j}(Z_0,\dots,Z_k;Y)\mathbf{1}_{\|(Z_0,\dots,Z_k;Y)\|\le R}\big\}\\ &\overset{\text{(b)}}{=} \mathbb{E}\big\{\psi_{k,j}(Z_0,\dots,Z_k;Y)\big\}\,.\end{aligned}$$

Here (a) follows since the induction hypothesis $\mathcal{A}_3(k)$ implies weak convergence of the empirical distribution of $(\boldsymbol{z}^0,\dots,\boldsymbol{z}^k;\boldsymbol{y})$ to the law of $(Z_0,\dots,Z_k;Y)$, and this implies convergence of expectations of bounded functions via Portmanteau's theorem. (Recall indeed that $(Z_1,\dots,Z_k)$ is a nondegenerate Gaussian vector, and the stated assumption about the joint law of (Z_0,Y).) Further, (b) follows by dominated convergence since $\psi_{k,j}$ is bounded by a polynomial. This implies

$$\operatorname*{p\text{-}limsup}_{n\to\infty}\frac{1}{n^{1/\ell}}\|\boldsymbol{v}_3\|_\ell \le C\sum_{j=1}^k \operatorname*{p\text{-}limsup}_{n\to\infty}\frac{1}{n^{1/\ell}}\|\boldsymbol{f}^j\|_\ell \le C\,.$$

4. Consider term $\boldsymbol{v}_4$. Define $\boldsymbol{Z}_j := [\boldsymbol{z}^1|\cdots|\boldsymbol{z}^{j+1}] \in \mathbb{R}^{n\times(j+1)}$, $\boldsymbol{R}_j := [\boldsymbol{0}|\boldsymbol{r}^1|\cdots|\boldsymbol{r}^j] \in \mathbb{R}^{n\times(j+1)}$, and $\boldsymbol{Y}_j := \boldsymbol{Z}_j + \boldsymbol{R}_j$. With these notations we have $\boldsymbol{A}\boldsymbol{F}_{k-1} = \boldsymbol{Y}_{k-1}$, and therefore,

$$\begin{aligned}\boldsymbol{v}_4 &= \boldsymbol{P}_{k-1}^{\perp}\boldsymbol{A}\boldsymbol{F}_{k-1}(\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{F}_{k-1})^{-1}\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{f}^k\\ &= \boldsymbol{Y}_{k-1}(\boldsymbol{w}^k)^{\mathsf{T}} - \boldsymbol{P}_{k-1}\boldsymbol{Y}_{k-1}(\boldsymbol{w}^k)^{\mathsf{T}} := \boldsymbol{v}_{4,a} + \boldsymbol{v}_{4,b}\,,\end{aligned}$$

where we defined the vector $\boldsymbol{w}^k := (\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{F}_{k-1})^{-1}\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{f}^k \in \mathbb{R}^k$. Using again—as done above—(A.13), and the induction hypothesis $\mathcal{A}(k)$, we obtain $\|\boldsymbol{w}^k\|_\infty = O(1)$. Further, again by the induction hypothesis, $\|\boldsymbol{z}^j\|_\ell/n^{1/\ell} = O(1)$ and (as proved in the analysis of term $\boldsymbol{v}_3$) $\|\boldsymbol{r}^j\|_\ell/n^{1/\ell} = O(1)$ for all $j \le k$. Therefore,

$$\operatorname*{p\text{-}limsup}_{n\to\infty}\frac{1}{n^{1/\ell}}\|\boldsymbol{v}_{4,a}\|_\ell \le \operatorname*{p\text{-}limsup}_{n\to\infty}\sum_{j=1}^{k+1}|w_j^k|\frac{1}{n^{1/\ell}}\big(\|\boldsymbol{z}^j\|_\ell + \|\boldsymbol{r}^{j-1}\|_\ell\big) \le C\,.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

For $\boldsymbol{v}_{4,b}$ we notice that

$$\boldsymbol{v}_{4,b} = \sum_{j=1}^{k} h_j \boldsymbol{f}^{j-1}\,, \qquad (h_1,\dots,h_k) = (\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{F}_{k-1})^{-1}(\boldsymbol{F}_{k-1}^{\mathsf{T}}\boldsymbol{Y}_{k-1})(\boldsymbol{w}^k)^{\mathsf{T}}\,.$$

Using again the induction hypothesis, we get $\|\boldsymbol{h}\|_\infty = O(1)$ and $\|\boldsymbol{f}^j\|_\ell/n^{1/\ell} = O(1)$, and the claim follows.

5. Finally, term $\boldsymbol{v}_5$ is treated almost exactly as term $\boldsymbol{v}_4$: we do not repeat the same argument.

This concludes the proof of $\mathcal{A}_1(k+1)$.

Proof of $\mathcal{A}_2(k+1)$. Fix any $j \le k$, and let $\boldsymbol{s}_i := (z_i^0,\dots,z_i^j,y_i)$, $\boldsymbol{s}_i^M := (z_i^{M,0},\dots,z_i^{M,j},y_i)$. We then have

$$\begin{aligned}
&\frac{1}{n}\big\|f_j^M(\boldsymbol{z}^{M,0},\dots,\boldsymbol{z}^{M,j};\boldsymbol{y}) - f_j(\boldsymbol{z}^0,\dots,\boldsymbol{z}^j;\boldsymbol{y})\big\|_2\\
&\le \frac{2}{n}\sum_{i=1}^n \big[f_j^M(\boldsymbol{s}_i^M) - f_j(\boldsymbol{s}_i^M)\big]^2 + \frac{2}{n}\sum_{i=1}^n \big[f_j(\boldsymbol{s}_i^M) - f_j(\boldsymbol{s}_i)\big]^2\\
&\le \frac{C}{n}\sum_{i=1}^n (1+\|\boldsymbol{s}_i^M\|_2)^m \mathbf{1}_{\|\boldsymbol{s}_i^M\|_2\ge M} + \frac{C}{n}\sum_{i=1}^n (1+\|\boldsymbol{s}_i^M\|_2+\|\boldsymbol{s}_i\|_2)^{m-1}\|\boldsymbol{s}_i-\boldsymbol{s}_i^M\|_2\\
&\le \frac{C}{nM}\sum_{i=1}^n (1+\|\boldsymbol{s}_i^M\|_2)^{m+1}\\
&+ C\left(\frac{1}{n}\sum_{i=1}^n (1+\|\boldsymbol{s}_i^M\|_2+\|\boldsymbol{s}_i\|_2)^{2(m-1)}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{s}_i-\boldsymbol{s}_i^M\|_2^2\right)^{1/2}\\
&\le O(1)\,\frac{1}{M} + O(1)\sum_{i=1}^j \frac{1}{\sqrt{n}}\|\boldsymbol{z}^{M,i}-\boldsymbol{z}^i\|_2\,,
\end{aligned}$$

where in the last step we used the induction hypotheses $\mathcal{A}_1(j)$, $j \le k$. Further, using the induction hypothesis $\mathcal{A}_2(i)$, $i \le k$, implies that

$$\lim_{M\to\infty} \text{p-limsup}_{n\to\infty} \frac{1}{n}\big\|f_j^M(\boldsymbol{z}^{M,0},\dots,\boldsymbol{z}^{M,j};\boldsymbol{y}) - f_j(\boldsymbol{z}^0,\dots,\boldsymbol{z}^j;\boldsymbol{y})\big\|_2 = 0\,. \tag{A.14}$$

By a similar argument we conclude that, for all $j \le k$,

$$\lim_{M\to\infty} \text{p-limsup}_{n\to\infty} \big|\mathsf{b}_{kj}^M - \mathsf{b}_{kj}\big| = 0\,. \tag{A.15}$$

Recall that $\|\boldsymbol{A}\|_{\rm op} \le 3$ with probability $1-\exp(-n/C)$, and therefore, with high probability,

$$\begin{aligned}
\big\|\boldsymbol{z}^{M,k+1}-\boldsymbol{z}^{k+1}\big\|_2 \le{}& C\big\|f_k^M(\boldsymbol{z}^{M,0},\dots,\boldsymbol{z}^{M,k};\boldsymbol{y}) - f_k(\boldsymbol{z}^0,\dots,\boldsymbol{z}^k;\boldsymbol{y})\big\|_2\\
&+\sum_{j=1}^k |\mathsf{b}_{kj}^M-\mathsf{b}_{kj}|\big\|f_j^M(\boldsymbol{z}^{M,0},\dots,\boldsymbol{z}^{M,j};\boldsymbol{y})\|_2\\
&+\sum_{j=1}^k |\mathsf{b}_{kj}|\big\|f_j^M(\boldsymbol{z}^{M,0},\dots,\boldsymbol{z}^{M,k};\boldsymbol{y}) - f_j(\boldsymbol{z}^0,\dots,\boldsymbol{z}^j;\boldsymbol{y})\big\|_2\,.
\end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Using (A.14), (A.15), together with the fact that $\|f_j^M(\boldsymbol{z}^{M,0},\dots,\boldsymbol{z}^{M,j};\boldsymbol{y})\|_2/\sqrt{n} = O(1)$ and $|\mathsf{b}_{kj}| = O(1)$ which follows by the induction hypothesis $\mathcal{A}_1(k)$, we get the desired claim $\mathcal{A}_2(k+1)$; cf. (A.8).

Proof of $\mathcal{A}_3(k+1)$. As above we let $\boldsymbol{s}_i := (z_i^0,\dots,z_i^{k+1},y_i)$, $\boldsymbol{s}_i^M := (z_i^{M,0},\dots,z_i^{M,k+1},y_i)$. We then have, with high probability,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^n \psi(\boldsymbol{s}_i) - \frac{1}{n}\sum_{i=1}^n \psi(\boldsymbol{s}_i^M)\right| \\
&\quad\le \frac{C}{n}\sum_{i=1}^n (1+\|\boldsymbol{s}_i^M\|_2+\|\boldsymbol{s}_i\|_2)^{m-1}\|\boldsymbol{s}_i-\boldsymbol{s}_i^M\|_2 \\
&\quad\overset{\text{(a)}}{\le} \left(\frac{1}{n}\sum_{i=1}^n (1+\|\boldsymbol{s}_i^M\|_2+\|\boldsymbol{s}_i\|_2)^{2(m-1)}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{s}_i-\boldsymbol{s}_i^M\|_2^2\right)^{1/2} \\
&\quad\overset{\text{(b)}}{\le} C\sum_{i=1}^{k+1}\frac{1}{\sqrt{n}}\|\boldsymbol{z}^{M,i}-\boldsymbol{z}^i\|_2\,,
\end{aligned}$$

where (a) is Cauchy–Schwarz and (b) follows from $\mathcal{A}_1(k+1)$. Using $\mathcal{A}_2(k+1)$ (i.e., (A.8) with k replaced by $k+1$), we obtain the desired claim.

Appendix B. A summary of the algorithm. In this appendix we provide a pseudo-code description of the algorithm of Theorem 2, for the reader's convenience. As usual, $\odot$ 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 $\|\partial_x^2\Phi(k\delta,\boldsymbol{x}^k)\|$ is the ℓ_2 norm of the vector whose ith component is $\partial_x^2\Phi(k\delta,x_i^k)$.

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. (1.3)). Compute solution Φ to PDE (1.2), with $\mu=\mu_\beta$. Compute $q_*(\beta)=\sup\{q:\ q\in\mathrm{supp}(\mu_\beta)\}$ (Edwards–Anderson parameter). Initialize $\boldsymbol{u}^{-1}=\boldsymbol{0}$, $\boldsymbol{u}^0\sim\mathsf{N}(0,\delta\boldsymbol{I}_n)$, $\boldsymbol{g}^{-1}=\boldsymbol{1}$, $\boldsymbol{g}^{-2}=\boldsymbol{0}$, $\mathsf{b}_0=0$. **for** $k\leftarrow 0$ **to** $\lfloor q_*/\delta\rfloor$ **do**
  $\boldsymbol{u}^{k+1}=\boldsymbol{A}(\boldsymbol{g}^{k-1}\odot\boldsymbol{u}^k)-\mathsf{b}_k\boldsymbol{g}^{k-2}\odot\boldsymbol{u}^{k-1}$ $\boldsymbol{x}^k=\boldsymbol{x}^{k-1}+\beta^2\mu(k\delta)\,\partial_x\Phi(k\delta,\boldsymbol{x}^{k-1})\,\delta+\beta\boldsymbol{u}^k$
  $\boldsymbol{g}^k=\sqrt{n}\partial_x^2\Phi(k\delta,\boldsymbol{x}^k)/\|\partial_x^2\Phi(k\delta,\boldsymbol{x}^k)\|$ $\mathsf{b}_{k+1}=\sum_{i=1}^n g_i^k/n$
end
Compute $\boldsymbol{m}=\sum_{k=1}^{\lfloor q_*/\delta\rfloor}\boldsymbol{g}^{k-1}\odot\boldsymbol{u}^k$. Round $\boldsymbol{m}$ 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 how to solve the PDE (1.2). 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 [CR02, OSS07, SO08].

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

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 $\tilde{z}_i \leftarrow \text{sign}(h_i)$;
end
return $\boldsymbol{\sigma}_* = \tilde{\boldsymbol{z}}$.

Acknowledgment. The author is grateful to Eliran Subag for an inspiring presentation of his work [Sub18] delivered at the workshop "Advances in Asymptotic Probability" in Stanford, and for a stimulating conversation.

REFERENCES

[ABE+05] S. Arora, E. Berger, H. Elad, G. Kindler, and M. Safra, *On non-approximability for quadratic programs*, in Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2005, pp. 206–215.

[ABM18] L. Addario-Berry and P. Maillard, *The Algorithmic Hardness Threshold for Continuous Random Energy Models*, preprint, https://arxiv.org/abs/1810.05129, 2018.

[AC15] A. Auffinger and W.-K. Chen, *The Parisi formula has a unique minimizer*, Comm. Math. Phys., 335 (2015), pp. 1429–1444.

[AC17] A. Auffinger and W.-K. Chen, *Parisi formula for the ground state energy in the mixed p-spin model*, Ann. Probab., 45 (2017), pp. 4617–4631.

[ACZ17] A. Auffinger, W.-K. Chen, and Q. Zeng, *The SK Model is Full-step Replica Symmetry Breaking at Zero Temperature*, preprint, https://arxiv.org/abs/1703.06872, 2017.

[AGZ09] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, Cambridge, UK, 2009.

[AMS20] A. El Alaoui, A. Montanari, and M. Sellke, *Optimization of Mean-field Spin Glasses*, preprint, https://arxiv.org/abs/2001.00904, 2020.

[BCKM98] J.-P. Bouchaud, L. F. Cugliandolo, J. Kurchan, and M. Mézard, *Out of equilibrium dynamics in spin-glasses and other glassy systems*, in Spin Glasses and Random Fields, World Scientific, 1998, pp. 161–223.

[BLM15] M. Bayati, M. Lelarge, and A. Montanari, *Universality in polytope phase transitions and message passing algorithms*, Ann. Probab., 25 (2015), pp. 753–822.

[BM11] M. Bayati and A. Montanari, *The dynamics of message passing on dense graphs, with applications to compressed sensing*, IEEE Trans. Inform. Theory, 57 (2011), pp. 764–785.

[BMN19] R. Berthier, A. Montanari, and P.-M. Nguyen, *State evolution for approximate message passing with non-separable functions*, Inf. Inference, 9 (2029), pp. 33–79.

[Bol14] E. Bolthausen, *An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model*, Comm. Math. Phys., 325 (2014), pp. 333–366.

[CH06] P. Carmona and Y. Hu, *Universality in Sherrington–Kirkpatrick's spin glass model*, Ann. Inst. H. Poincaré Probab. Statist., 42 (2006), pp. 215–222.

[Che17] W.-K. Chen, *Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound*, Ann. Probab., 45 (2017), pp. 3929–3966.

[CK94] L. F. Cugliandolo and J. Kurchan, *On the out-of-equilibrium relaxation of the Sherrington-Kirkpatrick model*, J. Phys. A, 27 (1994), pp. 5749–5772.

[CPS18] W.-K. Chen, D. Panchenko, and E. Subag, *The Generalized TAP Free Energy*, preprint, https://arxiv.org/abs/1812.05066, 2018.

[CPS19] W.-K. Chen, D. Panchenko, and E. Subag, *The Generalized Tap Free Energy* II, preprint, https://arxiv.org/abs/1903.01030, 2019.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

[CR02] A. Crisanti and T. Rizzo, *Analysis of the ∞-replica symmetry breaking solution of the Sherrington-Kirkpatrick model*, Phys. Rev. E, 65 (2002), 046137.

[DMM09] D. L. Donoho, A. Maleki, and A. Montanari, *Message passing algorithms for compressed sensing*, Proc. Natl. Acad. Sci. USA, 106 (2009), pp. 18914–18919.

[DMS17] A. Dembo, A. Montanari, and S. Sen, *Extremal cuts of sparse random graphs*, Ann. Probab., 45 (2017), pp. 1190–1217.

[EVdB01] A. Engel and C. Van den Broeck, *Statistical Mechanics of Learning*, Cambridge University Press, Cambridge, UK, 2001.

[Gam18] D. Gamarnik, *Computing the Partition Function of the Sherrington-Kirkpatrick Model is Hard on Average*, preprint, https://arxiv.org/abs/1810.05907, 2018.

[GS14] D. Gamarnik and M. Sudan, *Limits of local algorithms over sparse random graphs*, in Proceedings of the 5th Conference on Innovations in Theoretical Computer Science, ACM, 2014, pp. 369–376.

[Gue03] F. Guerra, *Broken replica symmetry bounds in the mean field spin glass model*, Comm. Math. Phys., 233 (2003), pp. 1–12.

[JM13] A. Javanmard and A. Montanari, *State evolution for general approximate message passing algorithms, with applications to spatial coupling*, Inf. Inference, 2 (2013), pp. 115–144.

[JT16] A. Jagannath and I. Tobasco, *A dynamic programming approach to the Parisi functional*, Proc. Amer. Math. Soc., 144 (2016), pp. 3135–3150.

[Kab03] Y. Kabashima, *A CDMA multiuser detection algorithm on the basis of belief propagation*, J. Phys. A, 36 (2003), pp. 11111–11121.

[KB19] D. Kunisky and A. S. Bandeira, *A Tight Degree 4 Sum-of-Squares Lower Bound for the Sherrington-Kirkpatrick Hamiltonian*, https://arxiv.org/abs/1907.11686, 2019.

[MM09] M. Mézard and A. Montanari, *Information, Physics and Computation*, Oxford University Press, Oxford, 2009.

[Mon16] A. Montanari, *Local Algorithms for Two Problems on Locally Tree-like Graphs*, talk given at Simons Institute for the Theory of Computing, February 22, 2016, https://simons.berkeley.edu/talks/andrea-montanari-02-22-2016.

[MPV87] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*, World Scientific, 1987.

[MPZ02] M. Mézard, G. Parisi, and R. Zecchina, *Analytic and algorithmic solution of random satisfiability problems*, Science, 297 (2002), pp. 812–815.

[MRX19] S. Mohanty, P. Raghavendra, and J. Xu, *Lifting Sum-of-Squares Lower Bounds: Degree-2 to Degree-4*, preprint, https://arxiv.org/abs/1911.01411, 2019.

[MS16] A. Montanari and S. 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, ACM, 2016, pp. 814–827.

[MV85] M. Mézard and M. A. Virasoro, *The microstructure of ultrametricity*, J. Physique, 46 (1985), pp. 1293–1307.

[Nis01] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*, Oxford University Press, 2001.

[Øks03] B. Øksendal, *Stochastic Differential Equations*, Springer-Verlag, Berlin, 2003.

[OSS07] R. Oppermann, M. J. Schmidt, and D. Sherrington, *Double criticality of the Sherrington-Kirkpatrick model at $T = 0$*, Phys. Rev. Lett., 98 (2007), 127201.

[Pan13a] D. Panchenko, *The Parisi ultrametricity conjecture*, Ann. of Math. (2), 177 (2013), pp. 383–393.

[Pan13b] D. Panchenko, *The Sherrington-Kirkpatrick Model*, Springer, New York, 2013.

[Par79] G. Parisi, *Infinite number of order parameters for spin-glasses*, Phys. Rev. Lett., 43 (1979), 1754.

[Par80] G. Parisi, *A sequence of approximated solutions to the SK model for spin glasses*, J. Phys. A, 13 (1980), L115.

[Sch08] M. J. Schmidt, *Replica Symmetry Breaking at Low Temperatures*, Ph.D. Thesis, Aachen University, Aachen, Germany, 2008.

[SD84] H.-J. Sommers and W. Dupont, *Distribution of frozen fields in the mean-field theory of spin glasses*, J. Phys. C, 17 (1984), 5785.

[SK75] D. Sherrington and S. Kirkpatrick, *Solvable model of a spin-glass*, Phys. Rev. Lett., 35 (1975), 1792.

[SO08] M. J. Schmidt and R. Oppermann, *Method for replica symmetry breaking at and near $T = 0$ with application to the Sherrington-Kirkpatrick model*, Phys. Rev. E, 77 (2008), 061104.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[Sub18] E. Subag, *Following the Ground-states of Full-RSB Spherical Spin Glasses*, https://arxiv.org/abs/1812.04588, 2018.
[Tal06a] M. Talagrand, *Parisi measures*, J. Funct. Anal., 231 (2006), pp. 269–286.
[Tal06b] M. Talagrand, *The Parisi formula*, Ann. of Math. (2), 163 (2006), pp. 221–263.
[Tal10] M. Talagrand, *Mean Field Models for Spin Glasses*, Springer-Verlag, Berlin, 2010.
[TAP77] D. J. Thouless, P. W. Anderson, and R. G. Palmer, *Solution of 'Solvable model of a spin glass,'* Philos. Mag., 35 (1977), pp. 593–601.
[Ton02] F. L. Toninelli, *About the Almeida-Thouless transition line in the Sherrington-Kirkpatrick mean-field spin glass model*, Europhys. Lett., 60 (2002), pp. 764–767.
[Vil08] C. Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer-Verlag, Berlin, 2009.
[WL12] P. G. Wolynes and V. Lubchenko, *Structural Glasses and Supercooled Liquids: Theory, Experiment, and Applications*, John Wiley & Sons, Hoboken, NJ, 2012.

Downloaded 08/03/21 to 171.66.160.111. Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/page/terms

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.