*The Annals of Probability*
2021, Vol. 49, No. 1, 180–205
https://doi.org/10.1214/20-AOP1448
© Institute of Mathematical Statistics, 2021

# THE OVERLAP GAP PROPERTY AND APPROXIMATE MESSAGE PASSING ALGORITHMS FOR $p$-SPIN MODELS

BY DAVID GAMARNIK[1] AND AUKOSH JAGANNATH[2]

[1]*Operations Research Center, and Sloan School of Management, Massachusetts Institute of Technology,* gamarnik@mit.edu
[2]*Department of Statistics and Actuarial Science, Department of Applied Mathematics, University of Waterloo,* a.jagannath@uwaterloo.ca

We consider the algorithmic problem of finding a near ground state (near optimal solution) of a $p$-spin model. We show that for a class of algorithms broadly defined as Approximate Message Passing (AMP), the presence of the Overlap Gap Property (OGP), appropriately defined, is a barrier. We conjecture that, when $p \geq 4$, the model does indeed exhibit OGP (and prove it for the space of binary solutions). Assuming the validity of this conjecture, as an implication the AMP fails to find near ground states in these models, per our result. We extend our result to the problem of finding pure states by means of Thouless, Anderson and Palmer (TAP) based iterations which is yet another example of AMP type algorithms. We show that such iterations fail to find pure states approximately, subject to the conjecture that the space of pure states exhibits the OGP, appropriately stated, when $p \geq 4$.

**1. Introduction.** Given an $N$-tensor $A = (A_{i_1,\dots,i_p}, 1 \leq i_1, \dots, i_p \leq N) \in (\mathbb{R}^N)^{\otimes p}$ of order $p$ and an $N$-vector $u \in \mathbb{R}^N$, define the usual inner tensor product by

$$A(u) \triangleq \sum_{1\leq i_1,\dots,i_p\leq N} A_{i_1,\dots,i_p} u_{i_1} \cdots u_{i_p}. \tag{1.1}$$

Consider the associated normalized variational problem over the binary cube $\mathcal{B}_N \triangleq \{-1, 1\}^N$,

$$\eta_N \triangleq \frac{1}{N} \min_{\sigma\in\mathcal{B}_N} A(\sigma). \tag{1.2}$$

The case when $A$ consists of i.i.d. zero mean Gaussian random entries with variance $1/N^{p-1}$, that is, $\mathcal{N}(0, \frac{1}{N^{p-1}})$ corresponds to the problem of finding a ground state of a $p$-spin model with Gaussian couplings and the (unique) vector $u^*$ achieving the minimization value is called the ground state [38]. The choice of variance $1/N^{p-1}$ and the normalization $1/N$ is dictated by the associated Gibbs distribution defined by assigning probability weight proportional to $\exp(-\beta A(\sigma))$ to each $\sigma \in \mathcal{B}_N$ for some fixed inverse temperature parameter $\beta \in \mathbb{R}_+$. In this case the partition function

$$Z \triangleq \sum_{\sigma\in\mathcal{B}_N} \exp(-\beta A(\sigma))$$

is well approximated by $N\eta_N$ as $\beta$ increases and $\eta_N$ is known to converge to a strictly negative limiting value $\eta^* < 0$ with high probability (w.h.p.) as $N \to \infty$. For us, though, the details of the choice of scaling are immaterial and the variational problem above is equivalent to the case when $A$ consists of i.i.d. standard normal entries and the normalization $1/N$ is skipped. Another standard assumption in the literature is to assume a symmetry of

Received December 2019; revised May 2020.
*MSC2020 subject classifications.* Primary 60K35, 68Q87; secondary 90C26.
*Key words and phrases.* Spin glasses, approximate message passing, overlap gap property.

$A$, for example, assuming that entries are fully determined by i.i.d. entries corresponding to $i_1 \leq \cdots \leq i_p$. This difference is again immaterial. Indeed, consider the tensor $\bar{A}$ defined by

$$\bar{A}_{i_1,\ldots,i_p} = \frac{1}{p!} A^{\pi},$$

where $A^{\pi}$ is defined by

$$A^{\pi}_{i_1,\ldots,i_p} = A_{\pi(i_1,\ldots,i_p)}, \quad 1 \leq i_1, \ldots, i_p \leq N$$

for any permutation $\pi$ of $1, \ldots, p$. Note that $\bar{A}$ is symmetric and satisfies $\bar{A}(u) = A_{\pi}(u)$ for every $\pi$.

In the present paper we focus on the algorithmic question of solving the minimization problem (1.2) approximately and efficiently (in polynomial time). That is, the question is one of existence of a polynomial time algorithm, which for every $\epsilon > 0$ produces a sequence of solution $\sigma_N \in \mathcal{B}_N$, satisfying

$$A(\sigma_N)/N \leq (1-\epsilon)\eta^*,$$

as $N \to \infty$, ideally w.h.p. as $N \to \infty$. This problem was successfully solved recently by Montanari [37] in the case of the Sherrington–Kirkpatrick model which is the special case corresponding to $p = 2$. The result though assumes the validity of a (widely believed) conjecture that the overlap distribution function is strictly increasing. In particular, it assumes the absence of an interval $[\nu_1, \nu_2]$ inside the support of the overlap distribution with zero mass, namely, that the overlap distribution does not exhibit the Overlap Gap Property (OGP). This approach was extended in [20] to general mixed $p$-spin models potentially exhibiting the OGP, where the optimal approximation ratio for their variant of AMP was obtained. The algorithms in [37] and [20] are based on a variant of Approximate Message Passing (AMP) type algorithms, which in the context of spin glasses is well motivated by the so-called Thouless, Anderson and Palmer (TAP) equation describing the magnetization of spins in spin glass models. AMP, as a class of algorithms was also found to be one of the most effective classes of algorithm in many models of signal processing [8, 9, 13, 14, 19, 32, 33], specifically models involving a "planted signal" (which does not apply to our $p$-spin model). The algorithmic result of [37] in its order was inspired by a similar result by Subag [40] regarding the problem of finding a near ground states in a spherical"mixed $p$-spin" model. Here, one considers a linear combination of objectives of the form (1.1), as one varies $p$ with the coefficients being fixed, and optimizing over the unit sphere $\{u : \|u\|_2 \leq 1\}$ instead of $\mathcal{B}_N$. Here, a polynomial time construction of near optimal solutions is provided, under the assumption that the model does not exhibit OGP (see part (2) of Proposition 1 in [40]). For the case of spherical models, the necessary and sufficient conditions for the OGP are known [7, 29, 41]. Both $p$-spin and spherical $p$-spin models are related to the Random Energy Model (REM) considered from the algorithmic perspective by Addario-Berry and Maillard [3], where, in contrast to [37] and [40], algorithmic hardness is established away from the ground state value. One should note, however, that REM is an oracle based optimization problem and thus does not fit classical input size based algorithmic complexity questions arising in the context of $p$-spin and spherical $p$-spin models.

At the same time it is known that the OGP does take place in $p$-spin models when $p \geq 4$, as was established in [15], Theorem 3. In particular, it was shown that, for every even $p \geq 4$, there exists $\mu > 0, 0 < \nu_1 < \nu_2 < 1$ such that w.h.p. for every pair of solutions $\sigma_1, \sigma_2$ satisfying $A(\sigma_j)/N \leq \eta^* + \mu$ for $j = 1, 2$, the associated normalized overlap, defined simply as the normalized absolute value of the inner product $(1/N)|\langle\sigma_1, \sigma_2\rangle|$, is either at most $\nu_1$ or at least some $\nu_2$. This naturally raises the question as to whether the OGP creates a barrier to the success of AMP when $p \geq 4$. The main result of our paper, Theorem 3.3, is to establish precisely this fact under the assumption that a certain relaxed version of the OGP takes place when $p \geq 4$. The relaxed version concerns the optimization problem $\min A(u)$, when $u$ is relaxed to be in Hilbert cube $u \in [-1, 1]^N \triangleq \mathcal{H}_N$, and otherwise is defined in the same way as for $\mathcal{B}_N$. This relaxed version would be a rather straightforward implication of the OGP for binary solutions if one could show that every nearly optimal solution in $\mathcal{H}_N$ is nearly binary. Unfortunately, even this fact is not known, and we leave it as an interesting, though, as we believe, an approachable open problem. As a consequence of our main result, we show that extension of the AMP result of [37] to the case $p \geq 4$ is not possible. As another implication, we show that a natural iterative scheme of computing the fixed point of the TAP equations fails as well in the case $p \geq 4$. We note that this iterative scheme is known to succeed in the high temperature regime due to the result of [14]. Another important class of algorithms ruled out by our negative result is gradient descent type algorithms. Since the gradient of $A(u)$ is a linear combination of the vectors of the form $A(\cdot, u)$, defined below in (1.3), then a discrete implementation of the gradient descent algorithm in the form $u^t = u^{t-1} + \eta_{t-1} \nabla A(u^{t-1})$ for some step choices $\eta_t$ is also a special case of our class of AMP algorithms.

One challenge in establishing our result formally is the formalization of the class of AMP algorithms to begin with. Unfortunately, there is no one formal definition for it, but rather there is a vaguely proposed scheme for a class of iterations inspired by the Belief Propagation type algorithms. The iterations take the form $u^{t+1} = F_t(G_t(u^t), G_{t-1}(u^{t-1}), \ldots, G_0(u^0))$, $t \geq 0$ and are performed for some constant number of rounds $t = 0, 1, \ldots, T$, where $F^t$ is an in general $t$-dependent function involving vector $A(\cdot, u) \in \mathbb{R}^N$, defined by

$$A(\cdot, u) \triangleq \left( \sum_{i_2, \ldots, i_p} A_{i, i_2, \ldots, i_p} u_{i_2} \cdots u_{i_p}, 1 \leq i \leq N \right), \tag{1.3}$$

for any $u \in \mathbb{R}^N$ as well other nonlinear operators $G_t : \mathbb{R}^N \to \mathbb{R}^N$, defined typically through some kind of univariate or $t$ variant nonlinear functions $g_t : \mathbb{R}^t \to \mathbb{R}$ applied coordinatewise in some way. We note that (1.3) is simply a matrix vector product $Au$ when $p = 2$.

Thus, as a first step we introduce a precise class of such iterative algorithms (functions) $F^t$ and associate it with a precise set of assumptions. We show separately that the algorithm of [37] is a special case. We assume that the results $U^t, 0 \leq t \leq T$ of each iteration, are truncated so that the resulting vector always belongs to $\| \cdot \|_\infty$ bounded region of the form $[-M, M]^N$ for some constant $M$. The rational for the truncation is as follows. In the implementation of the AMP, the iterations $F^t, 0 \leq t \leq T$ produce a real-valued vector $U^T \in \mathbb{R}^N$ which is then projected to a vector in $\mathcal{B}_N$ in a way discussed below. The idea here is that $U^T$ is a vector which is "close enough" to some vector $\sigma \in \mathcal{B}_N$ which is a near-ground state. In particular, the typical entries of $U^t$ are "not too far" from interval $[-1, 1]$ and, in particular, are bounded by $M$. We restrict every vector $U^t$ to be in $[-M, M]^N$ for technical convenience. The rounding scheme $[-M, M]^N \to \mathcal{B}_N$, assumed to be adopted by our class of AMP, is similar the one that was used in [37]: first, $U^T$ is projected to vector $V \in \mathcal{H}_N$ via a natural truncation $x \to \min(\max(x, -1), 1)$, and then some rounding scheme $\Pi : \mathcal{H}_N \to \mathcal{B}_N$ is adopted by the algorithm designer, which is guaranteed asymptotically to never lower the quality of the solution, that is, it guarantees that $A(\Pi(V))/N \leq A(V)/N + o(1)$. Our main result, stated more precisely, says that, for *any* AMP algorithm thus defined, the vector $V$ is w.h.p. suboptimal, namely, $A(V)/N$ exceeds $\eta^*$ by some fixed constant $\mu > 0$, w.h.p. as $N \to \infty$. Thus, we establish that the vector $V$ obtained in the penultimate (before $\Pi$) step of the AMP is suboptimal. We note that it is precisely this vector, which is shown to be nearly optimal in the case $p = 2$, in the argument of [37]. The last step of converting a real vector $V \in \mathcal{H}_N$ to $\Pi(V) \in \mathcal{B}_N$ is just used there in order to obtain a genuinely binary vector. We do not establish that the ultimate vector $\Pi(V) \in \mathcal{B}_N$ is sub-optimal, and this is a limitation of our technique. We note, however, that showing near optimality of $\pi(V)$ without showing near optimality of $V$ would amount to believing that the rounding $\Pi$ is somehow mysteriously capable of producing near optimal binary solution $\Pi(V)$ from a presumably far from optimal fractional solution $V \in \mathcal{H}_N$, which is something which does not seem to be plausible, and something which is not established in [37]. Nevertheless, it would be admittedly a more complete result to show directly that $\Pi(V)$ is far from optimal, without assuming the same for $V$, but we are currently unable to make this argument rigorous and leave it for further investigation.

*Proof of the main result. Outline.* We now describe the main ingredients of our proof. First, as a consequence of a result established in [15], we show that the OGP holds w.h.p. not just for one instance of $A$ but for a continuous family of sets of the form $\mathcal{A} = (\sqrt{1-\tau}A + \sqrt{\tau}\hat{A})$, $\tau \in [0, 1]$ where $\hat{A}$ is an independent instance of $A$. Note that, for each fixed $\tau$, the corresponding tensor has the same distribution as $A$. An easy consequence of the OGP result in [15] and the chaos result of [16] is the fact, which we prove in this paper (Theorem 3.4), that the OGP holds for $\mathcal{A}$ as well in the sense that for any two $A_{\tau_1}, A_{\tau_2} \in \mathcal{A}$ and any $\sigma_1, \sigma_2 \in \mathcal{B}_N$ satisfying $A_{\tau_j}(\sigma_j)/N \le \eta^* + \mu$, it is again the case that $N^{-1}|\langle\sigma_1, \sigma_2\rangle| \in [0, \nu_1] \cup [\nu_2, 1]$, for the same values $\nu_1, \nu_2, \mu$. The chaos property, roughly speaking, says for any fixed $\tau_1 \neq \tau_2$ near optimal solutions of $A_{\tau_1}$ and $A_{\tau_2}$ are nearly orthogonal; see Theorem 3.5 below. Our main conjecture regarding OGP (Conjecture 3.2), which we use as an assumption of the main result, is the conjecture that OGP holds, in fact, for near optimal solutions in $\mathcal{H}_N$, as opposed to those in $\mathcal{B}_N$ for the same family of instances. Establishing this conjecture is an interesting open question.

Our main ingredient of the proof is then to show that the iterations $U^T = U^T(A)$ as function of $A$ are sufficiently "continuous" to perturbation of the entries of $A$. Specifically, we obtain an upper bound on $N^{-1}\|U^T(A_\tau) - U^T(A_0)\|_2$ for the interpolation scheme $\mathcal{A}$ which is sufficiently continuous in $\tau$. This result is the subject of Theorem 6.1. A straightforward implication is that the same bound applies to $N^{-1}\|V(A_\tau) - V(A_0)\|_2$, where, as we recall, $V(A)$ is projection of $U^T(A)$ through the truncation $x \to \min(\max(x, -1), 1)$. Separately, we use the independence of $A$ and $\hat{A}$ to argue the near orthogonality of $V(A)$ and $V(\hat{A})$. The continuity result above then is used to show that, for an appropriate choice of $\tau$, it will hold that $N^{-1}\langle V(A_\tau), V(A)\rangle \in (\nu_1, \nu_2)$. The (conjectured) OGP property implies that this choice of $\tau$ corresponds to a sufficiently suboptimal solution $V(A_\tau)$, which contradicts concentration property of $A(V(A_\tau))$, which we establish separately using standard techniques, including Gaussian concentration of measure and Kirszbraun's theorem.

*Prior results on OGP and algorithmic implications.* The concept of OGP originates in the study of spin glass models, specifically the study of overlap distribution of replicas generated according to some associated Gibbs distribution, such as the one described above. Understanding the limiting distribution of overlaps is of an utmost importance to spin glass theory and has recieved significant attention [5, 6, 30, 31]. The first connections between the study of the overlaps and algorithms were made in the context of random constraint satisfaction problems, such as random K-SAT problem and many other similar problems. These problems exhibit an "infamous" gap between the range of parameters for which a satisfying assignment exists vs. those for which solutions can be found in polynomial time. The apparent hardness was linked conjecturally to the clustering (shattering) property of these models which were discovered to appear roughly in the regime where known polynomial time algorithms fail [1, 2, 17, 34, 36]. The clustering property says, roughly speaking, that a large part of the set of satisfying assignments can be partitioned into clusters separated by Hamming distance which is of the order of the size of the model itself. It is notable that the proof technique used to establish such a clustering property actually shows something more:

the overlaps between pairs of typical (random in some appropriate sense) satisfying assignments lie in a disconnected union of intervals $[0, \nu_1] \cup [\nu_2, 1]$. Thus, the set of solutions is disconnected not only with respect to its ambient metric space but also with respect to its one-dimensional projection onto the set of possible overlap values. The proof technique relies on fairly standard application of the moment method. The disconnectivity of overlaps (i.e., the presence of the overlap gaps of the form $(\nu_1, \nu_2)$) was later used as an obstruction to a class of local algorithms, defined as so-called Factors of IID in [15, 26, 27, 39], and for random walk type algorithms in [18]. It is this line of work, which is the closest in spirit to the present one, as one can think of AMP as a natural definition for "local" algorithms defined on dense instances—instances not defined on sparse graphs and hypergraphs.

It is important to note that, while OGP implies the clustering property, the converse in general is not true. Indeed, if the OGP takes place, then one can partition the set of all solutions of interest into those which have overlap at least $\nu_2$ with some arbitrarily marked solution $\sigma^o$ (thus marked "Cluster 1") vs. solutions with overlap at most $\nu_1$ with $\sigma^o$ (thus marked "other clusters"), leading to a set of at least two clusters separated by a significant distance. On the other hand, one can easily create a subset of $\mathcal{B}_N$ for which the set of all overlaps spans the entire interval $[0, 1]$, though at the same time admits clustering partition.

The OGP was further established for some other models, some involving planted signals [21, 22, 25]. It was shown in [22] to be an obstruction to Glauber Dynamics type algorithms by showing that OGP implies the existence of a free energy well, a property which was shown to be a barrier for Markov chain type algorithms in problems involving planted signals [12]. A related notion of free energy barriers associated with these gaps was also shown to be obstructions for local Markov chain type algorithms for problems of the class considered herein [11], where it was also shown that these free energy barriers occur in a broad class of models including both the $p$-spin and spherical $p$-spin models. It can be shown that OGP implies the existence of a free energy barrier at sufficiently low temperatures. It is of interest to establish the broadest class of algorithms for which OGP is a provable barrier. A step in this direction is a recent paper by coauthors and Wein [23], which establishes that OGP is a barrier for algorithms based on low-degree polynomials and algorithms described by the Langevin dynamics, for the $p$-spin, spherical $p$-spin models and the problem of finding a largest independent set of a sparse random graph.

The remainder of the paper is structured as follows. In the next section we introduce the formalism of the AMP algorithms. In Section 3 we give the definition of the OGP, state the corresponding conjecture and state our main result. The validity of OGP for binary solutions, that is, solutions in $\mathcal{B}_N$, is proven in the same section. Some preliminary technical results are established in Section 4. In Section 5 we show that the OGP conjecture for the Hilbert cube $\mathcal{H}_N$ follows from the OGP for the binary cube $\mathcal{B}_N$, provided that the validity of another very plausible Conjecture 3.6 holds which states that every nearly optimal solution in $\mathcal{H}_N$ should be nearly binary. Our main technical result is Theorem 6.1 which is stated and proven in Section 6. We note that it is a purely deterministic result showing that the output of the AMP depends on the values of the tensor $A$ sufficiently continuously. In Section 7 we establish the concentration property of the solution $V$ produced by the AMP around its expectation. Our main theorem is proven in Section 8. In Section 9 we consider TAP solutions and show that a natural class of iterations suggested by TAP fails to find the fixed point of TAP, modulo the same Conjecture 3.2, since the iterations are a special case of the class of AMP algorithms we define. This result is a direct implication of our main result, Theorem 3.3. It contrasts with the positive result of Bolthausen [14] which establishes that these iterations do converge to the solution of TAP equations in the high-temperature setting. In Section 10 we verify that the AMP algorithm constructed in [37] also fit the general definition of AMP introduced in this paper. Finally, we conclude in Section 11 where we state some open questions.

**2. Approximate message passing iterations formalism.** $\langle x, y\rangle = \sum_{1\le i\le N} x_i y_i$ denotes inner product of vectors $x, y \in R^N$. For any tensor $B \in (\mathbb{R}^N)^{\otimes p}$ $\|B\|_2$ denotes the Frobenius norm $\sqrt{\sum_{1\le i_1,\dots,i_p\le N} B^2_{i_1,\dots,i_p}}$, and $\|B\|_{\text{op}}$ denotes the operator norm

$$\max_{u_1,\dots,u_p} B(u_1,\dots,u_p),$$

where the maximum is over all $u_1,\dots,u_p \in \mathbb{R}^N$, $\|u_j\|_2 = 1$, $1 \le j \le p$. By Cauchy–Schwarz inequality $\|B\|_{\text{op}} \le \|B\|_2$.

Throughout the paper $A \in (\mathbb{R}^N)^{\otimes p}$ denotes $N$-size order $p$ tensor consisting of $\mathcal{N}(0, N^{-(p-1)})$ i.i.d. entries. For any $u_1,\dots,u_{p-1} \in \mathbb{R}^N$, let

$$A(u_1,\dots,u_p) = \sum_{1\le i_1,\dots,i_p\le N} A_{i_1,i_2,\dots,i_p} u^1_{i_1}\cdots u^p_{i_p},$$

so that, for any $u \in \mathbb{R}^N$, $A(u) = A(u,\dots,u)$ as in (1.1). Here, $u_r = (u^r_1,\dots,u^r_N)$. For any $u_1,\dots,u_{p-1} \in \mathbb{R}^N$, we also introduce

$$y = A(\cdot, u_1,\dots,u_{p-1}) \in \mathbb{R}^N \tag{2.1}$$

defined by

$$y_i = \sum_{1\le i_1,\dots,i_{p-1}\le N} A_{i,i_1,\dots,i_{p-1}} u^1_{i_1}\cdots u^{p-1}_{i_{p-1}}, \quad 1 \le i \le N.$$

Note that, for any $u_1,\dots,u_p \in \mathbb{R}^N$,

$$A(u_1,\dots,u_p) = \langle u_1, A(\cdot, u_2,\dots,u_p)\rangle. \tag{2.2}$$

Similarly, for any $u \in \mathbb{R}^N$ we write $A(\cdot, u)$ instead of $A(\cdot, u, u, \dots, u)$ for short. We recall the definition of $\eta_N$ from (1.2). Observe that we may view $A(u)$ as a centered Gaussian process indexed by $\mathcal{H}_N$ which has covariance

$$\mathbb{E}[A(u)A(v)] = N\left(\frac{\langle u, v\rangle}{N}\right)^p.$$

In particular, $|\mathbb{E}[A(u)A(v)]| \le N$ for any $u, v \in \mathbb{R}^N$ with $\|u\|_2, \|v\|_2 \le 1$. The following concentration result is then an immediate consequence of the Borell–TIS inequality, Theorem 2.1.11 of [4].

THEOREM 2.1. *For every $\delta > 0$*

$$\mathbb{P}(|\eta_N - \mathbb{E}[\eta_N]| \ge \delta) \le \exp(-(1/4)\delta^2 N),$$

*for all sufficiently large $N$.*

A major consequence of the development in spin glass theory is the existence of the limit

$$\lim_{N\to\infty} \mathbb{E}[\eta_N] = \eta^* < 0, \tag{2.3}$$

which, by Theorem 2.1, also implies that the limit $\eta_N \to \eta^*$ holds w.h.p. as $N \to \infty$.

We now introduce a set of assumptions which are used to define a class of AMP algorithms. Fix a positive integer $T$ and an $M > 0$. Consider two sequences of functions $f_t : [-M, M]^t \to \mathbb{R}$ and $F_t : \mathbb{R} \times [-M, M]^t \to \mathbb{R}$, $1 \le t \le T$.

ASSUMPTION 2.2. $f_t(0)=0$. Furthermore, functions $f_t$, $F_t$ are Lipschitz continuous on their respective domains. More precisely, there exists $\zeta \in \mathbb{R}_+$ such that for all $1 \le t \le T$

$$\sup_{u,v\in[-M,M]^t} \left|f_t(u) - f_t(v)\right| \le \zeta \|u - v\|_2, \tag{2.4}$$

$$\sup_{u,v\in\mathbb{R}\times[-M,M]^t} \left|F_t(u) - F_t(v)\right| \le \zeta \|u - v\|_2. \tag{2.5}$$

The assumption (2.5) says that the function $F_t$ is Lipschitz on an infinite rectangle $\mathbb{R} \times [-M, M]^t$ This will be required due to the special role played by the first variable of $F_t$.

Fix a positive constant $M > 1$. Let $x_M = \max(-M, \min(x, M))$ denote an $M$-truncation for any $x \in \mathbb{R}$. When $x$ is a vector, $x_M$ is assumed to be applied coordinatewise. We now define the iterations forming the basis of AMP. Fix $U^0 \in [-AnsM, M]^N$, and define the sequence $U^t \in \mathbb{R}^N$, $1 \le t \le T$ as follows:

$$U^t = \left[F_t\left(A\left(\cdot, f_t\left(U^0, \dots, U^{t-1}\right)\right), U^0, \dots, U^{t-1}\right)\right]_M \in [-M, M]^N, \tag{2.6}$$

where $F_t$, $f_t$ and $M$ are applied componentwise. In other words, in step $t$, first a vector $f_t(U^0, \dots, U^{t-1}) \in \mathbb{R}^N$ is formed by applying $f_t$ coordinatewise (recall that the domain of $f_t$ is $\mathbb{R}^t$). Then, this vector is used to define vector $A(\cdot, f_t(U^0, \dots, U^{t-1}))$ via (1.3). This vector is concatinated with prior vectors $U^0, \dots, U^{t-1}$ to form an $N \times (t + 1)$ matrix $(A(\cdot, f_t(U^0, \dots, U^{t-1})), U^0, \dots, U^{t-1}) \in \mathbb{R}^{N\times(t+1)}$. Then, function $F_t$ is applied coordinatewise. Finally the $M$-truncation is applied to each of the $N$ coordinates of the vector thus obtained, resulting in $U^t$.

We now describe an algorithm which uses AMP to generate a solution in $\mathcal{B}_N$. For this purposes we assume that the algorithm designer has access to some (computable) projection function $\Pi_N : \mathcal{H}_N \to \mathcal{B}_N$. We discuss this further below.

ALGORITHM 2.3 (AMP Algorithm). *The algorithm is parametrized by* $U^0, M, T, (f_t, 1 \le t \le T), (F_t, 1 \le t \le T), \Pi_N$:

*Input* $A \in (\mathbb{R}^N)^{\otimes p}$.
*Step 1 Compute* $U^T$ *using* (2.6).
*Step 2 Project* $U^T$ *to* $\mathcal{H}_N$ *by applying transformation* $x \to [x]_1 = \max(\min(x, 1), -1)$, *coordinatewise. Denote the resulting vector by* $V \in \mathcal{H}_N$.
*Step 3 Output* $\sigma = \Pi(V) \in \mathcal{B}_N$.

In some sense the details of the projection $\Pi_N$ are immaterial to us since our negative result will be concerned with the quality of the solution $V$ itself and not its projection. Nevertheless, for completeness we describe now the projection used in ([37]), which we denote by $\Pi_N^{\text{sign}}$. The projection was defined only for $p = 2$ which was the case of interest. But it is straightforward to extend the idea to the case of general $p$. Set $z^{(0)} = V$. For $j = 1, \dots, N$, construct $z^{(j)}$ by making all coordinates $\ell \ne j$ of $z^{(j)}$ to be the same as of $z^{(j-1)}$ and setting the $j$th coordinate of $z^{(j)}$ to be the sign opposite of

$$\sum_{j\ne i_1\ne i_2\ne\cdots\ne i_{p-1}} A_{j,i_1,i_2,\dots,i_{p-1}} z^{(j-1)}_{i_1,\dots,i_{p-1}}. \tag{2.7}$$

In particular, the first $j$ coordinates of $z^{(j)}$ are $\pm 1$, but the remaining coordinates are real valued in general. Set $\Pi_N^{\text{sign}}(V) = z^{(N)}$.

Let us comment on the meaning and motivation behind the steps of the AMP algorithm above and also the motivation behind the projection $\Pi_N$ described above and used in ([37]).

The idea is that, when the AMP algorithm succeeds, the vector $V$, while not being an element of the binary cube $\mathcal{B}_N$, should be nearly optimal in the sense that

$$A(V) \approx \inf_{w \in \mathcal{B}_N} A(w)$$

and should not be too far from $\mathcal{H}_N$, so that the projecting $U^T$ to $V \in \mathcal{H}_N$ does not change the objective value significantly. That is, $A(V) \approx A(U^T)$. Next, one observes that $\Pi_N^{\text{sign}}$ effectively rounds $V$ to a vector $z^{(N)}$ in $\mathcal{B}_N$ in such a way that the objective value only decreases asymptotically. This is verified by observing that, for each coordinate $j$, the dependence of $A(V)$ on variable $V_j$ is linear in $V_j$, except for terms $A_{i_1,\dots,i_p}$ with repeating coordinates (i.e., such that $i_\ell = i_r$ for some $1 \le \ell \ne r \le p$). Since $V \in \mathcal{H}_N$ and thus $|V_j| \le 1$, the linearity allows to round $V_j$ to $-1$ or 1 while only decreasing the objective value. This is done trivially by setting $V_j$ to be the sign opposite of the one of the multiplier of $V_j$ which is (2.7). This is done iteratively over all $N$ coordinates. The terms corresponding to repeating coordinates are easily shown to be of lower order of magnitude than the objective value. As a result one obtains a vector $z = z^{(N)} \in \mathcal{B}_N$, satisfying

$$A(z) \lesssim A(V) \approx \inf_{\sigma \in \mathcal{B}_N} A(\sigma).$$

But since $z$ belongs to the solution space itself (the binary cube $\mathcal{B}_N$), it must be the case that, in fact,

$$A(z) \approx A(V) \approx \inf_{\sigma \in \mathcal{B}_N} A(\sigma),$$

and thus the success of AMP is validated. Importantly, the near optimality of $z$ is argued from the near optimality of $V$ itself. This discussion is of key essence to the main result of our work which is stated in the next section.

**3. The OGP conjecture and the main result.** Consider an arbitrary set $\mathcal{A}$ of tensors $A \in (\mathbb{R}^N)^{\otimes p}$.

DEFINITION 3.1. The set $\mathcal{A}$ satisfies the Overlap Gap Property (OGP) with domain $\mathcal{S}_N \subset \mathbb{R}^N$ and parameters $\mu > 0, 0 < \nu_1 < \nu_2 < 1$ if for every pair $A_j \in \mathcal{A}$, $j = 1, 2$ and every $u_j$, $j = 1, 2$ satisfying

$$\frac{1}{N} A_j(u_j) \le \frac{1}{N} \inf_{w \in \mathcal{S}_N} A_j(w) + \mu, \quad j = 1, 2,$$

it holds

$$\frac{|\langle u_1, u_2 \rangle|}{\|u_1\|_2 \|u_2\|_2} \in [0, \nu_1] \cup [\nu_2, 1]. \tag{3.1}$$

Namely, every pair of nearly ($\mu$-close) optimal solutions with respect to any two members of $\mathcal{A}$ cannot have normalized inner product in the interval $(\nu_1, \nu_2)$.

Consider two independent random tensors $A$ and $\hat{A}$ in $(\mathbb{R}^N)^{\otimes p}$ both with i.i.d. $\mathcal{N}(0, 1/N^{p-1})$ entries. Introduce the interpolated set of tensors $A_\tau \triangleq \sqrt{1-\tau} A + \sqrt{\tau} \hat{A}$ with $\tau$ varying in [0, 1]. Note that, for each fixed $\tau$, $A_\tau$ is distributed as $A$. Our main conjecture regarding the OGP concerns the set $\mathcal{A} \triangleq (A_\tau, 0 \le \tau \le 1)$.

CONJECTURE 3.2. *For every even $p \ge 4$ here exists $\mu > 0, 0 < \nu_1 < \nu_2 < 1$ such that $\mathcal{A}$ described above satisfies the OGP with domain $\mathcal{S}_N = \mathcal{H}_N$ and parameters $\mu, \nu_1, \nu_2$, with probability at least $1 - \exp(-cN)$, for some $c > 0$ for all sufficiently large $N$. Furthermore, for every $\delta > 0$ and every $v_1, v_2 \in \mathcal{H}_N$ satisfying $A(v_1)/N \le (1-\delta)\mathbb{E}[\eta_N]$, $\hat{A}(v_2)/N \le (1-\delta)\mathbb{E}[\eta_N]$, it holds $|\langle v_1, v_2 \rangle| \le \delta N$ with probability at least $1 - \exp(-cN)$ for some $c > 0$ and all large $N$.*

Our main result, stated below, assumes the validity of this conjecture. To state this result, let us introduce the following. Let $\mathcal{M}_1([-M, M]^N)$ denote the space of probability measures on $[-M, M]^N$. Let $V(A, T, U^0)$ denote the output of the first two steps of Algorithm 2.3 after $T$ steps with coefficient matrix $A$ and initial data $U^0$, where the entires of $A \in (\mathbb{R}^N)^{\otimes p}$ are i.i.d. $\mathcal{N}(0, N^{-(p-1)})$. Then, the following holds.

THEOREM 3.3. *Let $p \geq 4$ be even. Let $M \geq 1$ and $\zeta > 0$. Assume that $(f_t)$, $(F_t)$ satisfy Assumption 2.2 with Lipschitz constant $\zeta$ and that Conjecture 3.2 holds. Then, there exists $\bar{\mu} > 0$ and $c > 0$, such that for $N$ sufficiently large and any $\nu \in \mathcal{M}_1([-M, M]^N)$, if $U^0 \sim \nu$, then $V = V(A, T, U^0)$ satisfies*

$$\mathbb{P}\left(\frac{A(V)}{N} \geq \frac{\min_{\sigma \in \mathcal{B}_N} A(\sigma)}{N} + \bar{\mu}\right) \geq 1 - \exp(-cN).$$

Thus, we argue the failure of the AMP to find a vector $V \in \mathcal{H}_N$ which is a near optimizer of $A$. As discussed earlier, this is a negative result regarding the performance of AMP, since finding such near optimal $V$ is a key step toward finding a near optimal member $z$ of the binary cube $\mathcal{B}_N$. Ideally, one would establish that the vector $z$ obtained from $V$ via any projection scheme, such as the one described above, is also $\mu$-away from optimality. Unfortunately, our proof technique stops short of that due to the potential sensitivity of the sign function used on obtaining $z$ to perturbation of $A$, thus potentially violating stability used crucially in the proof of our main result. We leave this as an interesting open question.

A partial support to the validity of Conjecture 3.2 above is its validity for the domain $\mathcal{S}_N = \mathcal{B}_N$ as we now establish.

THEOREM 3.4. *For every even $p \geq 4$ here exists $\mu > 0, 0 < \nu_1 < \nu_2 < 1$ such that $\mathcal{A}$ described above satisfies the OGP with domain $\mathcal{S} = \mathcal{B}_N$ and parameters $\mu, \nu_1, \nu_2$, with probability at least $1 - \exp(-cN)$, for some $c > 0$ for all sufficiently large $N$. Furthermore, for every $\delta > 0$ and every $\sigma_1, \sigma_2 \in \mathcal{B}_N$ satisfying $A(\sigma_1)/N \leq (1 - \delta)\mathbb{E}[\eta_N]$, $\hat{A}(\sigma_2)/N \leq (1 - \delta)\mathbb{E}[\eta_N]$, it holds $|\langle \sigma_1, \sigma_2 \rangle| \leq \delta N$ with probability at least $1 - \exp(-cN)$ for some $c > 0$ and all large $N$.*

PROOF. We note that in the case $\mathcal{S}_N = \mathcal{B}_N$, since $\|\sigma\|_2 = N$ for each $\sigma \in \mathcal{B}_N$, the requirement (3.1) in definition of OGP simplifies to

$$\frac{|\langle \sigma_1, \sigma_2 \rangle|}{N} \in [0, \nu_1] \cup [\nu_2, 1].$$

It was established in [15] Theorem 3, that the OGP holds for a single instance of a tensor $A$, that is, $\mathcal{A} = \{A\}$, with probability at least $1 - \exp(-cN)$ for some $c > 0$ and all large $N$. At the same time the following chaos property was established in [16].

THEOREM 3.5 ([16], Theorem 2). *For every $\epsilon > 0$ and $\tau \in (0, 1)$, there exists $C, \tilde{\mu} > 0$ such that with probability $1 - \exp(-CN)/C$, for every $\sigma_1, \sigma_2 \in \mathcal{B}_N$ satisfying $A(\sigma_1)/N \leq \mathbb{E}[\eta_N] + \tilde{\mu}$, $A_\tau(\sigma_2)/N \leq \mathbb{E}[\eta_N] + \tilde{\mu}$ it holds $|\langle \sigma_1, \sigma_2 \rangle| \leq \epsilon N$.*

We now combine these two results. We first claim that it suffices to establish OGP for a discrete finite subsets. Fix $\delta > 0$ such that $1/\delta$ is an integer, and consider $A_\tau$ for $\tau = 0, \delta, 2\delta, \ldots, \delta(1/\delta)$. We assume OGP holds for this set for some $\mu, \nu_1, \nu_2$ for every sufficiently small such $\delta$. Now, for any $\sigma \in \mathcal{B}_N$

$$A_\tau(\sigma) - A(\sigma) = (\sqrt{1 - \tau} - 1)A(\sigma) + \sqrt{\tau} A(\sigma).$$

In light of concentration bound of Theorem 2.1, for any $\epsilon > 0$ we can find small enough $\delta$ so that

$$\max_{0\le k\le (1/\delta)-1} \sup_{0\le\tau\le\delta} \max_{\sigma\in\mathcal{B}_N} \left|A_{k\delta+\tau}(\sigma) - A_{k\delta}(\sigma)\right| \le \epsilon N$$

with probability at least $1-\exp(-cN)$, for some $c > 0$ and large $N$. This means that modulo exponentially small probability, every $\sigma$ satisfying $A_{k\delta+\tau}(\sigma)/N \le \mathbb{E}[\eta_N] + \epsilon$ also satisfies $A_{k\delta}(\sigma)/N \le \mathbb{E}[\eta_N] + 2\epsilon$. Thus, if $\epsilon < \mu - 2\epsilon$, then the set $(A_\tau, \tau \in [0,1])$ satisfies OGP with $\hat{\mu} = \mu - 2\epsilon > 0$ and the same $\nu_1, \nu_2$, provided that the discrete set $(A_{k\delta}, 0 \le k \le 1/\delta)$ satisfies OGP with $\mu, \nu_1, \nu_2$. Thus, we now prove OGP for this discrete set.

Let $\mu, \nu_1, \nu_2$ be OGP parameters for a single instance $A$. By the union bounds over $k = 0, 1, \dots, 1/\delta$, the OGP holds for each $A_{k\delta}$ modulo exponentially small probability. Fix $\delta > 0$. Applying Theorem 3.5, we find $\tilde{\mu}$ so that the theorem claim holds for $\epsilon = \nu_1$ and $\tau = \delta$. By union bounds this also holds for all pairs $A_{k_1\delta}, A_{k_2\delta}, k_1 \ne k_2$ modulo exponentially small probability. Then, OGP holds for $\bar{\mu} \triangleq \min(\tilde{\mu}, \mu), \nu_1, \nu_2$ by considering separately the cases $k_1 = k_2$ and $k_1 \ne k_2$, where in the latter case for every $\sigma_j$, $j = 1, 2$ satisfying $A_{k_j\delta}(\sigma_j)/N \le \mathbb{E}[\eta_N] + \bar{\mu}$, $j = 1, 2$, we simply have $|\langle\sigma_1, \sigma_2\rangle| \le \nu_1 N$.

The second part of the theorem follows immediately from the chaos property of Theorem 3.5 in the special case $\tau = 1$. $\square$

Conjecture 3.2 follows from Theorem 3.4 if we could establish that every nearly optimal solution in $\mathcal{H}_N$ is actually close to a point in $\mathcal{B}_N$. This is quite plausible as one does not expect nearly optimal solutions to exist "deep" inside the Hilbert cube $\mathcal{H}_N$. Unfortunately, we are not able to show this and thus state it as an interesting open problem.

CONJECTURE 3.6. *Suppose the entries of $A \in (\mathbb{R}^N)^{\otimes p}$ are generated i.i.d. according to $\mathcal{N}(0, N^{-(p-1)})$. For every $\epsilon > 0$, there exists $\delta > 0$ such that with probability at least $1 - \exp(-cN)$ for some $c > 0$ and large enough $N$, every $u \in \mathcal{H}_N$ satisfying $A(u)/N \le (1-\delta)\mathbb{E}[\eta_N]$ also satisfies $\min_{v\in\mathcal{B}_N} \|u - v\|_2 \le \epsilon\sqrt{N}$.*

PROPOSITION 3.7. *If Conjecture 3.6 holds, then Conjecture 3.2 holds as well.*

The proof of this implication is found in Section 5.

**4. Preliminary technical results.** In this section we establish several preliminary results. Recall the interpolation set $\mathcal{A}$. We begin by establishing the following uniform version of Theorem 2.1.

THEOREM 4.1. *For every $\delta > 0$, there exists $c > 0$ such that*

$$\mathbb{P}\Big(\sup_{0\le\tau\le 1}\Big|\min_{w\in\mathcal{B}_N} A_\tau(w)/N - \mathbb{E}[\eta_N]\Big| \ge \delta\Big) \le \exp(-cN),$$

*for all sufficiently large $N$.*

PROOF. Consider discretization $\tau_i$, $1 \le i \le N$ of interval $[0,1]$ with equal length $\tau_{i+1} - \tau_i = 1/N$. The proof below actually reveals that any scaling $\tau_{i+1} - \tau_i = o(1)$ suffices. By the union bound the assertion holds for $\max_i$ replacing $\sup_{0\le\tau\le 1}$. Now, fix any $\tau \in [0,1]$, and find $i$ with $\tau_i \le \tau \le \tau_{i+1}$. Fix any $u \in \mathcal{B}_N$. We have $\|u\|_2 = \sqrt{N}$. Then,

$$\left|A_\tau(u) - A_{\tau_i}(u)\right| \le \sup_{\tau_i\le\tau\le\tau_{i+1}} \|A_\tau - A_{\tau_i}\|_{\mathrm{op}} N^{\frac{p}{2}}$$

$$= \sup_{\tau_i \le \tau \le \tau_{i+1}} \left\| (\sqrt{1-\tau} - \sqrt{1-\tau_i})A - (\sqrt{\tau} - \sqrt{\tau_i})\hat{A} \right\|_{\mathrm{op}} N^{\frac{p}{2}}$$
$$\le N^{\frac{p}{2}} (\sqrt{1-\tau_i} - \sqrt{1-\tau_{i+1}}) \|A\|_{\mathrm{op}}$$
$$+ N^{\frac{p}{2}} (\sqrt{\tau_{i+1}} - \sqrt{\tau_i}) \|\hat{A}\|_{\mathrm{op}}.$$

Since $\sqrt{\tau_{i+1}} - \sqrt{\tau_i} \le 1/\sqrt{N}$ for all $i$, applying Lemma 4.3, with probability at least $1 - \exp(-cN)$, the expression above is at most $2C\sqrt{N}$. We conclude that modulo exponentially small in $N$ probability

$$\max_{1 \le i \le N} \sup_{\tau_i \le \tau \le \tau_{i+1}} \max_{u \in \mathcal{B}_N} \left| A_\tau(u) - A_{\tau_i}(u) \right| \le 2C\sqrt{N},$$

from which the assertion follows. $\square$

Next, we show that optimization over the Hilbert cube $\mathcal{H}_N$ results in the asymptotically same optimal value as when optimizing over the binary cube $\mathcal{B}_N$, uniformly over the set $\mathcal{A}$.

THEOREM 4.2. *The assertion of Theorem 4.1 when $\mathcal{B}_N$ is replaced by $\mathcal{H}_N$.*

PROOF. Fix any realization $A$. Clearly, $\min_{w \in \mathcal{B}_N} A(w) \ge \min_{w \in \mathcal{H}_N} A(w)$. For any $A$, write $A(w)$ as a sum $A_{\mathrm{hom}}(w) + A_{\mathrm{rest}}(w)$ where $A_{\mathrm{hom}}$ corresponds to the terms of the form $A_{i_1,\dots,i_p}$ with the property that all $i_1,\dots,i_p$ are distinct (corresponding to a homogeneous polynomial in $w$), and $A_{\mathrm{rest}}(w)$ corresponds to the remaining terms. A simple observation is that

$$\min_{w \in \mathcal{H}_N} A_{\mathrm{hom}}(w) \ge \min_{w \in \mathcal{B}_N} A_{\mathrm{hom}}(w).$$

Indeed, each fractional solution $w \in \mathcal{H}_N$ can be rounded to a binary solution $\bar{w} \in \mathcal{B}_N$ sequentially over coordinates without ever increasing the value. For the second term we use the fact that $\|A_{\mathrm{rest}}\|_{\mathrm{op}} \le CN^{-\frac{p}{2}}$ modulo exponentially small in $N$ probability. The proof of this fact follows similarly to the proof of Lemma 4.3, with only adjusting to the size of the corresponding net being at most $p(4/\epsilon)^{pN-1}$ adjusting for repeated coordinates appearing in terms of $A_{\mathrm{rest}}$. Thus,

$$\left| \min_{w \in \mathcal{H}_N} A_{\mathrm{rest}}(w) \right| \le CN^{-\frac{p}{2}} \|w\|_2^p$$
$$\le CN^{-\frac{p}{2}} N^{\frac{p}{2}}$$
$$= C,$$

which is dominated by $N\mathbb{E}[\eta_N] = \Theta(N)$. This proves that

$$\mathbb{P}\Big( \Big| \min_{w \in \mathcal{H}_N} A(w)/N - \mathbb{E}[\eta_N] \Big| \ge \delta \Big) \le \exp(-cN)$$

for a fix random $A$.

The proof of the uniform over $\mathcal{A}$ version follows similarly to the proof of the uniform version for $\mathcal{B}_N$. $\square$

Recall the following operator norm bound.

LEMMA 4.3. *There exist constants $C, c > 0$, such that*

$$\mathbb{P}(\|A\|_{\mathrm{op}} > CN^{1-p/2}) \le e^{-cN}$$

*for all sufficiently large $N$.*

PROOF. The proof of this result is verbatim that from [10], Lemma 4.7. We include this for completeness.

Let $\mathbb{S}^N = \{x : \|x\|_2 = 1\}$ denote the unit $\ell_2$- ball. We may then view $A$ as a centered Gaussian process on $(\mathbb{S}^N)^{\times p}$, with covariance

$$\mathbb{E}[A(x_1, \dots, x_p)A(y_1, \dots, y_p)] = \frac{1}{N^{(p-1)}} \prod_{1 \le i \le p} \langle x_i, y_i \rangle.$$

This process is rotationally invariant. Fix an $\epsilon > 0$, let $\Sigma_\epsilon$ denote an $\epsilon-$ net for $\mathbb{S}^N$ with respect to $\| \cdot \|_2$ norm and let $\Sigma_\epsilon^p$ denote is $p$-fold cartesian product. By multilinearity of $A$,

$$\|A\|_{\mathrm{op}} \le \sup_{(x_1,\dots,x_k) \in \Sigma_\epsilon^p} A(x_1, \dots, x_p) + \epsilon p \|A\|_{\mathrm{op}}.$$

If we choose $\epsilon$ so that $2p\epsilon \le 1$, we have

$$\mathbb{P}(\|A\|_{\mathrm{op}} > \lambda) \le \mathbb{P}\Big( \bigcup_{\mathbf{x} \in \Sigma_\epsilon^p} \{A(x_1, \dots, x_p) \ge \lambda/2\} \Big).$$

To bound the right-hand side, note that for any fixed $(x_1, \dots, x_p) \in (\mathbb{S}^N)^p$, $A(x_1, \dots, x_p)$ is a centered Gaussian with variance $N^{-p+1}$, so that

$$\mathbb{P}(A(x_1, \dots, x_p) \ge \lambda N^{1-p/2}) \le e^{-N\lambda^2/2},$$

where, in the second line, $(x_1, \dots, x_p)$ is any point in $\Sigma^p$, $|\Sigma^k|$ denotes its cardinality and the final inequality comes from a Gaussian tail bound since $A(x_1, \dots, x_p)$ is a centered Gaussian with variance $N^{-p+1}$. Note furthermore that we may choose this net so that $|\Sigma_\epsilon| \le (4/\epsilon)^N$, [44], Lemma 5.1. Thus, by rotation invariance and a union bound we see that

$$\mathbb{P}(\|A\|_{\mathrm{op}} \ge \lambda N^{1-p/2}) \le \left(\frac{4}{\epsilon}\right)^{pN} e^{-N\lambda^2/2}.$$

Choosing $\lambda$ sufficiently large yields the result. □

Recall the set $\mathcal{A} = \sqrt{1-\tau} A + \sqrt{\tau} \hat{A}$, introduced earlier. Since $\|\sqrt{1-\tau} A + \sqrt{\tau} \hat{A}\|_{\mathrm{op}} \le \|A\|_{\mathrm{op}} + \|\hat{A}\|_{\mathrm{op}}$, we obtain the following immediate extension.

LEMMA 4.4. *There exist constants $C, c > 0$, such that*

$$\mathbb{P}\Big( \sup_{0 \le \tau \le 1} \|A_\tau\|_{\mathrm{op}} > CN^{1-p/2} \Big) \le e^{-cN}$$

*for all sufficiently large $N$.*

PROPOSITION 4.5. *There exists $c_2, c > 0$, which depend on $M$, such that*

$$\mathbb{P}\Big( \sup_{0 \le \tau \le 1} \max_{u,v \in [-M,M]^N} \frac{\|A(\cdot, u) - A(\cdot, v)\|_2}{\|u - v\|_2} \ge c_2 \Big) \le \exp(-cN)$$

*for all sufficiently large $N$.*

PROOF. By multilinearity of $A$, the triangle inequality and the definition of the operator norm,

$$\begin{aligned} \|A(\cdot, u, \dots, u) - A(\cdot, v, \dots, v)\|_2 &\le \|A(\cdot, u - v, \dots, u)\|_2 + \dots + \|A(\cdot, v, \dots, v, u - v)\|_2 \\ &\le \|A\|_{\mathrm{op}} \max\{\|v\|_2^{p-2}, \|u\|_2^{p-2}\} \|u - v\|_2. \end{aligned}$$

Since $u, v \in [-M, M]^N$, it follows that $\|v\|_2, \|u\|_2 \le M\sqrt{N}$. Applying the bound from Lemma 4.4, we obtain the result. □

LEMMA 4.6. *There exist* $c, C > 0$ *such that with probability at least* $1 - \exp(-cN)$ *for all large* $N$ *the following holds*: *for every* $\eta > 0$, *every* $u \in H_N$ *satisfying* $A(u) \leq -\eta N$ *also satisfies* $\|u\|_2 \geq C\eta^{1/p}\sqrt{N}$.

PROOF. Note that by Lemma 4.3, with probability at least $1 - \exp(-cN)$

$$\Big|\max_{\|u\|_2 \leq \delta\sqrt{N}} A(u)\Big| \leq \delta^p N^{\frac{p}{2}} \cdot \|A\|_{\mathrm{op}} \leq C\delta^p N.$$

Thus, if $A(u) \leq -\eta N$, it must be that $\|u\|_2^p \|A\|_{\mathrm{op}} \geq \eta N$, so that

$$\|u\|_2 \geq \frac{1}{C}\eta^{1/p} N^{1/2},$$

where $C$ is as in Lemma 4.3. □

**5. Conjecture 3.2 is implied by Conjecture 3.6.** Now, we show that Conjecture 3.6 implies Conjecture 3.2 and thus, ultimately, the main result Theorem 3.3.

PROOF OF PROPOSITION 3.7. We first show that the property described in Conjecture 3.6 holds *uniformly* for members of the set $\mathcal{A}$.

LEMMA 5.1. *If Conjecture* 3.6 *holds*, *then for every* $\epsilon > 0$ *there exists a* $\delta > 0$ *such that with probability at least* $1 - \exp(-cN)$ *for some* $c > 0$ *and all large enough* $N$ *the following holds*. *For every* $A \in \mathcal{A}$, *if* $u \in \mathcal{H}_N$ *satisfies* $A(u)/N \leq (1-\delta)\mathbb{E}[\eta_N]$, *then it also satisfies* $\min_{v\in\mathcal{B}_N} \|u - v\|_2 \leq \epsilon\sqrt{N}$.

PROOF. The proof is similar to the proof of Theorem 4.1. We provide it for convenience. Consider discretization $\tau_i, 1 \leq i \leq N$ of interval $[0, 1]$ with equal length $\tau_{i+1} - \tau_i = 1/N$. The proof below actually reveals that any scaling $\tau_{i+1} - \tau_i = o(1)$ suffices. By assumption of the conjecture and the union bound, with probability at least $1 - \exp(-cN)$ for large enough $N$, for every $i$ and every $u \in \mathcal{H}_N$ satisfying $A_{\tau_i}(u)/N \leq (1-\delta)\mathbb{E}[\eta_N]$ there exists $v \in \mathcal{B}_N$ with $\|u - v\|_2 \leq \epsilon\sqrt{N}$. (The constant $c > 0$ might need to be adjusted for the union bound purposes.) Fix any $\tau \in [0, 1]$, and find $i$ with $\tau_i \leq \tau \leq \tau_{i+1}$. Let $u \in \mathcal{H}_N$ be such that $A_\tau(u)/N \leq (1 - \delta/2)\mathbb{E}[\eta_N]$. We have $\|u\|_2 \leq \sqrt{N}$. Then,

$$\begin{aligned}
|A_\tau(u) - A_{\tau_i}(u)| &\leq \sup_{\tau_i \leq \tau \leq \tau_{i+1}} \|A_\tau - A_{\tau_i}\|_{\mathrm{op}} N^{\frac{p}{2}} \\
&= \sup_{\tau_i \leq \tau \leq \tau_{i+1}} \big\|(\sqrt{1-\tau} - \sqrt{1-\tau_i})A - (\sqrt{\tau} - \sqrt{\tau_i})\hat{A}\big\|_{\mathrm{op}} N^{\frac{p}{2}} \\
&\leq N^{\frac{p}{2}}(\sqrt{1-\tau_i} - \sqrt{1-\tau_{i+1}})\|A\|_{\mathrm{op}} \\
&\quad + N^{\frac{p}{2}}(\sqrt{\tau_{i+1}} - \sqrt{\tau_i})\|\hat{A}\|_{\mathrm{op}}.
\end{aligned}$$

Since $\sqrt{\tau_{i+1}} - \sqrt{\tau_i} \leq 1/\sqrt{N}$ for all $i$, applying Lemma 4.3, with probability at least $1 - \exp(-cN)$ the expression above is at most $C\sqrt{N} < (\delta/2)|\mathbb{E}[\eta_N]|N$ for all $i$, and thus $A_{\tau_i}(u)/N \leq (1-\delta)\mathbb{E}[\eta_N]$. By the above there exists $v \in \mathcal{B}_N$ with $\|u - v\|_2 \leq \epsilon\sqrt{N}$. □

We now return to the proof of the conjecture. Let $\mu, \nu_1, \nu_2$ be parameters for the OGP for $\mathcal{B}_N$ as in Theorem 3.4. Recall the constant $c_2$ from Proposition 4.5. For any $\epsilon > 0$, choose $\delta = \delta(\epsilon)$ as per the assumption of the conjecture and its implication in Lemma 5.1. As we are free to decrease $\delta$ by construction, let us assume that $\epsilon$ and $\delta$ are small enough to satisfy

$$\delta|\mathbb{E}[\eta_N]| + 2c_2\epsilon \leq (11/12)\mu. \tag{5.1}$$

Since $\overline{\lim}\,\mathbb{E}[\eta_N] < 0$, we may fix $\bar{\mu} > 0$ small enough so that for all large enough $N$

$$\mathbb{E}[\eta_N] + \bar{\mu} \leq (1 - \delta/2)\mathbb{E}[\eta_N].$$

Fix any $\tau_1, \tau_2 \in [0, 1]$ and $u_1, u_2 \in \mathcal{H}_N$ with

$$A_{\tau_j}(u_j)/N \leq \min_{w \in \mathcal{H}_N} A_{\tau_j}(w)/N + \bar{\mu}, \quad j = 1, 2.$$

By the uniform concentration bound Theorem 4.2 we have

$$A_{\tau_j}(u_j)/N \leq (1 - \delta)\mathbb{E}[\eta_N], \quad j = 1, 2.$$

Thus, by Lemma 5.1 we can find $v_j \in \mathcal{B}_N$, $j = 1, 2$ with $\|u_j - v_j\|_2 \leq \epsilon\sqrt{N}$.

Now, let $A$ be either $A_{\tau_1}$ or $A_{\tau_2}$, and $(u, v)$ be either $(u_1, v_1)$ or $(u_2, v_2)$, respectively. Applying (2.2), we have that

$$\begin{aligned} A(v) - A(u) &= \langle v, A(\cdot, v)\rangle - \langle u, A(\cdot, u)\rangle \\ &= \langle v, A(\cdot, v) - A(\cdot, u)\rangle + \langle v - u, A(\cdot, u)\rangle. \end{aligned}$$

Applying the Cauchy–Schwarz inequality,

$$\begin{aligned} \left|A(v) - A(u)\right| &\leq \|v\|_2 \left\|A(\cdot, v) - A(\cdot, u)\right\|_2 + \|v - u\|_2 \left\|A(\cdot, u)\right\|_2 \\ &\leq \sqrt{N}\left\|A(\cdot, v) - A(\cdot, u)\right\|_2 + \|v - u\|_2 \left\|A(\cdot, u)\right\|_2. \end{aligned}$$

Applying Proposition 4.5 for the case $M = 1$, for a certain $c_2$ we have

$$\left\|A(\cdot, v) - A(\cdot, u)\right\|_2 \leq c_2\|v - u\|_2 \leq c_2\epsilon\sqrt{N},$$

modulo $\exp(-cN)$ probability. Similarly, $\|A(\cdot, u)\|_2 \leq c_2\|u\|_2 \leq c_2\sqrt{N}$, modulo same probability. We conclude that modulo, an event with probability $\exp(-cN)$ (with $c > 0$ adjusted appropriately),

$$\left|A(v) - A(u)\right| \leq 2c_2\epsilon N,$$

and thus

$$A(v)/N \leq (1 - \delta)\mathbb{E}[\eta_N] + 2c_2\epsilon.$$

Applying (5.1), we find that $v_1, v_2 \in \mathcal{B}_N$ satisfy $A_{\tau_j}(v_j) \leq \mathbb{E}[\eta_N] + (11/12)\mu$. By the uniform concentration property of Theorem 4.1,

$$A_{\tau_j}(v_j) \leq \min_{w \in \mathcal{B}_N} A_{\tau_j}(w) + \mu.$$

Then, by the OGP, $|\langle v_1, v_2\rangle|/N \in [0, \nu_1] \cup [\nu_2, 1]$. We have

$$\langle u_1, u_2\rangle = \langle u_1 - v_1, u_2\rangle + \langle v_1, u_2 - v_2\rangle + \langle v_1, v_2\rangle.$$

Applying the Cauchy–Schwarz inequality for the first and second term on the right-hand side above and triangle inequality, we obtain

$$\left|\left|\langle u_1, u_2\rangle\right| - \left|\langle v_1, v_2\rangle\right|\right| \leq \left|\langle v_1, v_2\rangle\right| + 2\epsilon,$$

and thus the OGP holds for the set $\mathcal{A}$ over the domain $\mathcal{H}_N$ with parameters $\bar{\mu}$ and $\nu_1 + 2\epsilon, \nu_2 - 2\epsilon$. This completes the proof of the first assertion of Conjecture 3.2.

The proof of the second part is similar. $\square$

**6. Continuous dependence.** When we view Algorithm 2.3 as a discrete time dynamical system, it is natural to expect that this admits a similar dependence on the tensor $A$ as a time-inhomogenous differential equation of the same form. Thus, our proof of continuous dependence of iterations on the tensor $A$ can be viewed as a discrete analogue of similar standard result for differential equations; see, for example, [42], Section 2.4.

Given any tensor $B \in (\mathbb{R}^N)^{\otimes p}$, let

$$c_2(B) \triangleq \sup_{u \neq v \in [-M,M]^N} \frac{\|B(\cdot, u) - B(\cdot, v)\|_2}{\|u - v\|_2}. \tag{6.1}$$

We now state the main result of this section.

THEOREM 6.1. *Let $B, \hat{B} \in (\mathbb{R}^N)^{\otimes p}$, and let $\hat{V}^t, V^t$ denote the corresponding sequences output by Step* 2 *of Algorithm* 2.3 *with the same initial vector $U^0 = \hat{U}^0$. Under Assumption* 2.2 *there is some constant $K$, which depends only on $\zeta$ and $c_2(\hat{B})$, such that for every $T \geq 1$ and $U^0$,*

$$\sup_{1 \leq t \leq T} \|\hat{V}^t - V^t\| \leq K^T \|\hat{B} - B\|_{\mathrm{op}} (\zeta M \sqrt{NT})^{p-1}.$$

PROOF. Define $U^t$ and $\hat{U}^t$ as in Step 1 of Algorithm 2.3, and let $\mathbf{U}_t = (U^s)_{0 \leq s \leq t}$ and $\hat{\mathbf{U}}_t = (\hat{U}^s)_{0 \leq s \leq t}$. Since the map $f(x) = [x]_M$ is 1-Lipschitz for any $M$, we see that the claim of the theorem follows, provided that

$$\beta_N(t) \triangleq \|\mathbf{U}_t - \hat{\mathbf{U}}_t\|_2 = \sqrt{\sum_{s \leq t} \|U^s - \hat{U}^s\|_2^2}$$

satisfies

$$\beta_N(t) \leq \bigl(K\bigl(c_2(\hat{B}) + 1\bigr)\bigr)^t \|\hat{B} - B\|_{\mathrm{op}} (\zeta M \sqrt{Nt})^p, \tag{6.2}$$

as trivially $\|\hat{U}^t - U^t\|_2 \leq \beta_N(t)$.

Thus, we establish (6.2). By 1-Lipschitz continuity of $[\cdot]_M$, we have

$$\begin{aligned}
&\|\hat{U}^{t+1} - U^{t+1}\|_2 \\
&\quad \leq \|F_{t+1}(\hat{B}(\cdot, f_{t+1}(\hat{U}^t, \hat{U}^{t-1}, \dots, \hat{U}_0)), \hat{U}^0, \dots, \hat{U}^t) \\
&\qquad - F_{t+1}(B(\cdot, f_{t+1}(U^0, \dots, U^t)), U_0, \dots, U^t)\|_2.
\end{aligned}$$

Applying the part of Assumption 2.2 regarding $F_t$, we see that

$$\|\hat{U}^{t+1} - U^{t+1}\|_2 \leq \zeta \sqrt{\beta_N^2(t) + \|\hat{B}(\cdot, f_{t+1}(\hat{\mathbf{U}}_t)) - B(\cdot, f_{t+1}(\mathbf{U}_t))\|_2^2},$$

so that

$$\beta_N(t+1) \leq (1 + \zeta^2) \sqrt{\beta_N^2(t) + \|\hat{B}(\cdot, f_{t+1}(\hat{\mathbf{U}}_t)) - B(\cdot, f_{t+1}(\mathbf{U}_t))\|_2^2}.$$

By the triangle inequality,

$$\begin{aligned}
&\|\hat{B}(\cdot, f_{t+1}(\hat{\mathbf{U}}_t)) - B(\cdot, f_{t+1}(\mathbf{U}_t))\|_2 \\
&\quad \leq \|\hat{B}(\cdot, f_{t+1}(\hat{\mathbf{U}}_t)) - \hat{B}(\cdot, f_{t+1}(\mathbf{U}_t))\|_2 + \|\hat{B}(\cdot, f_{t+1}(\mathbf{U}_t)) - B(\cdot, f_{t+1}(\mathbf{U}_t))\|_2 \\
&\quad = I + II.
\end{aligned}$$

By definition of $c_2(\hat{B})$,

$$I \leq c_2(\hat{B}) \|f_{t+1}(\hat{\mathbf{U}}_t) - f_{t+1}(\mathbf{U}_t)\|_2 \leq \zeta c_2(\hat{B}) \beta_N(t).$$

We now analyze $II$. Note that, by Assumption 2.2,

$$\|f_{t+1}(\mathbf{U}_t)\|_2^2 \le \zeta^2 \sum_{0\le i\le t} \|U^i\|_2^2 \le \zeta^2 (t+1)M^2 N.$$

Then,

$$II \le \|\hat{B} - B\|_{\mathrm{op}} \big(M\zeta\sqrt{N(t+1)}\big)^{p-1}.$$

Combining these bounds, we obtain,

$$\|B(\cdot, f_{t+1}(\mathbf{U}_t)) - \hat{B}(\cdot, f_{t+1}(\hat{\mathbf{U}}_t))\|_2 \le \zeta c_2(\hat{B})\beta_N(t) + \|\hat{B} - B\|_{\mathrm{op}} \big(\zeta M\sqrt{N(t+1)}\big)^{p-1}.$$

Plugging this in to the above yields

$$\beta_N(t+1) \le (1+\zeta^2)\sqrt{\beta_N^2(t) + \big(\zeta c_2(\hat{B})\beta_N(t) + \|\hat{B} - B\|_{\mathrm{op}}(\zeta M\sqrt{Nt})^{p-1}\big)^2}.$$

We can write the inequality above in the form

$$\beta_N(t+1) \le K\beta_N(t) + b(t),$$

where $b(t)$ is nondecreasing and $K > 1$ which depends only on $c_2(\hat{B})$ and $\zeta$. The inequality above is a discrete version of Gronwall's inequality and, using $\beta_N(0) = 0$, easily leads to a bound

$$\beta_N(t) \le K^t b(t) = K^t \|\hat{B} - B\|_{\mathrm{op}}(\zeta M\sqrt{Nt})^{p-1} \qquad \square$$

**7. Concentration property of the AMP solution.** In this section we establish that the value associated with the solution $V$ produced by the AMP is concentrated around its expectation.

THEOREM 7.1. *Suppose that Assumption 2.2 holds. For any $\epsilon, M, T, \zeta$, there exists $c > 0$ such that the value $A(V)$ associated with the solution $V$ produced in Step 2 of Algorithm 2.3 satisfies*

$$\sup_{U^0 \in [-M,M]^N} \mathbb{P}\big(|A(V) - \mathbb{E}[A(V)|U^0]| \ge \epsilon N | U^0\big) \le \exp(-cN)$$

*for all sufficiently large $N$.*

PROOF. Fix $U^0$. Our approach is based on Gaussian concentration combined with Kirszbraun's theorem. Let $Z \in (\mathbb{R}^N)^{\otimes p}$ denote a tensor consisting of i.i.d. standard normal entries, so that $A = Z/N^{\frac{p-1}{2}}$ in distribution. We let $f(Z) = A(V(A)) = Z(V(Z/N^{\frac{p-1}{2}}))/N^{\frac{p-1}{2}}$, where $V = V(Z)$ is again the solution produced by AMP viewed as a function of $Z$ and thus introduce $f : (\mathbb{R}^N)^{\otimes p} \to \mathbb{R}$, defined by $f(z) = z(V(z/N^{\frac{p-1}{2}}))/N^{\frac{p-1}{2}}$. We first establish that this function is Lipschitz with an appropriate constant on an appropriate subspace of $(\mathbb{R}^N)^{\otimes p}$. Recall the constant $c_2$ introduced in Proposition 4.5. Let

$$K_{2,N} = \left\{ z \in (\mathbb{R}^N)^{\otimes p} : c_2\left(\frac{z}{N^{\frac{p-1}{2}}}\right) \le c_2 \right\}.$$

In particular, a random $Z$ with i.i.d. standard normal entries satisfies

$$\mathbb{P}(Z \in K_{2,N}) \ge 1 - \exp(-cN) \tag{7.1}$$

for all large enough $N$, where $c$ is as in the proposition.

LEMMA 7.2. *There exists a constant $c = c(M, c_2, \zeta, T)$ such that, for every $z_1, z_2 \in K_{2,N}$,*

$$|f(z_2) - f(z_1)| \leq c\sqrt{N}\|z_2 - z_1\|_2.$$

PROOF. Applying Theorem 6.1, for any $z_1, z_2 \in K_{2,N}$ we have

$$\begin{aligned} \|V(z_2/N^{\frac{p-1}{2}}) - V(z_1/N^{\frac{p-1}{2}})\|_2 &\leq cN^{\frac{p-1}{2}}\|N^{-\frac{p-1}{2}}(z_2 - z_1)\|_{\mathrm{op}} \\ &= c\|z_2 - z_1\|_{\mathrm{op}}, \end{aligned} \tag{7.2}$$

where $c = c(M, c_2, \zeta, T)$ is an appropriate constant.

Next,

$$\begin{aligned} |f(z_1) - f(z_2)| &= N^{-\frac{p-1}{2}}|z_2(V(z_2/N^{\frac{p-1}{2}})) - z_1(V(z_1/N^{\frac{p-1}{2}}))| \\ &\leq N^{-\frac{p-1}{2}}|z_2(V(z_2/N^{\frac{p-1}{2}})) - z_1(V(z_2/N^{\frac{p-1}{2}}))| \\ &\quad + N^{-\frac{p-1}{2}}|z_1(V(z_2/N^{\frac{p-1}{2}})) - z_1(V(z_1/N^{\frac{p-1}{2}}))|. \end{aligned} \tag{7.3}$$

We first analyze the second summand above. For simplicity, we use $v_1, v_2$ in place of $V(z_1/N^{\frac{p-1}{2}})$ and $V(z_2/N^{\frac{p-1}{2}})$. Note $z_1(u) = \langle u, z(\cdot, u)\rangle$. Thus,

$$\begin{aligned} |z_1(v_2) - z_1(v_1)| &= |\langle v_2, z_1(\cdot, v_2)\rangle - \langle v_1, z_1(\cdot, v_1)\rangle| \\ &\leq |\langle v_2, z_1(\cdot, v_2)\rangle - \langle v_2, z_1(\cdot, v_1)\rangle| + |\langle v_2, z_1(\cdot, v_1)\rangle - \langle v_1, z_1(\cdot, v_1)\rangle|. \end{aligned}$$

Then,

$$\begin{aligned} |\langle v_2, z_1(\cdot, v_2)\rangle - \langle v_2, z_1(\cdot, v_1)\rangle| &= |\langle v_2, z_1(\cdot, v_2) - z_1(\cdot, v_1)\rangle| \leq \|v_2\|_2\|z_1(\cdot, v_2) - z_1(\cdot, v_1)\|_2 \\ &\leq M\sqrt{N}N^{\frac{p-1}{2}}c_2(z_1/N^{\frac{p-1}{2}})\|v_2 - v_1\|_2. \end{aligned}$$

Since $z_1 \in K_{2,N}$, we obtain instead a bound

$$M\sqrt{N}N^{\frac{p-1}{2}}c_2\|v_2 - v_1\|_2 \leq M\sqrt{N}N^{\frac{p-1}{2}}c_2c\|z_2 - z_1\|_{\mathrm{op}},$$

where the inequality follows from (7.2).

For the second term we have

$$|\langle v_2, z_1(\cdot, v_1)\rangle - \langle v_1, z_1(\cdot, v_1)\rangle| \leq \|v_2 - v_1\|_2\|z_1(\cdot, v_1)\|_2.$$

Since $z_1 \in K_{2,N}$ and $z_1(\cdot, 0) = 0$, then

$$\|z_1(\cdot, v_1)\|_2 \leq N^{\frac{p-1}{2}}c_2\|v_1\|_2 \leq N^{\frac{p-1}{2}}c_2M\sqrt{N}.$$

Using (7.2) to $\|v_2 - v_1\|_2$, we obtain a bound $c\|z_2 - z_1\|_{\mathrm{op}}N^{\frac{p-1}{2}}c_2M\sqrt{N}$.

Applying both bounds to (7.3) and using $\|\cdot\|_{\mathrm{op}} \leq \|\cdot\|_2$, we complete the proof. □

We now complete the proof of the theorem. For every $z \in (\mathbb{R}^N)^{\otimes p}$, define

$$g(z) = \inf_{\hat{z} \in K_{2,N}} (f(\hat{z}) + c\sqrt{N}\|\hat{z} - z\|_2),$$

where $c$ is as in Lemma 7.2. Kirszbraun's theorem says that $g$ is a Lipschitz continuous function with constant $c\sqrt{N}$ and $g = f$ on $K_{2,N}$. This is easy to verify. Indeed, fix any $z_1, z_2 \in (\mathbb{R}^N)^{\otimes p}$ and $\epsilon > 0$. Find $\hat{z}_1 \in K_{2,N}$ such that

$$|g(z_1) - (f(\hat{z}_1) + c\sqrt{N}\|\hat{z}_1 - z_1\|_2)| \leq \epsilon.$$

Then,

$$g(z_2) - g(z_1) \le f(\hat{z}_1) + c\sqrt{N}\|\hat{z}_1 - z_2\|_2 - \big(f(\hat{z}_1) + c\sqrt{N}\|\hat{z}_1 - z_1\|_2\big) + \epsilon$$
$$= c\sqrt{N}\big(\|\hat{z}_1 - z_2\|_2 - \|\hat{z}_1 - z_1\|_2\big) + \epsilon \le c\sqrt{N}\|z_2 - z_1\|_2.$$

Using a similar reversed inequality, the Lipschitz continuity of $g$ is established. Now, if $z \in K_{2,N}$, then by Lemma 7.2 for every $\hat{z} \in K_{2,N}$,

$$f(\hat{z}) + c\sqrt{N}\|\hat{z} - z\|_2 \ge f(z),$$

implying that the infimum is achieved by $\hat{z} = z$, establishing the Kirszbraun's theorem.

In conclusion, $g$ is a Lipschitz continuous function with constant $c\sqrt{N}$. Thus, by Gaussian concentration (see, e.g., [44]), for every $t \ge 0$,

$$\mathbb{P}\big(\big|g(Z) - \mathbb{E}\big[g(Z)|U^0\big]\big| \ge tN|U^0\big) \le \exp\bigg(-\frac{t^2 N}{4c^2}\bigg).$$

We now use the fact that $f = g$ on the high probability set $K_{2,N}$. Specifically,

$$\mathbb{E}\big[g(Z)|U^0\big] = \mathbb{E}\big[f(Z)|U^0\big] - \mathbb{E}\big[f(Z)\mathbf{1}(Z \notin K_{2,N})|U^0\big] + \mathbb{E}\big[g(Z)\mathbf{1}(Z \notin K_{2,N})|U^0\big].$$

Using $g(Z) \le f(0) + c\sqrt{N}\|Z\|_2 = c\sqrt{N}\|Z\|_2$, we have

$$\begin{aligned}\mathbb{E}\big[g(Z)\mathbf{1}(Z \notin K_{2,N})|U^0\big] &\le c\sqrt{N}\mathbb{E}\big[\|Z\|_2\mathbf{1}(Z \notin K_{2,N})\big] \\ &\le c\sqrt{N}\big(\mathbb{E}\big[\|Z\|_2^2\big]\big)^{\frac{1}{2}}\mathbb{P}^{\frac{1}{2}}(Z \notin K_{2,N}) \\ &\le \exp(-c_4 N)\end{aligned}$$

for some appropriately chosen $c > 0$ and all sufficiently large $N$, where, in the second line, we used that $U^0$ and $A$ are independent and the last inequality follows from (7.1) and from $\mathbb{E}[\|Z\|_2^2] = N^{O(1)}$. Similarly, since $f(Z) \le N^{-\frac{p-1}{2}}\sum|Z_{i_1,\dots,i_p}|$, we also have $\mathbb{E}[f^2(Z)|U^0] = N^{O(1)}$, and thus $\mathbb{E}[f(Z)\mathbf{1}(Z \notin K_{2,N})|U^0]$ is at most $\exp(-c_4 N)$ for all large enough $N$, where we used the same notation for constant $c_4$ as above for convenience. We conclude

$$\big|\mathbb{E}\big[g(Z)|U^0\big] - \mathbb{E}\big[f(Z)|U^0\big]\big| \le \exp(-c_5 N),$$

for some $c_5 > 0$ and all large $N$.

Thus, for any $t > 0$,

$$\begin{aligned}&\mathbb{P}\big(\big|f(Z) - \mathbb{E}\big[f(Z)|U^0\big]\big| \ge tN\big) \\ &\quad\le \mathbb{P}\big(\big|g(Z) - \mathbb{E}\big[f(Z)|U^0\big]\big| \ge tN, \mathbf{1}(Z \in K_{2,N})\big) + \mathbb{P}(Z \notin K_{2,N}) \\ &\quad\le \mathbb{P}\big(\big|g(Z) - \mathbb{E}\big[g(Z)|U^0\big]\big| \ge tN - \big(\mathbb{E}\big[g(Z)|U^0\big] - \mathbb{E}\big[f(Z)|U^0\big]\big)\big) + \exp(-CN) \\ &\quad\le \mathbb{P}\big(\big|g(Z) - \mathbb{E}\big[g(Z)|U^0\big]\big| \ge (t/2)N\big) + \exp(-CN) \le \exp(-c_6 N),\end{aligned}$$

for all large enough $N$ and appropriately chosen $c_6 > 0$. As $U^0$ was arbitrary, the result then follows. □

**8. OGP is an obstruction to AMP. Proof of the main result.** In this section we complete the proof of the main result, Theorem 3.3. Let us begin by first conditioning on the value of $U^0$. Let $A \in (\mathbb{R}^N)^{\otimes p}$ be a tensor with i.i.d. $\mathcal{N}(0, 1/N^{p-1})$ entries. Recall that by assumption $A$ and $U^0$ are independent. Let $V = V(A)$ be the result of the Step 2 of Algorithm

2.3 after $T$ steps. Applying the concentration properties given by Theorems 2.1 and 7.1, it suffices to show that, for every $\epsilon > 0$,

$$\frac{\mathbb{E}[A(V)|U^0]}{N} \geq \mathbb{E}[\eta_N] + \mu - \epsilon$$

for all large enough $N$, where $\mu$ is as in Conjecture 3.2, as in this case the main result would be established for $\bar{\mu} = \mu - 2\epsilon$ for every $\epsilon > 0$.

Thus, for the purposes of contradiction, assume

$$\mathbb{E}[A(V)|U^0]/N \leq \mathbb{E}[\eta_N] + \mu_2 \tag{8.1}$$

for some $\mu_2 < \mu$ for infinitely many $N$.

Generate a tensor $\hat{A} \in (\mathbb{R}^N)^{\otimes p}$ distributed as $A$ and independent from $A$ and $U^0$. Consider the interpolated set $\mathcal{A} = (A_\tau, \tau \in [0,1])$ described in Section 3. Denote by $\mathcal{E}_{\text{OGP}}$ the high-probability OGP event defined in Conjecture 3.2 with parameters $\mu, \nu_1, \nu_2$. Let $V_\tau$ be the vector produced by AMP when run on tensor $A_\tau$, $\tau \in [0,1]$. For any $\tau \in [0,1]$, we have

$$\begin{aligned} c_2(A_\tau) &= \sup_{u \neq v \in [-M,M]^N} \frac{\|\sqrt{1-\tau}A(\cdot,u) + \sqrt{\tau}\hat{A}(\cdot,u) - (\sqrt{1-\tau}A(\cdot,v) + \sqrt{\tau}\hat{A}(\cdot,v))\|_2}{\|u-v\|_2} \\ &\leq \sqrt{1-\tau}c_2(A) + \sqrt{\tau}c_2(\hat{A}) \leq c_2(A) + c_2(\hat{A}). \end{aligned}$$

Applying Proposition 4.5, we have $c_2(A) + c_2(\hat{A}) \leq 2c_2$ modulo exponentially small in $N$ probability. We conclude that $\sup_{\tau \in [0,1]} c_2(A_\tau) \leq 2c_2$ modulo exponentially small probability.

By Theorem 6.1 then modulo exponentially small probability, using $\|\cdot\|_{\text{op}} \leq \|\cdot\|_2$, we have that, for any $\tau_1, \tau_2 \in [0,1]$,

$$\|V^{\tau_1} - V^{\tau_2}\|_2 \leq C^T N^{\frac{p-1}{2}} \|A_{\tau_1} - A_{\tau_2}\|_2$$

for some constant $C > 0$ which incorporates $c_2, \zeta$, and $M$ (and which may change from line to line). We have

$$\begin{aligned} \|A_{\tau_1} - A_{\tau_2}\|_2 &= \|\sqrt{1-\tau_1}A - \sqrt{1-\tau_2}A + \sqrt{\tau_1}\hat{A} - \sqrt{\tau_2}\hat{A}\|_2 \\ &\leq \big(|\sqrt{1-\tau_1} - \sqrt{1-\tau_2}| + |\sqrt{\tau_1} - \sqrt{\tau_2}|\big)\big(\|A\|_2 + \|\hat{A}\|_2\big). \end{aligned}$$

Since $\|A\|_2^2$ is distributed as $N^{-p-1}\sum_{1 \leq i \leq N^p} Z_i^2$, which is $N$ in expectation, then by standard large deviations estimates, $\mathbb{P}(\|A\|_2 \geq cN)$ is exponentially small for any $c > 1$. In particular, $\|A\|_2 + \|\hat{A}\|_2 \leq 4\sqrt{N}$, modulo exponentially small probability.

Combining and letting $h(\tau_1,\tau_2) = |\sqrt{1-\tau_1} - \sqrt{1-\tau_2}| + |\sqrt{\tau_1} - \sqrt{\tau_2}|$, we obtain

$$\big|\|V^{\tau_1}\|_2 - \|V^{\tau_2}\|_2\big| \leq \|V^{\tau_1} - V^{\tau_2}\|_2 \leq C^T N^{\frac{p}{2}} h(\tau_1,\tau_2) \tag{8.2}$$

for all $\tau_1, \tau_2$ modulo exponentially small in $N$ probability.

Next, for any $\tau_1, \tau_2 \in [0,1]$,

$$\frac{\langle V^0, V^{\tau_2}\rangle}{\|V^0\|_2\|V^{\tau_2}\|_2} - \frac{\langle V^0, V^{\tau_1}\rangle}{\|V^0\|_2\|V^{\tau_1}\|_2} = \frac{\|V^{\tau_1}\|_2\langle V^0, V^{\tau_2}\rangle - \|V^{\tau_2}\|_2\langle V^0, V^{\tau_1}\rangle}{\|V^0\|_2\|V^{\tau_1}\|_2\|V^{\tau_2}\|_2}.$$

For the numerator, applying the Cauchy–Schwarz inequality,

$$\begin{aligned} &\big|\|V^{\tau_2}\|_2\langle V^0, V^{\tau_1}\rangle - \|V^{\tau_1}\|_2\langle V^0, V^{\tau_2}\rangle\big| \\ &\quad = \big|\|V^{\tau_2}\|_2\langle V^0, V^{\tau_1}\rangle - \|V^{\tau_1}\|_2\langle V^0, V^{\tau_1}\rangle + \|V^{\tau_1}\|_2\langle V^0, V^{\tau_1}\rangle - \|V^{\tau_1}\|_2\langle V^0, V^{\tau_2}\rangle\big| \\ &\quad \leq \big|\|V^{\tau_2}\|_2 - \|V^{\tau_1}\|_2\big| \, \|V^0\|_2\|V^{\tau_1}\|_2 + \|V^{\tau_1}\|_2\|V^0\|_2\|V^{\tau_2} - V^{\tau_1}\|_2 \end{aligned}$$

Provided (8.2) holds, we obtain

$$\left|\frac{\langle V^0, V^{\tau_2}\rangle}{\|V^0\|_2\|V^{\tau_2}\|_2} - \frac{\langle V^0, V^{\tau_1}\rangle}{\|V^0\|_2\|V^{\tau_1}\|_2}\right| \le \frac{2C^T N^{\frac{p}{2}} h(\tau_1, \tau_2)}{\|V^{\tau_2}\|_2}.$$

Next, we fix $\alpha > 0$, to be specified later, let $\delta = N^{-\alpha}$. We assume for convenience that $1/\delta = N^\alpha$ is an integer. Introduce a discrete sequence $\tau_n = n\delta, 0 \le n \le 1/\delta$. By Lemma 4.6, applying union in $N^\alpha$ terms bound, we have that $\|V^{\tau_n}\| \ge C_2\sqrt{N}$ for some $C_2 > 0$ for all sufficiently large $N$, modulo exponentially small probability. Provided this holds, the bound above can be replaced by

$$\frac{2C^T N^{\frac{p}{2}} h(\tau_1, \tau_2)}{C_2\sqrt{N}} = C^T h(\tau_1, \tau_2) N^{\frac{p-1}{2}}.$$

Now, we have

$$h(\tau_{n_1}, \tau_{n_2}) \le C^T N^{\frac{\alpha}{2}}(\tau_{n_2} - \tau_{n_1}) = C^T N^{\frac{\alpha}{2}} N^{-\alpha}(n_2 - n_1) = C^T N^{-\frac{\alpha}{2}}(n_2 - n_1)$$

for all $n_1, n_2$ and some $C > 0$.

Combining, we conclude that modulo exponentially small in $N$ probability for all $n = 0, \dots, N^\alpha$,

$$\left|\frac{\langle V^0, V^{\tau_{n+1}}\rangle}{\|V^0\|_2\|V^{\tau_{n+1}}\|_2} - \frac{\langle V^0, V^{\tau_n}\rangle}{\|V^0\|_2\|V^{\tau_n}\|_2}\right| \le C^T N^{\frac{p-1}{2}} N^{-\frac{\alpha}{2}}$$

and, provided $\alpha > p - 1$, the bound above is $o(1)$ and, in particular, is smaller than $\nu_2 - \nu_1$ for $N$ sufficiently large.

Next, we examine $\frac{\langle V^0, V^{\tau_n}\rangle}{\|V^0\|_2\|V^{\tau_n}\|_2}$ in the extreme case $n = 0$ and $n = N^\alpha$. The value is clearly 1 when $n = 0$. Applying the second part of Conjecture 3.2, we have that for every $\epsilon$, this value is at most $\epsilon/C^2$ modulo exponentially small probability, where $C$ is the constant from Lemma 4.6. In particular, at $n = N^\alpha$ this value is at most $\nu_1$. It follows that there must exist an index $n^*$, (which is random in general) such that

$$\left|\frac{\langle V^0, V^{\tau_{n^*}}\rangle}{\|V^0\|_2\|V^{\tau_{n^*}}\|_2}\right| \in (\nu_1, \nu_2).$$

Now, in the event that OGP holds, which by Conjecture 3.2 holds modulo exponentially small probability, this implies $A(V^{\tau_{n^*}})/N \ge \mathbb{E}[\eta_N] + \mu$ and, therefore, the larger event

$$\max_{0 \le n \le N^\alpha} A\big(V^{\tau_n}\big)/N \ge \mathbb{E}[\eta_N] + \mu.$$

However, this contradicts assumption (8.1) and the concentration bound of Theorem 7.1 applied in the union over $n = 0, \dots, N^\alpha$ bound. This yields the result conditionally on $U^0$. Since the lower bound we obtain does not depend on $U^0$, we can take the expectation in $U^0$ and obtain the main result.

**9. TAP-type iteration schemes.** One motivations for the AMP algorithm discussed in the Introduction is the prediction that the minimizers of $A(u)$ satisfy a selfconsistent equation, called a "mean-field" equation. In this setting, equations of this type are called Thouless–Anderson–Palmer (TAP) equations, after the work of those three authors in [43] on mean-field equations in the case $p = 2$ on $\mathcal{B}_N$, in a certain physically motivated relaxation. For a discussion of these and related results, see also [35]. In this section we show that, as an implication of Theorem 3.3, the iterative methods designed to produce solutions to TAP-like equations fail, modulo Conjecture3.2.

More precisely, consider the following modification of the objective. Recall the Bernoulli entropy, $S : [-1, 1] \to \mathbb{R}_+$

$$S(x) = \frac{1}{2}(1+x)\log(1+x) + \frac{1}{2}(1-x)\log(1-x).$$

For any $\beta > 0$, let $f_\beta : [-1, 1] \to \mathbb{R}$

$$f_\beta(x) = \frac{\beta^2}{2}\big(1 - x^p - px^{p-1}(1-x)\big).$$

Finally, define the one-parameter family of functions $F_\beta : \mathcal{H}_N \to \mathbb{R}^N$ given by

$$F_\beta(x) = \beta A(x) - S(x) + f_\beta\left(\frac{\|x\|^2}{N}\right).$$

Observe that, as $\beta \to \infty$,

$$\frac{F_\beta(u)}{\beta} \to \begin{cases} A(x) & x \in \mathcal{B}_N, \\ \infty & x \in \mathcal{H}_N \setminus \mathcal{B}_N. \end{cases}$$

Thus, one expects that for $\beta$ very large, minimizers of $F_\beta$ are near minimizers for $A(u)$. In particular, one approach to computing near minimizers for (1.2) would be to produce minimizers of $F$.

Differentiating $F_\beta$, we see that the critical points of $F$ satisfy the fixed point equation

$$x = \tanh\left[\beta \nabla A(x) + 2f'\left(\frac{\|x\|^2}{N}\right)x\right]. \tag{9.1}$$

Thus, one approach to produce these minimizers is to construct solutions of these fixed point equations. The AMP algorithm is one such method, based off of a deep intuition in the physics literature that suggests that in the case $p = 2$.

Another approach would be a more naive approach and, in the spirit of standard AMP iterations, would be to simply iteratively construct solutions to (9.1) as in [14]. It is expected that the critical points of this equation satisfy

$$\frac{\|x\|^2}{N} = q_*(\beta),$$

where $q_*(\beta)$ is an explicit constant, called the Edwards–Anderson order parameter and is given by as in [37]. Motivated by this, consider the following class of AMP iterations:

$$U^t = \tanh\big(\beta A(\cdot, U^{t-1}) + a_{t-2}U^{t-2}\big), \quad U_0 = 1_N q, U_{-1} = 0, 1 \le t \le T,$$

where $q > 0$ is a fixed constant and $a_t$ is any bounded sequence. For instance, we may take $a_t = 2f'(q_*)$ and $q = q_*$. One might make the replacement $a_t \mapsto f'(\|U^t\|_2^2/N)$; however, one can show that this will not change the performance if the original sequence was chosen appropriately. See Section 10 for a similar argument in the more detailed case of the AMP iteration from [37].

As a consequence of OGP, we see that the above iteration will fail to produced fixed points which are also near optimizers of $A(u)$ for $\beta$ large. More precisely, we have the following.

COROLLARY 9.1. *Suppose that the entries of $A \in (\mathbb{R}^N)^{\otimes p}$ are i.i.d. $\mathcal{N}(0, N^{-p+1})$ and $p \ge 4$ is even. Assuming the validity of Conjecture 3.2, there exists a $\bar{\mu} > 0$ such that, for any $M, T > 0$ and $\beta$ sufficiently large, if $V = V(A)$ is the result of the firs two steps of the AMP algorithm after $T$ iterations, then*

$$\frac{1}{\beta}F_\beta(V) \ge \frac{\min_{x\in\mathcal{H}_N} F(x)}{N} + \bar{\mu}.$$

PROOF. First note that this iteration is of the form (2.6), for some functions $F_t$, $f_t$ satisfying Assumption 2.2. Indeed, let

$$f_t(u_1, \dots, u_t) = u_t,$$
$$F_t(u_0, \dots, u_t) = \tanh(\beta u_0 + a_{t-2} u_{t-2}).$$

These functions are Lipschitz on the relevant domains as $\tanh(x)$ is smooth with bounded derivatives. Thus, by Theorem 3.3,

$$A(V) \geq \min_{x \in \mathcal{B}_N} A(x)/N + \bar{\mu}.$$

Now, observe that $F_\beta(x)$ satisfies $F_\beta(x) \geq \beta A(x) - \log(2)$ on $\mathcal{H}_N$. In particular, this is an equality on $\mathcal{B}_N$. As a result,

$$\begin{aligned}\frac{1}{N\beta} F_\beta(V) &\geq \frac{A(V)}{N} - \frac{\log 2}{\beta N} \geq \min_{x \in \mathcal{B}_N} \frac{A(x)}{N} + \bar{\mu} \\ &\geq \min_{x \in \mathcal{B}_N} \frac{F_\beta(x)}{\beta N} + \bar{\mu} + \frac{\log(2)}{2\beta} \geq \min_{x \in \mathcal{H}_N} \frac{F_\beta(x)}{\beta N} + \bar{\mu}/2,\end{aligned}$$

where in the last line we take $\beta > \frac{\log(2)}{2\bar{\mu}}$ by assumption. □

**10. Verification for AMP for *p*-spin models.** In this section we show that the AMP algorithm defined in [37] is a special case of the AMP defined in Section 2, modulo some truncation and averaging steps which we discuss below. Here, $p = 2$ so $A \in \mathbb{R}^{N \times N}$ is a matrix. The algorithm constructed in [37] is as follows. A one-dimensional measure $\mu$ is fixed which is a solution of the minimization problem of the Parisi functional. A function $\Phi : [0, 1] \times \mathbb{R} \to \mathbb{R}$ is a solution of the associated PDF. It is known that $\partial_x \Phi(t, x)$ and $\partial_{xx} \Phi(t, x)$ of this function are Lipschitz continuous. A certain value $q_* \in [0, 1]$ is fixed (it is Edward–Anderson parameter). Let $u^{-1} = 0 \in \mathbb{R}^N$, $u^0$ be i.i.d. standard normal vector in $\mathbb{R}^N$: $u^0 \stackrel{d}{=} \mathcal{N}(0, I_N)$, $g^{-2} = 0 \in \mathbb{R}^N$, $g^{-1} = 1_N \in \mathbb{R}^N$, $b^0 = 0 \in \mathbb{R}^N$. Given $a, b \in \mathbb{R}^N$, $a \cdot b \in \mathbb{R}^N$ denotes a coordinatewise product of $a$ and $b$. Then, for $t = 0, 1, \dots, \lfloor q_*/\delta \rfloor \triangleq T$:

$$u^{t+1} = A(g^{t-1} \cdot u^t) - b^t g^{t-2} \cdot u^{t-1}, \tag{10.1}$$

$$x^t = x^{t-1} + \beta^2 \mu(t\delta) \partial_x \Phi(t\delta, x^{t-1}) \delta + \beta \sqrt{\delta} u^t, \tag{10.2}$$

$$g^t = \sqrt{N} \partial_{xx} \Phi(t\delta, x^t) / \| \partial_{xx} \Phi(t\delta, x^t) \|_2, \tag{10.3}$$

$$b^t = N^{-1} \sum_{1 \leq i \leq N} g_i^t, \tag{10.4}$$

where everywhere the functions are applied coordinatewise.

We first consider modifications of these iterations and justify them. First set $M$ to be a large constant. Since $u^0 \stackrel{d}{=} \mathcal{N}(0, I_N)$, then the fraction of coordinates of $u^0$ with absolute values larger than $N$ decreases to zero as a function of $M$. Replace (10.1) by

$$u^{t+1} = [A(g^{t-1} \cdot u^t) - b_t g^{t-2} \cdot u^{t-1}]_M. \tag{10.5}$$

In the final step of algorithm in [37], the resulting vectors $u^1, \dots, u^t$ are used to construct

$$z = \sqrt{\delta} \sum_{1 \leq t \leq \lfloor q_*/\delta \rfloor} g^{t-1} \cdot u^t,$$

and then $z$ is rounded to a vector in $\mathcal{H}_N$ via $\max(-1, \min(1, \cdot))$ operator. It is thus expected that the truncation of $u^t$ in (10.5) by a value $M$ does not affect the result significantly, provided $M$ is large, though we do not prove this fact.

Next, as an implication of the analysis in [37], as $N \to \infty$, the norm $\|\Phi(t\delta, x^t)\|_2$ is concentrated around a deterministic function of $t$ which has value $\Theta(\sqrt{N})$. In particular, there it is argued that the empirical measure $\mathcal{E}_N^t = \frac{1}{N}\sum \delta_{x_i^t}$ converges to some deterministic limit $\mathcal{E}^t$ weakly almost surely. Since $\partial_{xx}\Phi$ is smooth and bounded ([28], Theorem 4), it then follows that

$$\|\partial_{xx}\Phi(t\delta, x^t)\|/\sqrt{N} = \sqrt{\int \partial_{xx}\Phi(t\delta, y)\, d\mathcal{E}_N^t(y)} \to h(t).$$

Denoting this function by $\sqrt{N}h(t)$, $t = 0, 1, \dots, T$, we thus rewrite (10.3) as

$$g^t = h^{-1}(t)\partial_{xx}\Phi(t\delta, x^t). \tag{10.6}$$

Similarly, $b^t$, which per (10.4) is defined as coordinatewise average of $g^t$, as $N \to \infty$ is concentrated around a deterministic function of $t$, which we denote by $\eta(t)$, $t = 0, 1, \dots, T$. Thus, we replace (10.4) by $b^t = \eta^t$.

We now fit these iterations into our framework, as defined in Section 2. We begin by defining $f_t$. In light of (10.6) replacing (10.3), we may define $f_t : \mathbb{R}^{t+1} \to \mathbb{R}$ as follows:

$$f_t(u^0, \dots, u^t) = g^{t-1}(u^0, \dots, u^{t-1})u^t, \tag{10.7}$$

where the function $g_t : \mathbb{R}^t \to \mathbb{R}$ is a one-dimensional version of $g^t$, namely,

$$g_t(u^0, \dots, u^t) = h^{-1}(t)\partial_{xx}\Phi(t\delta, x^t),$$

where $x^t$ is defined through a one-dimensional version of (10.2),

$$x^t = x^{t-1} + \beta^2\mu(t\delta)\partial_x\Phi(t\delta, x^{t-1})\delta + \beta\sqrt{\delta}u^t.$$

Since $\partial_{xx}\Phi$ and $\partial_{xx}\Phi$ are Lipschitz continuous in the second argument for each fixed first argument, it is then immediate to verify that $g_t : \mathbb{R}^t \to \mathbb{R}$ is Lipschitz continuous as a function of $u^0, \dots, u^t$, wrt $\|\cdot\|_\infty$ norm, say with constant $C_t$ and satisfies $g_t(0) = 0$. This implies $\|g_t(x)\|_\infty \le C_t\|x\|_\infty$. Then, by (10.7) we have, for every $u^0, \dots, u^t$ and $v^0, \dots, v^t$,

$$\begin{aligned}|f_t(u^0, \dots, u^t) - f_t(v^0, \dots, v^t)| &= |g^{t-1}(u^0, \dots, u^{t-1})u_t - g^{t-1}(v^0, \dots, v^{t-1})v^t| \\ &\le C_t \max_j(\max(|u^j|, |v^j|)\|(u^0, \dots, u^t) - (v^0, \dots, v^t)\|_\infty.\end{aligned}$$

Thus, $f_t$ is Lipschitz continuous on $[-M, M]^{t+1}$ and thus satisfies the first part of Assumption 2.2.

Now, motivated by (10.1) or rather (10.5), we define $F_t : \mathbb{R}^{t+1} \to \mathbb{R}$ by

$$\begin{aligned}F_t(y, u^0, \dots, u^t) &= y - b_t g^{t-2} \cdot u^{t-1} \\ &= y - \eta(t)h^{-1}(t-2)\partial_{xx}\Phi((t-2)\delta, x_t) \cdot u^{t-1}.\end{aligned}$$

We see that $F_t$ is Lipschitz continuous on $\mathbb{R} \times [-M, M]^{t+1}$ and thus satisfies the second part of Assumption 2.2.

## 11. Some open questions.

We now list some questions which remain open. First, validating Conjectures 3.2 and 3.6 is of interest. It is conceivable that the first of these conjectures can be approached by analyzing the Parisi measure directly on the Hilbert cube $\mathcal{H}_N$, as opposed to the binary cube $\mathcal{B}_N$. Carrying out the corresponding technical analysis of the associated variational problem could be quite daunting though. Another interesting question left open in this work is establishing the negative result for the binary output $\Pi(V)$ of the AMP scheme (Step 3) directly, as opposed to one for the penultimate state $V$. Lifting the truncation

$[\cdot]_M$ assumption adopted by our class of AMP algorithms is another question which remains open.

Next, it would be interesting to extend the main result of this paper to the $p$-spin spherical spin glass model and complement the positive result of Subag [40]. We expect that our negative result extends to this model within the same scope of algorithms almost verbatim. Furthermore, by similar arguments one can show that (continuous time) gradient flow started from a uniform at random point fails to reach near minimizers of the spherical $p$-spin model in $\log(N)$ time with probability tending to 1. That being said, it is important to note that Subag's algorithm is based on an iterative sequence of computations which involve linear projections of gradient and Hessians of $A(u)$ to the linear space orthogonal to $u$. As such, this computational scheme does not formally fit our framework of algorithms. It is conceivable though that the projection step can be approximated well by iterations of the form we consider, say perhaps by imitating the power iteration approach for spectral computations. Perhaps as an easier challenge, one could try to show that Subag's scheme specifically fails to find near ground states in models exhibiting OGP. A related question is whether there exists a connection between the OGP and the algorithmic hardness of the CREM model discussed in [3].

Our approach was formulated in terms of bounded ($N$-independent) number of iterations $T$. It is easy to see though that the proof method extends without a change to the case $T \leq c \log N$ for small enough constant $c$. At the same time we believe that AMP scheme is not effective in computing near ground states, regardless of the scale of the number of iterations. Thus, an interesting open question is to see whether an AMP scheme achieving near ground states can be designed, say when $T = N^{O(1)}$.

Finally, perhaps the most intriguing question which remains open is one regarding the genuine hardness of the problem of finding ground states in models exhibiting the OGP. While formal hardness of problems associated with spin glass models is known, in particular, it is shown in [24] that computing the partition function of the $p$-spin models is hard on average, even in $p = 2$ regime; these results are established using more "standard" average case hardness proof approaches and do not take advantage of the intricate solution space topology, such as the one expressed by OGP. At the same time, as of now we have very compelling consistence of the presence of OGP and the apparent hardness of the associated optimization problem in many models. What is lacking, however, is the formal link between the two within a class of algorithms which is broader than AMP. An interesting and challenging conjecture is that the OGP implies formal average case hardness of the underlying optimization problem, perhaps even within the class of all polynomial time algorithms.

**Acknowledgments.** The first author acknowledges the support from the Office of Naval Research Grant N00014-17-1-2790. The second author acknowledges the support of the National Science Foundation Grant NSF OISE-1604232 and Natural Sciences and Engineering Research Council of Canada [RGPIN-2020-04597, DGECR-2020-00199]. Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG).

## REFERENCES

[1] ACHLIOPTAS, D. and COJA-OGHLAN, A. (2008). Algorithmic barriers from phase transitions. In *Foundations of Computer Science*, 2008. *FOCS'*08. *IEEE* 49*th Annual IEEE Symposium on* 793–802. IEEE, New York.

[2] ACHLIOPTAS, D., COJA-OGHLAN, A. and RICCI-TERSENGHI, F. (2011). On the solution-space geometry of random constraint satisfaction problems. *Random Structures Algorithms* **38** 251–268. MR2663730 https://doi.org/10.1002/rsa.20323

[3] ADDARIO-BERRY, L. and MAILLARD, P. (2019). The algorithmic hardness threshold for continuous random energy models. *Mathematical Statistics and Learning* **2** 77–101.

[4] ADLER, R. J. and TAYLOR, J. E. (2007). *Random Fields and Geometry*. *Springer Monographs in Mathematics*. Springer, New York. MR2319516

[5] AUFFINGER, A. and CHEN, W.-K. (2015). On properties of Parisi measures. *Probab. Theory Related Fields* **161** 817–850. MR3334282 https://doi.org/10.1007/s00440-014-0563-y

[6] AUFFINGER, A., CHEN, W.-K. and ZENG, Q. (2017). The SK model is full-step replica symmetry breaking at zero temperature. Preprint. Available at arXiv:1703.06872.

[7] AUFFINGER, A. CHEN, W.-K. and ZENG, Q. (2020). The SK model is infinite step replica symmetry breaking at zero temperature. *Comm. Pure Appl. Math.* **73** 921–943. https://doi.org/10.1002/cpa.21886

[8] BAYATI, M., LELARGE, M. and MONTANARI, A. (2015). Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.* **25** 753–822. MR3313755 https://doi.org/10.1214/14-AAP1010

[9] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. MR2810285 https://doi.org/10.1109/TIT.2010.2094817

[10] BEN AROUS, G., GHEISSARI, R. and JAGANNATH, A. (2020). Bounding flows for spherical spin glass dynamics. *Comm. Math. Phys.* **373** 1011–1048. MR4061404 https://doi.org/10.1007/s00220-019-03649-4

[11] BEN AROUS, G. and JAGANNATH, A. (2018). Spectral gap estimates in mean field spin glasses. *Comm. Math. Phys.* **361** 1–52. MR3825934 https://doi.org/10.1007/s00220-018-3152-6

[12] BEN AROUS, G., GHEISSARI, R. and JAGANNATH, A. (2020). Algorithmic thresholds for tensor PCA. *Ann. Probab.* **48** 2052–2087. https://doi.org/10.1214/19-AOP1415

[13] BERTHIER, R., MONTANARI, A. and NGUYEN, P.-M. (2020). State evolution for approximate message passing with non-separable functions. *Inf. Inference*: *A Journal of the IMA* **9** 33–79. MR4079177 https://doi.org/10.1093/imaiai/iay021

[14] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. MR3147441 https://doi.org/10.1007/s00220-013-1862-3

[15] CHEN, W.-K., GAMARNIK, D., PANCHENKO, D. and RAHMAN, M. (2019). Suboptimality of local algorithms for a class of max-cut problems. *Ann. Probab.* **47** 1587–1618. MR3945754 https://doi.org/10.1214/18-AOP1291

[16] CHEN, W.-K., HANDSCHY, M. and LERMAN, G. (2018). On the energy landscape of the mixed even $p$-spin model. *Probab. Theory Related Fields* **171** 53–95. MR3800830 https://doi.org/10.1007/s00440-017-0773-1

[17] COJA-OGHLAN, A. and EFTHYMIOU, C. (2011). On independent sets in random graphs. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms* 136–144. SIAM, Philadelphia, PA. MR2857116

[18] COJA-OGHLAN, A., HAQSHENAS, A. and HETTERICH, S. (2017). Walksat stalls well below satisfiability. *SIAM J. Discrete Math.* **31** 1160–1173. MR3656499 https://doi.org/10.1137/16M1084158

[19] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.

[20] EL ALAOUI, A., MONTANARI, A. and SELLKE, M. (2020). Optimization of mean-field spin glasses. Preprint. Available at arXiv:2001.00904.

[21] GAMARNIK, D. and ILIAS, Z. (2017). High dimensional regression with binary coefficients. Estimating squared error and a phase transtition. In *Conference on Learning Theory* 948–953.

[22] GAMARNIK, D., JAGANNATH, A. and SEN, S. (2019). The overlap gap property in principal submatrix recovery. Preprint. Available at arXiv:1908.09959.

[23] GAMARNIK, D., JAGANNATH, A. and WEIN, A. S. (2020). Low-degree hardness of random optimization problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. To appear.

[24] GAMARNIK, D. and KIZILDAG, E. (2018). Computing the partition function of the Sherrington–Kirkpatrick model is hard on average. Preprint. Available at arXiv:1810.05907.

[25] GAMARNIK, D. and LI, Q. (2018). Finding a large submatrix of a Gaussian random matrix. *Ann. Statist.* **46** 2511–2561. MR3851747 https://doi.org/10.1214/17-AOS1628

[26] GAMARNIK, D. and SUDAN, M. (2017). Limits of local algorithms over sparse random graphs. *Ann. Probab.* **45** 2353–2376. MR3693964 https://doi.org/10.1214/16-AOP1114

[27] GAMARNIK, D. and SUDAN, M. (2017). Performance of sequential local algorithms for the random NAE-$K$-SAT problem. *SIAM J. Comput.* **46** 590–619. MR3620150 https://doi.org/10.1137/140989728

[28] JAGANNATH, A. and TOBASCO, I. (2016). A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.* **144** 3135–3150. MR3487243 https://doi.org/10.1090/proc/12968

[29] JAGANNATH, A. and TOBASCO, I. (2017). Low temperature asymptotics of spherical mean field spin glasses. *Comm. Math. Phys.* **352** 979–1017. MR3631397 https://doi.org/10.1007/s00220-017-2864-3

[30] JAGANNATH, A. and TOBASCO, I. (2017). Some properties of the phase diagram for mixed $p$-spin glasses. *Probab. Theory Related Fields* **167** 615–672. MR3627426 https://doi.org/10.1007/s00440-015-0691-z

[31] JAGANNATH, A. and TOBASCO, I. (2018). Bounds on the complexity of replica symmetry breaking for spherical spin glasses. *Proc. Amer. Math. Soc.* **146** 3127–3142. MR3787372 https://doi.org/10.1090/proc/13875

[32] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. MR3311445 https://doi.org/10.1093/imaiai/iat004

[33] KABASHIMA, Y. (2003). A CDMA multiuser detection algorithm on the basis of belief propagation. *J. Phys. A: Math. Gen.* **36** 11111.

[34] MÉZARD, M., MORA, T. and ZECCHINA, R. (2005). Clustering of solutions in the random satisfiability problem. *Phys. Rev. Lett.* **94** 197205. https://doi.org/10.1103/PhysRevLett.94.197205

[35] MÉZARD, M., PARISI, G. and VIRASORO, M. A. (1987). *Spin Glass Theory and Beyond. World Scientific Lecture Notes in Physics* **9**. World Scientific Co., Inc., Teaneck, NJ. MR1026102

[36] MÉZARD, M., PARISI, G. and ZECCHINA, R. (2002). Analytic and algorithmic solution of random satisfiability problems. *Science* **297** 812–815.

[37] MONTANARI, A. (2019). Optimization of the Sherrington–Kirkpatrick Hamiltonian. In 2019 *IEEE* 60*th Annual Symposium on Foundations of Computer Science* (*FOCS*) 1417–1433. IEEE, New York.

[38] PANCHENKO, D. (2013). *The Sherrington–Kirkpatrick model*. Springer Science & Business Media.

[39] RAHMAN, M. and VIRÁG, B. (2017). Local algorithms for independent sets are half-optimal. *Ann. Probab.* **45** 1543–1577. MR3650409 https://doi.org/10.1214/16-AOP1094

[40] SUBAG, E. (2017). The geometry of the Gibbs measure of pure spherical spin glasses. *Invent. Math.* **210** 135–209. MR3698341 https://doi.org/10.1007/s00222-017-0726-4

[41] TALAGRAND, M. (2006). Free energy of the spherical mean field model. *Probab. Theory Related Fields* **134** 339–382. MR2226885 https://doi.org/10.1007/s00440-005-0433-8

[42] TESCHL, G. (2012). *Ordinary Differential Equations and Dynamical Systems. Graduate Studies in Mathematics* **140**. Amer. Math. Soc., Providence, RI. MR2961944 https://doi.org/10.1090/gsm/140

[43] THOULESS, D. J., ANDERSON, P. W. and PALMER, R. G. (1977). Solution of'solvable model of a spin glass'. *Philos. Mag.* **35** 593–601.

[44] VERSHYNIN, R. (2018). *High-Dimensional Probability. Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. MR3837109 https://doi.org/10.1017/9781108231596