

Universality of Approximate Message Passing Algorithms

Wei-Kuo Chen*

Wai-Kit Lam†

June 2, 2020

Abstract

We consider a broad class of Approximate Message Passing (AMP) algorithms defined as a Lipschitzian functional iteration in terms of an $n \times n$ random symmetric matrix A . We establish universality in noise for this AMP in the n -limit and validate this behavior in a number of AMPs popularly adapted in compressed sensing, statistical inferences, and optimizations in spin glasses.

1 Introduction

Motivated by the ideas from belief propagation algorithms, Approximate Message Passing (AMP) algorithms were initially introduced in the context of compressed sensing, see [13, 14, 15, 16]. Thereafter they have received great popularity in a number of emerging applications in data science, statistical physics, etc. concerning the development of efficient algorithms for some randomized estimations and optimizations with large complexity.

One major application has been laid on the subject of matrix estimations, in which one aims to extract the structure of a signal matrix in a randomized environment. A popular setting is the so-called spiked model, where the data arrives as the sum of a noise, an $n \times n$ symmetric random matrix A_n , and the signal, an $n \times n$ symmetric low-rank matrix Z_n ,

$$\hat{A}_n := A_n + Z_n.$$

The goal is to recover the structure of Z_n from the realization of the matrix \hat{A}_n . A typical example one considered in the literature is when A_n is the normalized Gaussian Wigner ensemble and Z_n is given by

$$Z_n = \frac{1}{n} \sum_{\ell=1}^r \gamma_\ell z^\ell \otimes z^\ell, \quad (1.1)$$

where z^1, \dots, z^r are non-random column vectors with $\|z^\ell\|_2 = \sqrt{n}$ and the parameters $\gamma_1, \dots, \gamma_r \geq 0$ are the signal-to-noise ratios (SNR's). In probability and statistics, this spiked model has been intensively studied by means of the spectral method, see [2, 7, 6, 9, 18, 20, 22, 32, 33]. In Bayesian optimal approach, the setting often considered in the literature is to assume that the vectors z^1, \dots, z^r are randomized and their i -th marginal vectors, (z_i^1, \dots, z_i^r) for $1 \leq i \leq n$, are independently sampled from a given prior distribution. It turns out that the corresponding Minimum Mean Square Error Estimator (MMSEE), $\mathbb{E}[(z^1, \dots, z^r) | \hat{A}_n]$, can be connected to the Gibbs expectation of the famous Sherrington-Kirkpatrick (SK) mean-field spin glass model arising from the statistical

*University of Minnesota. Email: wkchen@umn.edu. Partially supported by NSF grant DMS-17-52184

†University of Minnesota. Email: wlam@umn.edu

physics [35]. One peculiar feature within this connection is that this model satisfies the so-called Nishimori identity, namely, the conditional distribution of the vectors z^1, \dots, z^r given the data \hat{A}_n is equal to the distribution of the vector-valued spin configuration of the corresponding SK model. This allows one to fully understand the behavior of the MMSEE and its phase transition in terms of SNR's, see [3, 10, 11, 23, 24, 25, 26] for the recent progress.

The study of the above Bayesian estimation arises a challenging computational problem in searching for polynomial-time algorithms in simulating the MMSEE. To this end, AMP algorithms have been widely adapted [21, 28, 29, 30, 31, 34, 37] and known to achieve a good level of success; in some cases, it allows to obtain the Bayesian-optimal error estimates, see [11, 12, 29]. In addition to being useful in matrix estimations, AMP has also been applied to a number of randomized optimization problems in mean-field spin glass models in recent years. In particular, it was shown in [17, 27] that AMP allows to implement polynomial-time algorithms in the optimization of the SK Hamiltonian and its variants.

In these applications, the AMP algorithm is formulated as a sequence of n -dimensional vectors $(v^{[k]})_{k \geq 0}$ of the form

$$v^{[k+1]} = \hat{A}_n f_k(v^{[k]}, \dots, v^{[0]}) - \sum_{j=1}^k b_{k,j} f_{j-1}(v^{[j-1]}, \dots, v^{[0]}),$$

for

$$b_{k,j} = \frac{1}{n} \sum_{i=1}^n \frac{\partial f_k}{\partial v_i^{[j]}}(v_i^{[k]}, \dots, v_i^{[0]}),$$

where $f_k \in C^1(\mathbb{R}^{k+1})$ and the two vectors $f_k(v^{[k]}, \dots, v^{[0]})$ and $f_{j-1}(v^{[j-1]}, \dots, v^{[0]})$ above are defined coordinate-wise by $v^{[k]}, \dots, v^{[0]}$ and $v^{[j-1]}, \dots, v^{[0]}$, respectively. The key component here is the initialization $v^{[0]}$; it influences the convergence of the AMP in the large n limit.

When Z_n is a zero matrix and the initialization $v^{[0]}$ is independent of A_n , it was known [4, 5, 8, 19] that under mild assumptions on f_j 's, this iterative algorithm converges in the sense that for any Lipschitz function $\phi \in C(\mathbb{R}^{k+1})$, almost surely,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi(v_i^{[k]}, \dots, v_i^{[0]}) = \mathbb{E} \phi(V_k, \dots, V_0), \quad (1.2)$$

where (V_k, \dots, V_0) is a centered Gaussian random vector with covariance

$$\mathbb{E} V_{a+1} V_{b+1} = \mathbb{E} f_a(V_a, \dots, V_0) f_b(V_b, \dots, V_0), \quad \forall 0 \leq a, b \leq k-1.$$

When Z_n is of the form (1.1) and the spectrum of \hat{A}_n exhibit the so-called Baik-Ben Arous-Péché (BBP) phase transition [1], namely, the top eigenvalue of \hat{A}_n stays a gap away from the rest of the eigenvalues and the principal eigenvector is correlated to the prior, a recent paper [29] further showed that an analogous convergence remains valid when the AMP is initialized by the principal eigenvector, see Example 2.2 below. The typical way to use AMP is to select the functions ϕ, f_k, \dots, f_0 properly (usually are smooth and with bounded derivatives) so that the limit (1.2) converges to the desired quantities of interest by adjusting the number of iteration k , see, e.g., [15, 27, 29].

While the above convergences were known to be true when A_n is Gaussian, in this work we investigate their validity under general randomness. When the signal matrix is not presented, i.e., $Z_n \equiv 0$ (or equivalently, $\hat{A}_n = A_n$) and $u^{[0]}$ is independent of A_n , this question was answered earlier in the work [4], in which they showed that if the evolution functions f_k, \dots, f_0 of the AMP are polynomials, then the AMP converges to the same limit of (1.2) independent of the choice of

the randomness on A_n . In our setting, we consider a generalized AMP with Lipschitz evolution functions and let it iterate in the presence of the signal matrix. Our first main result validates the universality of the AMP. As a consequence, this implies that the universality established in [4] also holds for Lipschitz functions and under the presence of the signal matrix Z_n . (We note here that the work in [4] can be actually extended to Lipschitz evolution functions by [4, Proposition 6] and its proof, but only when $Z_n \equiv 0$.) Furthermore, we show that universality of AMP with spectral initialization remains valid when the system exhibits the BBP phase transition.

Our approach is based on a Gaussian interpolation argument. In doing so, the central ingredient relies on a novel control on the moments of the partial derivatives of the AMP orbit with respect to the entries of the noise matrix A_n . While our argument are formulated for the purpose of this paper, the same strategy is expected to be applicable in more general settings.

2 Main results

We begin with some notations. For any column vectors $u^0, u^1, \dots, u^k \in \mathbb{R}^n$ and a function $f : \mathbb{R}^{k+1} \rightarrow \mathbb{R}$, we define $f(u^k, u^{k-1}, \dots, u^0)$ as a column vector by

$$f(u^k, u^{k-1}, \dots, u^0)_i = f(u_i^k, u_i^{k-1}, \dots, u_i^0).$$

For $x, y \in \mathbb{R}^n$, set $\|x\|_2 = (\sum_{i=1}^n x_i^2)^{1/2}$ and $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$. Let $M_n(\mathbb{R})$ be the collection of all $n \times n$ real-valued symmetric matrices. For $X, Z \in M_n(\mathbb{R})$, denote

$$X_n = \frac{X}{\sqrt{n}}$$

and

$$\hat{X}_n = \frac{X}{\sqrt{n}} + \frac{Z}{n}.$$

The generalized approximate message passing is formulated as follows.

Definition 2.1. Let $u^{[0]} : M_n(\mathbb{R}) \rightarrow \mathbb{R}^n$ be a measurable function. For any $k \geq 0$, let $F_k \in C(\mathbb{R}^{k+1})$ be Lipschitz. The generalized AMP orbit corresponding to (X, Z) , $(F_k)_{k \geq 0}$, and $u^{[0]}$ is the sequence of vector-valued functions $u^{[k]} : M_n(\mathbb{R}) \rightarrow \mathbb{R}^n$ for $k \geq 0$ defined iteratively by

$$u^{[k+1]}(X) = F_k(\hat{X}_n u^{[k]}(X), u^{[k-1]}(X), u^{[k-2]}(X), \dots, u^{[0]}(X)).$$

We now specify the randomness on X , Z , and u^0 . Let $\sigma > 0$ be fixed. For any $n \geq 1$, let $u^0 = (u_i^0)_{i \in [n]}$ be an n -dimensional random vector and $Z = (z_{ii'})_{i, i' \in [n]}$ be an $n \times n$ random symmetric matrix. Assume that there exists a constant $C(\sigma) > 0$ such that

$$\sup_{n \geq 1} \left(\mathbb{E} \exp\left(\frac{\|u^0\|_2^2}{\sigma n}\right), \max_{i \in [n]} \mathbb{E} \exp\left(\frac{|u_i^0|}{\sigma}\right), \max_{i, i' \in [n]} \mathbb{E} \exp\left(\frac{|z_{ii'}|}{\sigma}\right) \right) \leq C(\sigma). \quad (2.1)$$

Suppose that $A = (a_{ii'})_{i, i' \in [n]}$ is an $n \times n$ random symmetric matrix, whose upper triangular entries are independent with zero mean and unit variance and are σ -subgaussian, i.e., $\mathbb{E} e^{\lambda a_{ii'}} \leq e^{\lambda^2 \sigma^2 / 2}$ for all $\lambda \in \mathbb{R}$.

We further assume that A is independent of u^0 and Z , but allow u^0 and Z to be dependent on each other. An important example of A is when the entries $a_{ii'}$'s are standard normal. In this case, we denote A by G and we call Definition 2.1 associated to $X = G$ a Gaussian AMP. Our main result shows that if we initialize $u^{[0]}(X) = u^0$, then the AMP corresponding to any A is essentially the same as the Gaussian AMP.

Theorem 2.1. Let $u^{[0]}(X) = u^0$. For any $k \geq 0$ and Lipschitz function ϕ on \mathbb{R}^{k+1} , we have that in probability

$$\lim_{n \rightarrow \infty} |\Phi_{k,n}(A) - \Phi_{k,n}(G)| = 0,$$

where

$$\Phi_{k,n}(X) := \frac{1}{n} \sum_{i=1}^n \phi(u_i^{[k]}(X), \dots, u_i^{[0]}(X)), \quad X \in M_n(\mathbb{R}). \quad (2.2)$$

Next we introduce the AMP used in the matrix estimation and some optimization problems in mean-field spin glasses.

Definition 2.2. Let $f_{-1} \equiv 0$. For $k \geq 0$, assume that $f_k \in C^1(\mathbb{R}^{k+1})$ is Lipschitz and its first-order partial derivatives are also Lipschitz. Let $X \in M_n(\mathbb{R})$. Starting from an initialization $v^{[0]}(X)$, define the AMP orbit for $k \geq 0$ iteratively by

$$v^{[k+1]}(X) = \hat{X}_n f_k(v^{[k]}(X), \dots, v^{[0]}(X)) - \sum_{j=1}^k b_{k,j}(X) f_{j-1}(v^{[j-1]}(X), \dots, v^{[0]}(X)), \quad (2.3)$$

where

$$b_{k,j}(X) = \frac{1}{n} \sum_{i=1}^n \frac{\partial f_k}{\partial v_i^{[j]}(X)}(v_i^{[k]}(X), \dots, v_i^{[0]}(X)).$$

Note that Definition 2.2 is not a direct example of the generalized AMP in Definition 2.1 due to the term $b_{k,j}(X)$. Nevertheless, since $b_{k,j}(X)$ is an average of the partial derivatives, this quantity is essentially indistinguishable between different randomness and this allows us to establish the following universality.

Theorem 2.2. Let $v^{[0]}(X) = u^0$. For any $k \geq 0$, if ϕ is Lipschitz on \mathbb{R}^{k+1} , then in probability,

$$\lim_{n \rightarrow \infty} |\phi_{k,n}(A) - \phi_{k,n}(G)| = 0,$$

where

$$\phi_{k,n}(X) := \frac{1}{n} \sum_{i=1}^n \phi(v_i^{[k]}(X), \dots, v_i^{[0]}(X)), \quad X \in M_n(\mathbb{R}).$$

Remark 1. As pointed out in the introduction, it was known [4, 5, 8, 19] that if $Z = 0$, the Gaussian AMP in Definition 2.2 converges, see (1.2). If f_0, \dots, f_k are polynomials and $Z = 0$, it was further understood in [4] that this convergence is independent of the choice of the randomness of A . Theorem 2.2 here extends this universality to Lipschitz functions and in the presence of Z . We refer the reader to check [17, 27] for the usage of this AMP in the optimization of the SK Hamiltonian and related models.

Example 2.1. For $\gamma \geq 0$, set $Z = \gamma u^0 \otimes u^0$. In this case, \hat{A}_n is a rank-one spiked matrix,

$$\hat{A}_n = \frac{A}{\sqrt{n}} + \gamma \frac{u^0 \otimes u^0}{n}.$$

In matrix estimation, one would like to recover the vector u^0 from the realization of \hat{A}_n . When $A = G$, the MMSEE, $\mathbb{E}[u^0 | \hat{A}_n]$, is popularly adapted for this purpose and it can be simulated via the AMP in Definition 2.2 with specifically chosen functions f_k 's, see, e.g., [11]. Theorem 2.2 here indicates that in a non-Gaussian noise environment, the AMP in Definition 2.2 still allows to implement the same simulation for u^0 as the Gaussian AMP.

Recall that the initialization u^0 and the signal matrix Z are assumed to be independent of the noise. In Example 2.1, since the MMSEE is a measurable function of the spiked matrix \hat{A}_n , it is often more desirable that the initialization depends on \hat{A}_n , as it should provide a better estimate for the MMSEE. When $A = G$, an attempt along this line has been successfully carried out in [29]. Our last main result addresses universality towards this direction. For any $X \in M_n(\mathbb{R})$, denote by $\lambda_1(\hat{X}_n) \geq \lambda_2(\hat{X}_n) \geq \dots \geq \lambda_n(\hat{X}_n)$ the eigenvalues of \hat{X}_n and by $\psi^1(\hat{X}_n)$ the top eigenvector of \hat{X}_n with $\|\psi^1(\hat{X}_n)\|_2 = \sqrt{n}$. Set

$$\psi(X) = \text{sign}(\langle \psi^1(\hat{X}_n), u^0 \rangle) \psi^1(\hat{X}_n) \quad (2.4)$$

whenever $\langle \psi^1(\hat{X}_n), u^0 \rangle \neq 0$. Note that although there are two possible choices of $\psi^1(\hat{X}_n)$ up to a sign, the definition $\psi(X)$ here is not influenced by this difference.

Theorem 2.3. *Assume that*

$$\begin{aligned} \liminf_{n \rightarrow \infty} \lambda_1(\hat{G}_n) &> \max\left(\limsup_{n \rightarrow \infty} \max_{2 \leq \ell \leq n} |\lambda_\ell(\hat{G}_n)|, 1\right), \\ \liminf_{n \rightarrow \infty} \lambda_1(\hat{A}_n) &> \max\left(\limsup_{n \rightarrow \infty} \max_{2 \leq \ell \leq n} |\lambda_\ell(\hat{A}_n)|, 1\right), \end{aligned} \quad (2.5)$$

and

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \min(|\langle \psi^1(\hat{G}_n), u^0 \rangle|, |\langle \psi^1(\hat{A}_n), u^0 \rangle|) > 0. \quad (2.6)$$

Consider the AMP orbit $(v^{[\ell]})_{0 \leq \ell \leq k}$ defined in Definition 2.2. Let $v^{[0]}(X) = \psi(X)$. If $\phi \in C(\mathbb{R}^{k+1})$ is Lipschitz, then in probability,

$$\lim_{n \rightarrow \infty} |\phi_{k,n}(A) - \phi_{k,n}(G)| = 0.$$

The assumptions (2.5) and (2.6) say that the top eigenvalue stays a gap away from the rest of the eigenvalues and the principal eigenvector is correlated to the prior vector u^0 . These behaviors are not only required in our proofs for technical purposes, but also appear to be quite typical in the BBP phase transition, see the following example.

Example 2.2. Recall \hat{A}_n from Example 2.1. Let $u^0 = (u_1^0, \dots, u_n^0)$ for $u_1^0, \dots, u_n^0 \stackrel{i.i.d.}{\sim} w$, where w is a centered random variable with compact support and unit variance. Recall from [6] that the BBP transition point is equal to 1: If $\gamma < 1$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \lambda_1(\hat{G}_n) &= 2, \text{ a.s.}, \\ \lim_{n \rightarrow \infty} \frac{1}{n} \langle \psi^1(\hat{G}_n), u^0 \rangle &= 0, \text{ a.s.}; \end{aligned} \quad (2.7)$$

if $\gamma > 1$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \lambda_1(\hat{G}_n) &= \gamma + \gamma^{-1} > 2, \text{ a.s.}, \\ \lim_{n \rightarrow \infty} \frac{1}{n} \langle \psi^1(\hat{G}_n), u^0 \rangle &= \sqrt{1 - \gamma^{-2}} > 0, \text{ a.s.} \end{aligned} \quad (2.8)$$

These imply that the spectral method can be used to gain useful information about u^0 only if the SNR exceeds the critical threshold, i.e., $\gamma > 1$, as in this case the principal eigenvector is positively correlated to u^0 . In [29], the convergence of AMP in Definition 2.1 initialized by the top eigenvector was investigated, which states that again when $\gamma > 1$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi(u_i^{[0]}(G), u_i^{[k]}(G)) = \mathbb{E} \phi(w, \mu_k w + \sigma_k g). \quad (2.9)$$

Here, starting from $\mu_0 = \sqrt{1 - \gamma^2}$ and $\sigma_0 = 1/\gamma$, $(\mu_k)_{k \geq 1}$ and $(\sigma_k)_{k \geq 1}$ are defined through

$$\begin{aligned}\mu_{k+1} &= \gamma \mathbb{E}[w f_k(\mu_k w + \sigma_k g)], \\ \sigma_{k+1}^2 &= \mathbb{E}[f_k(\mu_k w + \sigma_k g)^2],\end{aligned}$$

where $g \sim N(0, 1)$ is independent of w . Note that \hat{G}_n is a perturbation of G_n by a rank-one matrix. The eigenvalue interlacing property implies that for any small $\delta > 0$, asymptotically

$$-\sqrt{2} - \delta \leq \lambda_n(G_n) \leq \lambda_i(\hat{G}_n) \leq \lambda_1(G_n) \leq \sqrt{2} + \delta$$

for all $2 \leq i \leq n$, where $\lambda_1(G_n)$ and $\lambda_n(G_n)$ are the largest and smallest eigenvalues of G_n , respectively. Note that this inequality, (2.7), and (2.8) are also valid for A . Hence, the assumptions of (2.5) and (2.6) are valid and as a result, the convergence of (2.9) is universal in probability.

Our approach to proving Theorem 2.1 is to match the first and second moments of $\Phi_{k,n}$ between A and G , respectively. To this end, we define a Gaussian interpolation $X = A(t) := \sqrt{t}A + \sqrt{1-t}G$ for $0 \leq t \leq 1$ and control the t -derivatives of $\mathbb{E}\Phi_{k,n}(A(t))$ and $\mathbb{E}\Phi_{k,n}(A(t))^2$. The hope is that if the total number of the terms as well as their orders appearing in these derivatives are small enough, then we anticipate that an application of the approximate Gaussian integration by parts would make the derivatives small. However, due to the iteration of the AMP, these derivatives involve highly complicated Hadamard products of a large number of column vectors in terms of the higher order partial derivatives of $u^{[\ell]}(X)$ and $\hat{X}_n u^{[\ell]}(X)$. As a result, the control of their p -th moments are extremely delicate, especially for those of $\hat{X}_n u^{[\ell]}(X)$. The novelty of our analysis adapts a Taylor expansion of the derivatives up to the p -th order, which allows us to extract the dependence of the i -th row of X out of the derivatives. This combining with a subtle moment computation in this expansion perfectly cancels out the majority of the smaller order terms and yields the following moment controls (see Proposition 5.2 and Lemma 5.3) that for any $p \geq 1$, there exists a universal constant $C > 0$ such that for any collection P of variables $x_{i'}$ for $i, i' \in [n]$ counting multiplicities with $|P| = m$, we have

$$\begin{aligned}\sup_{n \geq 2} \sup_{i \in [n]} \left(\mathbb{E} |\partial_P u^{[k]}(A)_i|^p \right)^{1/p} &\leq \frac{C}{n^{m/2}}, \\ \sup_{n \geq 2} \sup_{i \in [n]} \left(\mathbb{E} |\partial_P (\hat{A} u^{[k]}(A))_i|^p \right)^{1/p} &\leq \frac{C}{n^{m/2}},\end{aligned}$$

where ∂_P is the partial derivatives with respect to the variables in P . Using the Markov inequality and the union bound, these yield a uniform control on the derivatives that for any P with $|P| = m$ and $\delta > 0$, with probability at least $1 - Cn^{-\delta}$,

$$\begin{aligned}\max_{i \in [n]} |\partial_P u^{[k]}(A)_i| &\leq \frac{1}{n^{\frac{m}{2} - \delta - \frac{1}{p}}}, \\ \max_{i \in [n]} |\partial_P (\hat{A} u^{[k]}(A))_i| &\leq \frac{1}{n^{\frac{m}{2} - \delta - \frac{1}{p}}}.\end{aligned}$$

Once Theorem 2.1 is established, the proof of Theorem 2.2 follows essentially by a special choice of the functions F_0, \dots, F_k, \dots . Although the term $b_{k,j}(X)$ in (2.3) relies on all coordinates, its form averages out the partial derivatives and consequently, $b_{k,j}(A)$ and $b_{k,j}(G)$ are asymptotically equal in probability, which is already enough to establish Theorem 2.2 following an induction argument. Lastly to show Theorem 2.3, recall that while both AMP's in Theorems 2.2 and 2.3 share the same iteration procedure, their initializations are of different kind; the former is initialized independent of A_n , but the latter adapts the principal eigenvector of \hat{A}_n . We show that this eigenvector can

be approximated very well by the power method (see Lemma 8.1). In view of this method, it is essentially a special case of our generalized AMP with the choice $F_k(x_k, \dots, x_0) = \hat{X}_n x_k$ and an analogous argument as that for Theorem 2.2 allows to establish Theorem 2.3. One technicality here is that in order to guarantee the convergence of the power method, one would have to choose the initialization carefully and ensure that the principle eigenvalue of \hat{A}_n is well-separated from the other eigenvalues. This explains why the assumptions (2.5) and (2.6) need to be in position.

We mention that many works in the literature, e.g., [19], also established high-dimensional version of the AMP, in which the functions f_k are allowed to be of vector-valued. In addition, it was also discussed in [29] that one can initialize the AMP in Definition 2.2 via other leading eigenvectors, whose corresponding eigenvalues exhibit the BBP phase transition. In view of the present approach, it is plausible that universality under these settings can be obtained from our results by a similar line of derivation. We do not address these directions here.

The rest of the paper is organized as follows. Sections 3-6 are the preparation for the proof of Theorem 2.1. Section 3 establishes a Gaussian concentration inequality for the function $\Phi_{k,n}(X)$ as well as a number of prior controls on the AMP orbit. In Section 4, we show that in proving Theorem 2.1, it suffices to assume that ϕ and F_k 's are smooth with uniformly bounded derivatives. Section 5 provides the main estimates on the moments of the derivatives of the AMP orbits. In Section 6, we carry out our interpolation argument and present the proof of Theorem 2.1. The proofs of Theorems 2.2 and 2.3 are provided in Sections 7 and 8, respectively. Finally, the Appendix gathers error estimates of some approximate Gaussian integration by parts.

3 Lipschitz property and concentration inequality

Consider the AMP in Definition 2.1 with initialization $u^{[0]}(X) = u^0$. In this section, we establish a Lipschitz property for this AMP and a concentration inequality for $\Phi_{k,n}(G)$. These will be used later in the proof of Theorem 2.1. Recall that the functions F_k in Definition 2.1 are Lipschitz. Let η_k be the Lipschitz constant of F_k . For any $X \in M_n(\mathbb{R})$, denote by $\|X\|_2$ the $\ell_2 - \ell_2$ operator norm of X .

Proposition 3.1. *If $\phi \in C(\mathbb{R}^{k+1})$ is Lipschitz with Lipschitz constant $\eta > 0$, then we have that*

$$\begin{aligned} |\Phi_{k,n}(X) - \Phi_{k,n}(Y)| &\leq k\eta \|X_n - Y_n\|_2 \left(\frac{\|u^{[0]}(X)\|_2}{\sqrt{n}} + 1 \right) \sum_{\ell=1}^k \Theta_\ell(Y) \Delta_\ell(X) \\ &\quad + \frac{k\eta \|u^{[0]}(X) - u^{[0]}(Y)\|_2}{\sqrt{n}} \sum_{\ell=1}^k \Theta_\ell(Y), \end{aligned}$$

where

$$\Delta_k(X) := 2^k \left((\eta_0 + \dots + \eta_{k-1})(\|\hat{X}_n\|_2 + 1) + |F_0(0)| + \dots + |F_{k-1}(0)| + 1 \right)^k \quad (3.1)$$

and

$$\Theta_k(Y) := ((1 + \|\hat{Y}_n\|_2) \max(\eta_0, \dots, \eta_{k-1}) + 2)^k. \quad (3.2)$$

This proposition says that the AMP orbits behave stably subject to a small perturbation to the matrix X and the initialization. From this, we show that the Gaussian AMP is concentrated.

Theorem 3.1. *Let $u^{[0]}(X) = u^0$. For any $k \geq 0$, if $\phi \in C(\mathbb{R}^{k+1})$ is Lipschitz, then*

$$\lim_{n \rightarrow \infty} \mathbb{E} |\Phi_{k,n}(G) - \tilde{\mathbb{E}} \Phi_{k,n}(G)|^2 = 0,$$

where $\tilde{\mathbb{E}}$ is the expectation conditionally on u^0 and Z .

For the rest of this section, we establish these results.

3.1 Proof of Proposition 3.1

The proof of this proposition relies on two lemmas on the boundedness and the Lipschitz property of the vector $u^{[k]}(X)$ following an iterative argument.

Lemma 3.1. *For every $k \geq 1$,*

$$\|u^{[k]}(X)\|_2 \leq \Delta_k(X)(\|u^{[0]}(X)\|_2 + \sqrt{n}), \quad (3.3)$$

where $\Delta_k(X)$ is defined in (3.1).

Proof. Write

$$\|u^{[\ell]}(X)\|_2^2 = \sum_{i=1}^n F_{\ell-1}(\hat{X}_n u^{[\ell-1]}(X), u^{[\ell-2]}(X), \dots, u^{[0]}(X))_i^2.$$

Using the Lipschitz property of $F_{\ell-1}$ and the trivial bound $(a+b)^2 \leq 4(a^2 + b^2)$ yields

$$\begin{aligned} \|u^{[\ell]}(X)\|_2^2 &\leq \sum_{i=1}^n \left(\eta_{\ell-1} \left(|\hat{X}_n u^{[r]}(X)_i|^2 + \sum_{r=0}^{\ell-2} |u_i^{[r]}(X)|^2 \right)^{1/2} + |F_{\ell-1}(0)| \right)^2 \\ &\leq 4 \sum_{i=1}^n \left(\eta_{\ell-1}^2 |\hat{X}_n u^{[\ell-1]}(X)_i|^2 + \eta_{\ell-1}^2 \sum_{r=0}^{\ell-2} |u_i^{[r]}(X)|^2 + F_{\ell-1}(0)^2 \right) \\ &= 4\eta_{\ell-1}^2 \|\hat{X}_n u^{[\ell-1]}(X)\|_2^2 + 4\eta_{\ell-1}^2 \sum_{r=0}^{\ell-2} \|u^{[r]}(X)\|_2^2 + 4nF_{\ell-1}(0)^2 \end{aligned}$$

so that from the Minkowski inequality,

$$\begin{aligned} \|u^{[\ell]}(X)\|_2 &\leq 2\eta_{\ell-1} \|\hat{X}_n\|_2 \|u^{[\ell-1]}(X)\|_2 + 2\eta_{\ell-1} \sum_{r=0}^{\ell-2} \|u^{[r]}(X)\|_2 + 2n^{1/2} |F_{\ell-1}(0)| \\ &\leq C \left(\sum_{r=0}^{\ell-1} \|u^{[r]}(X)\|_2 + n^{1/2} \right), \end{aligned}$$

where

$$C := 2(\eta_0 + \dots + \eta_{k-1})(\|\hat{X}_n\|_2 + 1) + 2(|F_0(0)| + \dots + |F_{k-1}(0)|).$$

If we let $t_\ell := \|u^{[\ell]}(X)\|_2 + n^{1/2}$ and $C' := 1 + C$, then the above inequality implies that

$$t_\ell \leq C' \sum_{r=0}^{\ell-1} t_r, \quad \forall 1 \leq \ell \leq k.$$

Using induction yields that

$$t_\ell \leq C'(1 + C')^{\ell-1} t_0, \quad \forall 1 \leq \ell \leq k,$$

which implies that for $\Delta_k(X) := (1 + C')^k = (2 + C)^k$,

$$\|u^{[k]}(X)\|_2 \leq C'(1 + C')^{k-1} (\|u^{[0]}(X)\|_2 + \sqrt{n}) \leq \Delta_k(X)(\|u^{[0]}(X)\|_2 + \sqrt{n})$$

and this completes our proof. \square

Lemma 3.2. For any $X, Y \in M_n(\mathbb{R})$,

$$\begin{aligned} \|u^{[k]}(X) - u^{[k]}(Y)\|_2 &\leq \Theta_k(Y) \Delta_k(X) \|X_n - Y_n\|_2 (\|u^{[0]}(X)\|_2 + \sqrt{n}) \\ &\quad + \Theta_k(Y) \|u^{[0]}(X) - u^{[0]}(Y)\|_2, \end{aligned}$$

where $\Delta_k(X)$ is defined in (3.2).

Proof. From the Lipschitz property of $F_{\ell-1}$,

$$\|u^{[\ell]}(X) - u^{[\ell]}(Y)\|_2 \leq \eta_{\ell-1} \left(\|\hat{X}_n u^{[\ell-1]}(X) - \hat{Y}_n u^{[\ell-1]}(Y)\|_2 + \sum_{r=0}^{\ell-2} \|u^{[r]}(X) - u^{[r]}(Y)\|_2 \right).$$

Here, for any $1 \leq \ell \leq k$,

$$\begin{aligned} \|\hat{X}_n u^{[\ell-1]}(X) - \hat{Y}_n u^{[\ell-1]}(Y)\|_2 &\leq \|u^{[\ell-1]}(X)\|_2 \|X_n - Y_n\|_2 \\ &\quad + \|\hat{Y}_n\|_2 \|u^{[\ell-1]}(X) - u^{[\ell-1]}(Y)\|_2. \end{aligned}$$

If we let

$$\begin{aligned} C &= (1 + \|\hat{Y}_n\|_2) \max(\eta_0, \dots, \eta_{k-1}), \\ D &= \|X_n - Y_n\|_2 \max(\|u^{[0]}(X)\|_2, \dots, \|u^{[k-1]}(X)\|_2), \end{aligned}$$

then

$$\|u^{[\ell]}(X) - u^{[\ell]}(Y)\|_2 \leq C \left(D + \sum_{r=0}^{\ell-1} \|u^{[r]}(X) - u^{[r]}(Y)\|_2 \right).$$

If we let

$$t_\ell := D + \|u^{[\ell]}(X) - u^{[\ell]}(Y)\|_2,$$

then for $C' := C + 1$,

$$t_\ell \leq C' \sum_{r=0}^{\ell-1} t_r, \quad 1 \leq \ell \leq k$$

and by induction, $t_\ell \leq C'(1 + C')^{\ell-1} t_0$. Consequently, for $\Theta_k(Y) := (1 + C')^k$,

$$\|u^{[k]}(X) - u^{[k]}(Y)\|_2 \leq C'(1 + C')^{k-1} t_0 \leq \Theta_k(Y) t_0$$

and this completes our proof by noting that

$$D \leq \|X_n - Y_n\|_2 (\|u^{[0]}(X)\|_2 + \sqrt{n}) \Delta_{k-1}(X) \leq \|X_n - Y_n\|_2 (\|u^{[0]}(X)\|_2 + \sqrt{n}) \Delta_k(X).$$

□

Proof of Proposition 3.1. From the Lipschitz property of ϕ and Lemmas 3.1 and 3.2, our assertion follows immediately. □

3.2 Some prior bounds

Lemma 3.3. *For any integer $p \geq 1$, we have that*

$$\sup_{n \geq 1} \left(\frac{\mathbb{E} \|u^0\|_2^p}{n^{p/2}}, \frac{\mathbb{E} \|Z\|_2^p}{n^p}, \mathbb{E} \|A_n\|_2^p, \mathbb{E} \|\hat{A}_n\|_2^p \right) < \infty.$$

Proof. Note that $x^{2p} \leq p!e^{x^2}$ and $|x|^p \leq p!e^{|x|}$. The inequality (2.1) implies that

$$\sup_{n \geq 1} \frac{\mathbb{E} \|u^0\|_2^p}{n^{p/2}} \leq \sup_{n \geq 1} \left(\frac{\mathbb{E} \|u^0\|_2^{2p}}{n^p} \right)^{1/2} \leq \sup_{n \geq 1} \sqrt{p! \sigma^p \mathbb{E} e^{\|u^0\|_2^2 / \sigma n}} < \sqrt{p! \sigma^p C(\sigma)}.$$

Next, note that the operator norm is no larger than the Frobenius norm. This and the Jensen inequality lead to

$$\begin{aligned} \sup_{n \geq 1} \frac{\mathbb{E} \|Z\|_2^p}{n^p} &\leq \sup_{n \geq 1} \mathbb{E} \left(\frac{\sum_{i,i'=1}^n |z_{ii'}|^2}{n^2} \right)^{p/2} \leq \sup_{n \geq 1} \frac{\mathbb{E} \sum_{i,i'=1}^n |z_{ii'}|^p}{n^2} \\ &\leq \sup_{n \geq 1} \frac{p! \sigma^p}{n^2} \sum_{i,i'=1}^n \mathbb{E} e^{|z_{ii'}|/\sigma} \leq p! \sigma^p C(\sigma). \end{aligned}$$

Finally, since the entries of A are independent σ -subgaussian with zero mean, it is well-known (see, for instance, Corollary 4.4.8 in [36]) that $\sup_{n \geq 1} \mathbb{E} \|A_n\|_2^p < \infty$. Putting these bounds together yields the uniform integrability of $\|\hat{A}_n\|_2^p$ and this completes our proof. \square

Lemma 3.4. *For any $k \geq 0$, we have that*

$$\sup_{n \geq 1} \mathbb{E} \left(\frac{\|u^{[k]}(A)\|_2}{\sqrt{n}} \right)^4 < \infty \quad (3.4)$$

and if $\phi \in C(\mathbb{R}^{k+1})$ is Lipschitz,

$$\sup_{n \geq 1} \mathbb{E} |\Phi_{k,n}(A)|^4 < \infty. \quad (3.5)$$

Proof. The proof follows directly from Lemmas 3.1 and 3.3. \square

3.3 Proof of Theorem 3.1

First of all, we establish a Gaussian concentration inequality for the functional $\Phi_{k,n}$.

Lemma 3.5. *Let $u^{[0]}(X) = u^0$. For any $k \geq 0$, if $\phi \in C(\mathbb{R}^{k+1})$ is Lipschitz, then there exists a constant $c > 0$ such that for every $t > 0$,*

$$\tilde{\mathbb{P}}(|\Phi_{k,n}(G) - \tilde{\mathbb{E}}[\Phi_{k,n}(G)]| \geq t - c(\Omega_n + 1)e^{-n/c}) \leq ce^{-nt^2/(c\Omega_n^2)} + ce^{-n/c},$$

where $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{E}}$ are the probability and expectation conditionally on u^0 and Z , and

$$\Omega_n := \left(1 + \frac{\|u^0\|_2}{\sqrt{n}}\right) \left(1 + \frac{\|Z\|_2}{n}\right)^k. \quad (3.6)$$

Proof. From $u^{[0]}(X) = u^{[0]}(Y) = u^0$ and $\|\hat{X}_n\|_2 \leq \|X_n\|_2 + \|Z\|_2/n$, Proposition 3.1 implies

$$|\Phi_{k,n}(X) - \Phi_{k,n}(Y)| \leq c_0 \Omega_n \|X_n - Y_n\|_2 \sum_{\ell=1}^k \left(1 + \|X_n\|_2\right)^\ell \left(1 + \|Y_n\|_2\right)^\ell,$$

where Ω_n is defined in (3.6) and c_0 is a constant independent of n . Observe that for any $M > 0$, if $\|X_n\|_2, \|Y_n\|_2 \leq M$, then

$$\Phi_{k,n}(X) \leq \Phi_{k,n}(Y) + c_0 \Omega_n (\|X_n - Y_n\|_2 \wedge (2M)) \sum_{\ell=1}^k (1 + \|X_n\|_2 \wedge M)^\ell (1 + \|Y_n\|_2)^\ell.$$

This implies that if

$$T(X) := \inf_{Y \in M_n(\mathbb{R}) : \|Y_n\|_2 \leq M} \left(\Phi_{k,n}(Y) + c_0 \Omega_n (\|X_n - Y_n\|_2 \wedge (2M)) \cdot \sum_{\ell=1}^k (1 + \|X_n\|_2 \wedge M)^\ell (1 + \|Y_n\|_2)^\ell \right),$$

then $T(X) \geq \Phi_{k,n}(X)$ if $\|X_n\|_2 \leq M$ and consequently, $T(X) = \Phi_{k,n}(X)$ if $\|X_n\|_2 \leq M$. Next, note that for any $Y_n \in M_n(\mathbb{R})$ with $\|Y_n\|_2 \leq M$ and $X, X' \in M_n(\mathbb{R})$,

$$\begin{aligned} \|X_n - Y_n\|_2 \wedge (2M) &\leq (\|X_n - X'_n\|_2 + \|X'_n - Y_n\|_2) \wedge (2M) \\ &\leq \|X_n - X'_n\|_2 \wedge (2M) + \|X'_n - Y_n\|_2 \wedge (2M) \\ &\leq \|X_n - X'_n\|_2 + \|X'_n - Y_n\|_2 \wedge (2M) \end{aligned}$$

and

$$\begin{aligned} (1 + \|X_n\|_2 \wedge M)^\ell &\leq (1 + \|X_n - X'_n\|_2 \wedge M + \|X'_n\|_2 \wedge M)^\ell \\ &= (1 + \|X'_n\|_2 \wedge M)^\ell + \sum_{a=1}^{\ell} \binom{\ell}{a} (\|X_n - X'_n\|_2 \wedge M)^a (1 + \|X'_n\|_2 \wedge M)^{\ell-a} \\ &\leq (1 + \|X'_n\|_2 \wedge M)^\ell + \sum_{a=1}^{\ell} \binom{\ell}{a} \|X_n - X'_n\|_2 M^{a-1} (1 + M)^{\ell-a} \\ &= (1 + \|X'_n\|_2 \wedge M)^\ell + \|X_n - X'_n\|_2 \sum_{a=1}^{\ell} \binom{\ell}{a} M^{a-1} (1 + M)^{\ell-a}. \end{aligned}$$

From these and noting that the ℓ_2 -operator norm of a matrix is less than its Frobenius norm, we see that $T(X)$ is Lipschitz with respect to the Frobenius norm with Lipschitz constant $c_1 \Omega_n / n^{1/2}$ for some constant c_1 independent of n . Hence, the usual Gaussian concentration inequality for Lipschitz functions implies that

$$\tilde{\mathbb{P}}(|T(G) - \tilde{\mathbb{E}}[T(G)]| \geq t) \leq 2e^{-nt^2/(4c_1^2 \Omega_n^2)}, \quad \forall t > 0.$$

Now note that as long as we fix M large enough at the beginning,

$$\begin{aligned} \tilde{\mathbb{P}}(|T(G) - \tilde{\mathbb{E}}T(G)| \geq t) &\geq \tilde{\mathbb{P}}(|\Phi_{k,n}(G) - \tilde{\mathbb{E}}T(G)| \geq t, \|G_n\|_2 \leq M) \\ &\geq \tilde{\mathbb{P}}(|\Phi_{k,n}(G) - \tilde{\mathbb{E}}T(G)| \geq t) - \mathbb{P}(\|G_n\|_2 \geq M) \\ &\geq \tilde{\mathbb{P}}(|\Phi_{k,n}(G) - \tilde{\mathbb{E}}T(G)| \geq t) - c_2 e^{-n(M-c_3)^2/c_2}, \end{aligned}$$

where the last inequality used the well-known bound that the largest eigenvalue of G is concentrated around its mean with exponential tail bound, which follows by the Borell-TIS inequality and c_2, c_3 are two constants independent of n and M . On the other hand,

$$\begin{aligned}\tilde{\mathbb{E}}T(G) &= \tilde{\mathbb{E}}[\Phi_{k,n}(G); \|G_n\|_2 \leq M] + \tilde{\mathbb{E}}[T(G); \|G_n\|_2 \geq M] \\ &= \tilde{\mathbb{E}}[\Phi_{k,n}(G)] + \tilde{\mathbb{E}}[-\Phi_{k,n}(G) + T(G); \|G_n\|_2 \geq M].\end{aligned}$$

Here,

$$\begin{aligned}|\tilde{\mathbb{E}}[-\Phi_{k,n}(G) + T(G); \|G_n\|_2 \geq M]| &\leq (\tilde{\mathbb{E}}(|\Phi_{k,n}(G)| + |T(G)|)^2)^{1/2} \mathbb{P}(\|G_n\|_2 \geq M)^{1/2} \\ &\leq c_4(\Omega_n + 1)e^{-n(M-c_3)^2/2c_2}\end{aligned}$$

for some c_4 independent of n and M . From these,

$$\begin{aligned}\tilde{\mathbb{P}}(|\Phi_{k,n}(G) - \tilde{\mathbb{E}}[\Phi_{k,n}(G)]| \geq t - c_4(\Omega_n + 1)e^{-n(M-c_3)^2/2c_2}) \\ \leq 2e^{-nt^2/(4c_1^2\Omega_n^2)} + c_2e^{-n(M-c_3)^2/c_2}.\end{aligned}$$

This completes our proof. \square

Proof of Theorem 3.1. From Lemmas 3.3 and 3.5 and the Markov inequality, in probability \mathbb{P} ,

$$\lim_{n \rightarrow \infty} |\Phi_{k,n}(G) - \tilde{\mathbb{E}}[\Phi_{k,n}(G)]| = 0.$$

In addition, from

$$(\mathbb{E}|\Phi_{k,n}(G) - \tilde{\mathbb{E}}[\Phi_{k,n}(G)]|^4)^{1/4} \leq (\mathbb{E}\Phi_{k,n}(G)^4)^{1/4} + (\mathbb{E}(\tilde{\mathbb{E}}[\Phi_{k,n}(G)])^4)^{1/4} \leq 2(\mathbb{E}\Phi_{k,n}(G)^4)^{1/4},$$

the uniform upper bound (3.5) gives

$$\sup_{n \geq 1} (\mathbb{E}|\Phi_{k,n}(G) - \tilde{\mathbb{E}}[\Phi_{k,n}(G)]|^4)^{1/4} < \infty.$$

Hence, the assertion follows. \square

4 Smooth approximation

Recall that the functions F_k in Definition 2.1 and ϕ in Theorem 2.1 are Lipschitz. In this section, we show that to prove Theorem 2.1, it suffices to assume that these functions are smooth and their derivatives of any nonzero orders are uniformly bounded.

Proposition 4.1. *For any $k \geq 0$ and $\varepsilon > 0$, there exist a constant C independent of n and some functions $\bar{\phi} \in C^\infty(\mathbb{R}^{k+1})$ and $\bar{F}_\ell \in C^\infty(\mathbb{R}^{\ell+1})$ for $0 \leq \ell \leq k-1$, whose partial derivatives of any nonzero orders are uniformly bounded such that*

$$|\Phi_{k,n}(X) - \bar{\Phi}_{k,n}(X)| \leq \varepsilon C \sum_{\ell=0}^{k-1} \|\hat{X}_n\|_2^\ell, \quad (4.1)$$

where

$$\bar{\Phi}_{k,n}(X) = \frac{1}{n} \sum_{i=1}^n \bar{\phi}(\bar{u}_i^{[k]}(X), \bar{u}_i^{[k-1]}(X), \dots, \bar{u}_i^{[0]}(X))$$

and $\bar{u}^{[k]}$ is the k -th AMP orbit in Definition 2.1 associated to the functions $\bar{F}_0, \dots, \bar{F}_{k-1}$ and the initial condition $\bar{u}^{[0]}(X) = u^{[0]}(X)$.

Proof. Denote by η_ℓ the Lipschitz constant of F_ℓ . Let $\varepsilon > 0$ be fixed. Assume that $\zeta_\ell \in C^\infty(\mathbb{R}^{\ell+1})$ is a mollifier with $\zeta_\ell \geq 0$ and $\int \zeta_\ell dx = 1$ and it is supported on the unit ball $\{x \in \mathbb{R}^{\ell+1} : \|x\|_2 \leq 1\}$. Define $\zeta_{\ell,\varepsilon}(x) = \varepsilon^{-(\ell+1)}\zeta_\ell(x/\varepsilon)$. Set

$$\bar{F}_{\ell,\varepsilon}(x) = F_\ell * \zeta_{\ell,\varepsilon}(x) = \int \zeta_{\ell,\varepsilon}(x-y)F_\ell(y)dy.$$

Note that for any $\varepsilon > 0$ and $x \in \mathbb{R}^{\ell+1}$,

$$|\bar{F}_\ell(x) - F_\ell(x)| = \left| \int \zeta_\ell(z)(F_\ell(x - \varepsilon z) - F_\ell(x))dz \right| \leq \eta_\ell \varepsilon \int \|z\|_2 \zeta_\ell(z)dz \leq \eta'_\ell \varepsilon$$

for some constant $\eta'_\ell > 0$. In addition, for any index $\alpha = (\alpha_\ell, \dots, \alpha_0) \in (\{0\} \cup \mathbb{N})^{\ell+1}$ with $|\alpha| := \sum_{r=0}^\ell \alpha_r \geq 1$, if $\alpha_{r_0} \geq 1$ for some $0 \leq r_0 \leq \ell$, then

$$\partial^\alpha \bar{F}_\ell(x) = \frac{1}{\varepsilon^{\ell+|\alpha|}} \int \partial^{\alpha'} \zeta_\ell\left(\frac{x-y}{\varepsilon}\right) \partial_{y_{r_0}} F_\ell(y) dy = \varepsilon^{1-|\alpha|} \int \partial^{\alpha'} \zeta_\ell(z) \partial_{y_{r_0}} F_\ell(x - \varepsilon z) dz,$$

where

$$\alpha' := (\alpha_\ell, \dots, \alpha_{r_0+1}, \alpha_{r_0} - 1, \alpha_{r_0-1}, \dots, \alpha_0).$$

Since F_ℓ is Lipschitz and ζ_ℓ is supported on the unit ball, it follows that the partial derivatives of all nonzero orders of \bar{F}_ℓ are uniformly bounded. In particular, $\sup_x \|\nabla \bar{F}_\ell(x)\|_2 \leq \eta''_\ell$, independent of ε . Let $\eta = \max_{1 \leq j \leq \ell} \{\eta'_j, \eta''_j\}$. To show (4.1), note that

$$\begin{aligned} \|u^{[\ell+1]} - \bar{u}^{[\ell+1]}\|_2 &\leq \|F_\ell(\hat{X}_n u^{[\ell]}, u^{[\ell-1]}, \dots, u^{[0]}) - \bar{F}_\ell(\hat{X}_n \bar{u}^{[\ell]}, \bar{u}^{[\ell-1]}, \dots, \bar{u}^{[0]})\|_2 \\ &\leq \|F_\ell(\hat{X}_n u^{[\ell]}, u^{[\ell-1]}, \dots, u^{[0]}) - \bar{F}_\ell(\hat{X}_n u^{[\ell]}, u^{[\ell-1]}, \dots, u^{[0]})\|_2 \\ &\quad + \|\bar{F}_\ell(\hat{X}_n u^{[\ell]}, u^{[\ell-1]}, \dots, u^{[0]}) - \bar{F}_\ell(\hat{X}_n \bar{u}^{[\ell]}, \bar{u}^{[\ell-1]}, \dots, \bar{u}^{[0]})\|_2 \\ &\leq \eta \varepsilon + \eta \left(\|\hat{X}_n\|_2 \|u^{[\ell]} - \bar{u}^{[\ell]}\|_2 + \sum_{r=0}^{\ell-1} \|u^{[r]} - \bar{u}^{[r]}\| \right). \end{aligned}$$

Since $\bar{u}^{[0]} = u^{[0]}$, an induction argument implies that

$$\|u^{[\ell]}(X) - \bar{u}^{[\ell]}(X)\|_2 \leq \varepsilon C \sum_{j=0}^{\ell-1} \|\hat{X}_n\|_2^j, \quad (4.2)$$

where C is a constant depending only on ℓ and η . Finally, by the same argument, for any $\varepsilon > 0$, there exists a $\bar{\phi} \in C^\infty(\mathbb{R}^{k+1})$ with uniformly bounded partial derivatives of any nonzero orders such that $\|\phi - \bar{\phi}\|_\infty < \varepsilon$. From (4.2) and the Lipschitz property of ϕ , our proof is completed. \square

5 Bounding the derivatives

We establish uniform moment controls on the partial derivatives of the generalized AMP orbit in Definition 2.1. For $\sigma > 0$, recall the random vector u^0 and the random matrix Z from (2.1). Let $\mathcal{G}_n(\sigma)$ be the collection of $n \times n$ symmetric random matrices $A = (a_{ii'})_{i,i' \in [n]}$, whose entries are centered independent and each of them are σ' -subgaussian for some $0 \leq \sigma' \leq \sigma$. We also assume that $\mathcal{G}_n(\sigma)$ is independent of u^0 and Z .

For any $m \in \mathbb{N}$, denote by $[m] = \{1, \dots, m\}$. For any $n \geq 2$, let \mathcal{T}_n be the collection of all sequences $(i_r, i'_r)_{r \geq 1} \subset [n]^2$ with $i_r < i'_r$ for all $r \geq 1$. Let P be an arbitrary finite subset of \mathbb{N} and let $m = |P|$. For $h \in C^m(M_n(\mathbb{R}))$ and $(i_r, i'_r)_{r \geq 1} \in \mathcal{T}_n$, denote by

$$\partial_P h(X) \in \mathbb{R}$$

the partial derivative of h with respect to the variables x_{i_r, i'_r} for all $r \in P$ counting multiplicities. For a vector-valued function $H = (h_1, \dots, h_n)$ for $h_1, \dots, h_n \in C^m(M_n(\mathbb{R}))$, we also set the partial derivative of H by

$$\partial_P H(X) = (\partial_P h_1(X), \dots, \partial_P h_n(X)) \in \mathbb{R}^n$$

and denote

$$\partial_P H(X)_i = \partial_P h_i(X), \quad 1 \leq i \leq n.$$

For any $n \geq 2$, $k \geq 0$, $p \geq 1$, and $m \geq 0$, denote by $\mathcal{B}_n(k, p, m)$ the collection of all

$$(P, (i_r, i'_r)_{r \geq 1}, A, i)$$

for $P \subset \mathbb{N}$ with $|P| = m$, $(i_r, i'_r) \in \mathcal{T}_n$, $A \in \mathcal{G}_n(\sigma)$, and $i \in [n]$. The following is our main estimate.

Proposition 5.1. *Consider the AMP orbit $(u^{[k]}(X))_{k \geq 0}$ in Definition 2.1 with $u^{[0]}(X) = u^0$. Assume that the functions F_k in Definition 2.1 satisfy the following assumption:*

$$F_k \in C^\infty(\mathbb{R}^{k+1}) \text{ and its partial derivatives of all nonzero orders are uniformly bounded.} \quad (5.1)$$

Let $k \geq 0$, $p \geq 1$, and $m \geq 0$. Let $U \in C^\infty(\mathbb{R}^{2(k+1)})$. Assume that its partial derivatives of nonzero orders are uniformly bounded. Define a vector-valued random function on $M_n(\mathbb{R})$ by

$$U(X) = U(\hat{X}_n u^{[k]}(X), \dots, \hat{X}_n u^{[0]}(X), u^{[k]}(X), \dots, u^{[0]}(X)) \in \mathbb{R}^n.$$

There exists a universal constant $\Gamma_{k,p,m}^U$ such that

$$\sup_{\mathcal{B}_n(k,p,m)} (\mathbb{E} |\partial_P U(A)_i|^p)^{1/p} \leq \frac{\Gamma_{k,p,m}^U}{n^{m/2}}, \quad \forall n \geq 2.$$

As we shall see, this bound will be used to control the Gaussian interpolation between the first two moments of $\Phi_{k,n}(A)$ and $\Phi_{k,n}(G)$. For the rest of this section, we establish this proposition in three subsections. First of all, we derive explicit formulas for the derivatives of the AMP orbit. Next, we show that Proposition 5.1 is valid if U depends only on the marginal variables. The general case is treated in the last subsection.

5.1 Some auxiliary lemmas

Let $k \geq 0$ and $m \geq 1$ be fixed. Let $v_0, \dots, v_k \in C^m(M_n(\mathbb{R}))$ and $F \in C^m(\mathbb{R}^{k+1})$. Set

$$V(X) = F(v_k(X), \dots, v_0(X)) \in \mathbb{R}, \quad \forall X \in M_n(\mathbb{R}).$$

Let $(i_r, i'_r)_{r \geq 1} \in \mathcal{T}_n$. Let P be a finite subset of \mathbb{N} with $|P| = m$. For any $1 \leq r \leq m$, set $\mathcal{J}_r(k) = \{0, \dots, k\}^r$ and set $\mathcal{P}_r(P)$ the collection of all partitions $\mathcal{P} = \{P_1, \dots, P_r\}$ of P into r nonempty subsets. For $J = (j_1, \dots, j_r) \in \mathcal{J}_r(k)$, set

$$\partial_J F(y_k, \dots, y_0) = \partial_{y_{j_r} \dots y_{j_1}} F(y_k, \dots, y_0).$$

Lemma 5.1. *We have that*

$$\partial_P V = \sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_r(k)} \sum_{\mathcal{P} \in \mathcal{P}_r(P)} \partial_J F(v_k, \dots, v_0) \partial_{P_1} v_{j_1} \dots \partial_{P_r} v_{j_r}.$$

Proof. We argue by induction on the size of the set P . The case $|P| = 1$ is obvious. Suppose that the conclusion holds for some $m \geq 1$ and all $P \subset \mathbb{N}$ with $|P| = m$. Without loss of generality, it suffices to show that the conclusion holds for $P = [m+1]$. From induction hypothesis, we compute directly to get

$$\begin{aligned} \partial_{[m+1]}V &= \partial_{\{m+1\}}(\partial_{[m]}V) = \sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_r(k)} \sum_{\mathcal{P} \in \mathcal{P}_r([m])} \left(\partial_J F \partial_{\{m+1\}} (\partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r}) \right. \\ &\quad \left. + \left(\sum_{j=0}^k \partial_{j,J} F \partial_{\{m+1\}} v_j \right) \partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r} \right). \end{aligned}$$

To handle the first summation, note that

$$\partial_{\{m+1\}}(\partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r}) = \sum_{s=1}^r (\partial_{P_1} v_{j_1}) \cdots (\partial_{P_{s-1}} v_{j_{s-1}}) (\partial_{\{m+1\} \cup P_s} v_{j_s}) (\partial_{P_{s+1}} v_{j_{s+1}}) \cdots (\partial_{P_r} v_{j_r}),$$

which implies that

$$\begin{aligned} &\sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_r(k)} \sum_{\mathcal{P} \in \mathcal{P}_r([m])} \partial_J F \partial_{\{m+1\}} (\partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r}) \\ &= \sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_r(k)} \sum_{\mathcal{P} \in \mathcal{P}_r([m+1]): \{m+1\} \notin \mathcal{P}} \partial_J F (\partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r}). \end{aligned} \tag{5.2}$$

On the other hand, since

$$\left(\sum_{j=0}^k \partial_{j,J} F \partial_{\{m+1\}} v_j \right) \partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r} = \sum_{j_{r+1}=0}^k \partial_{j_{r+1},J} F \partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r} \partial_{\{m+1\}} v_{j_{r+1}},$$

it follows that

$$\begin{aligned} &\sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_r(k)} \sum_{\mathcal{P} \in \mathcal{P}_r([m])} \left(\sum_{j=0}^k \partial_{j,J} F \partial_{\{m+1\}} v_j \right) \partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r} \\ &= \sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_{r+1}(k)} \sum_{\mathcal{P} \in \mathcal{P}_{r+1}([m+1]): \{m+1\} \in \mathcal{P}} \partial_J F \partial_{P_1} v_{j_1} \cdots \partial_{P_{r+1}} v_{j_{r+1}}. \end{aligned}$$

To simplify this summation, we write

$$\begin{aligned} &\sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_{r+1}(k)} \sum_{\mathcal{P} \in \mathcal{P}_{r+1}([m+1]): \{m+1\} \in \mathcal{P}} \partial_J F \partial_{P_1} v_{j_1} \cdots \partial_{P_{r+1}} v_{j_{r+1}} \\ &= \sum_{2 \leq r \leq m+1} \sum_{J \in \mathcal{J}_r(k)} \sum_{\mathcal{P} \in \mathcal{P}_r([m+1]): \{m+1\} \in \mathcal{P}} \partial_J F \partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r} \\ &= \sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_r(k)} \sum_{\mathcal{P} \in \mathcal{P}_r([m+1]): \{m+1\} \in \mathcal{P}} \partial_J F \partial_{P_1} v_{j_1} \cdots \partial_{P_r} v_{j_r} \\ &\quad + \sum_{J \in \mathcal{J}_{m+1}(k)} \sum_{\mathcal{P} \in \mathcal{P}_{m+1}([m+1])} \partial_J F \partial_{P_1} v_{j_1} \cdots \partial_{P_{m+1}} v_{j_{m+1}}, \end{aligned}$$

where in the first equality we changed the variable $r+1 \rightarrow r$, while in the second equality we divide $2 \leq r \leq m+1$ into $2 \leq r \leq m$ and $r = m+1$ and use the observation that $\mathcal{P}_1([m+1])$ contains no element \mathcal{P} so that $\{m+1\} \in \mathcal{P}$. Combining this summation with (5.2) yields the desired formula. \square

Lemma 5.2. For any $H = (h_1, \dots, h_n)$ for $h_1, \dots, h_n \in C^m(M_n(\mathbb{R}))$, we have that

$$\partial_P(\hat{X}_n H(X)) = \frac{1}{\sqrt{n}} \sum_{r \in P} E_r \partial_{P \setminus \{r\}} H(X) + \hat{X}_n \partial_P H(X),$$

where $E_r \in M_n(\mathbb{R})$, whose entries are equal to 1 at (i_r, i'_r) and (i'_r, i_r) and are zero otherwise.

Proof. It suffices to assume that $P = [m]$. If $m = 1$, then $\partial_{\{1\}}(\hat{X}_n H(X)) = n^{-1/2} E_1 H(X) + \hat{X}_n \partial_{\{1\}} H(X)$. Assume that the assertion is valid for $m \geq 1$. Then

$$\begin{aligned} \partial_{[m+1]}(\hat{X}_n H(X)) &= \partial_{\{m+1\}} \left(\frac{1}{\sqrt{n}} \sum_{r \in [m]} E_r \partial_{[m] \setminus \{r\}} H(X) + \hat{X}_n \partial_{[m]} H(X) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{r \in [m]} E_r \partial_{[m+1] \setminus \{r\}} H(X) + \partial_{\{m+1\}}(\hat{X}_n \partial_{[m]} H(X)) \\ &= \frac{1}{\sqrt{n}} \sum_{r \in [m+1]} E_r \partial_{[m+1] \setminus \{r\}} H(X) + \hat{X}_n \partial_{[m+1]} H(X). \end{aligned}$$

□

5.2 Moment control

The most crucial ingredient of this paper lies on the following proposition, which establishes two special cases of Proposition 5.1.

Proposition 5.2. Consider the AMP orbits in Definition 2.1 with the initialization $u^{[0]}(X) = u^0$ and assume that $(F_k)_{k \geq 0}$ satisfies (5.1). For any $k \geq 0$, $p \geq 1$, and $m \geq 0$, there exist constants $\Gamma_{k,p,m}$ and $\Gamma'_{k,p,m}$ such that for any $n \geq 2$,

$$\sup_{\mathcal{B}_n(k,p,m)} (\mathbb{E} |\partial_P u^{[k]}(A)_i|^p)^{1/p} \leq \frac{\Gamma_{k,p,m}}{n^{m/2}} \quad (5.3)$$

and

$$\sup_{\mathcal{B}_n(k,p,m)} (\mathbb{E} |\partial_P(\hat{A}_n u^{[k]}(A))_i|^p)^{1/p} \leq \frac{\Gamma'_{k,p,m}}{n^{m/2}}. \quad (5.4)$$

Proof. We argue by induction on $k \geq 0$. First assume that $k = 0$. We aim to show that (5.3) and (5.4) are valid for all $p \geq 1$ and $m \geq 0$. For any P with $|P| = m$, since $\partial_P u^{[0]}(X) = u^0$ if $m = 0$ and $\partial_P u^{[0]}(X) = 0$ if $m \geq 1$, (2.1) obviously implies (5.3). To show (5.4), note that for any P with $|P| = m$,

$$\partial_P(\hat{X}_n u^{[0]}(X)) = \begin{cases} \frac{1}{\sqrt{n}} \hat{X}_n u^0, & \text{if } m = 0, \\ \frac{1}{\sqrt{n}} E_r u^0, & \text{if } m = 1 \text{ and } P = \{r\} \text{ for some } r \geq 1, \\ 0 & \text{if } m \geq 2, \end{cases}$$

where $E_r \in M_n(\mathbb{R})$ is equal to 1 on the entries (i_r, i'_r) and (i'_r, i_r) and is zero elsewhere. To control the first case, note that u^0 is independent of A and the entries in A are independent. From the subgaussianity of a_{ij} and (2.1), there exist positive constants $\lambda(\sigma)$ and $D(\sigma)$ such that for any $n \geq 1$ and $\lambda \in [-\lambda(\sigma), \lambda(\sigma)]$,

$$\mathbb{E} e^{\lambda n^{-1/2} \sum_{j=1}^n a_{ij} u_j^0} \leq \mathbb{E} e^{\lambda^2 \sigma^2 \|u^0\|^2 / 2n} \leq \mathbb{E} e^{\lambda(\sigma)^2 \sigma^2 \|u^0\|^2 / 2n} \leq D(\sigma).$$

Consequently, from $x^{2p} \leq (2p)! \cosh x$ and the Jensen inequality,

$$\mathbb{E} \left| \frac{\sum_{j=1}^n a_{ij} u_j^0}{\sqrt{n}} \right|^p \leq \frac{\sqrt{(2p)! D(\sigma)}}{\lambda(\sigma)^p}.$$

On the other hand, the Cauchy-Schwarz and Jensen inequalities imply that

$$\begin{aligned} \mathbb{E} \left| \sum_{j=1}^n \frac{z_{ij}}{n} u_j^0 \right|^p &\leq \mathbb{E} \left| \frac{\sum_{j=1}^n z_{ij}^2}{n} \right|^{p/2} \left(\frac{\|u^0\|_2}{\sqrt{n}} \right)^p \\ &\leq \left(\mathbb{E} \left| \frac{\sum_{j=1}^n z_{ij}^2}{n} \right|^p \right)^{1/2} \left(\mathbb{E} \left(\frac{\|u_0\|_2^2}{n} \right)^p \right)^{1/2} \\ &\leq \left(\mathbb{E} \left| \frac{\sum_{j=1}^n z_{ij}^{2p}}{n} \right| \right)^{1/2} \left(\mathbb{E} \left(\frac{\|u_0\|_2^2}{n} \right)^p \right)^{1/2} \leq D'(\sigma), \end{aligned}$$

where $D'(\sigma)$ is a constant independent of n and is guaranteed by the moment assumption (2.1). Combining these two inequalities validates (5.4) for $(k, p, m) = (0, p, 0)$. In the second case, since $E_r u^0$ has only two nonzero entries and they are $u_{i_r}^0$ and $u_{i'_r}^0$, the bound (2.1) implies (5.4) for $(k, p, m) = (0, p, 1)$. The third case is evident. In conclusion, (5.4) holds for $k = 0$, $p \geq 1$, and $m \geq 0$.

Next, we assume that there exists some $k_0 \geq 0$ such that (5.3) and (5.4) are valid for all $0 \leq k \leq k_0$, $p \geq 1$, and $m \geq 0$. Our goal is to show that they are also valid for $k = k_0 + 1$, $p \geq 1$, and $m \geq 0$. Denote by

$$v^{[k_0]}(X) = \hat{X}_n u^{[k_0]}(X), v^{[k_0-1]}(X) = u^{[k_0-1]}(X), \dots, v^{[0]}(X) = u^{[0]}(X).$$

First we verify (5.3). Assume that $m = 0$. From the Lipschitz property of F_{k_0} ,

$$\begin{aligned} (\mathbb{E} |u^{[k_0+1]}(A)_i|^p)^{1/p} &= (\mathbb{E} |F_{k_0}(v^{[k_0]}(A), \dots, v^{[0]}(A))_i|^p)^{1/p} \\ &\leq \eta_{k_0} \sum_{\ell=0}^{k_0} (\mathbb{E} |v^{[\ell]}(A)_i|^p)^{1/p} + |F_{k_0}(0)|, \end{aligned}$$

where η_{k_0} is the Lipschitz constant of F_{k_0} . By induction hypothesis, (5.3) follows when $(k, p, m) = (k_0 + 1, p, 0)$ for all $p \geq 1$. Now suppose that $m \geq 1$. For $n \geq 2$, consider an arbitrary P with $|P| = m$, $(i_r, i'_r)_{r \geq 1} \in \mathcal{T}_n$, $A \in \mathcal{G}_n(\sigma)$, and $i \in [n]$. From Lemma 5.1,

$$\begin{aligned} \partial_P u^{[k_0+1]}(X)_i &= \sum_{r \in P} \sum_{J \in \mathcal{J}_r(k_0), \mathcal{P} \in \mathcal{P}_r(P)} \partial_J F_{k_0}(v_i^{[k_0]}(X), \dots, v_i^{[0]}(X)) \\ &\quad \partial_{P_1} v_i^{[j_1]}(X) \cdots \partial_{P_r} v_i^{[j_r]}(X). \end{aligned}$$

From this, Minkowski's inequality, Hölder's inequality, and the boundedness of the partial derivatives of F_{k_0} , there exist constants $\eta_{k_0, J}$'s such that

$$(\mathbb{E} |\partial_P u^{[k_0+1]}(A)_i|^p)^{1/p} \leq \sum_{1 \leq r \leq m} \sum_{J \in \mathcal{J}_r(k_0), \mathcal{P} \in \mathcal{P}_r(P)} \eta_{k_0, J} \prod_{l=1}^r (\mathbb{E} |\partial_{P_l} v^{[j_l]}(A)_i|^{rp})^{1/rp}.$$

Since each $v^{[j]}(A)$ is either $\hat{A}_n u^{[j]}(A)$ or $u^{[j]}(A)$, this implies that

$$\begin{aligned} (\mathbb{E}|\partial_P u^{[k_0+1]}(A)_i|^p)^{1/p} &\leq \sum_{r \in P} \sum_{J \in \mathcal{J}_r(k_0), \mathcal{P} \in \mathcal{P}_r(P)} \prod_{l=1}^r \frac{\max(\Gamma_{k_0, rP, |P_l|}, \Gamma'_{k_0, rP, |P_l|})}{n^{|P_l|/2}} \\ &= \frac{1}{n^{m/2}} \sum_{r \in P} \sum_{J \in \mathcal{J}_r(k_0), \mathcal{P} \in \mathcal{P}_r(P)} \prod_{l=1}^r \max(\Gamma_{k_0, rP, |P_l|}, \Gamma'_{k_0, rP, |P_l|}) \\ &=: \frac{1}{n^{m/2}} \Gamma_{k_0+1, p, m}. \end{aligned}$$

Hence, (5.3) is valid for $k = k_0 + 1$, $p \geq 1$, and $m \geq 1$. Putting these two cases together yields the validity of (5.3) for $k = k_0 + 1$, $p \geq 1$, and $m \geq 0$.

Now we verify (5.4). Let $P \subset \mathbb{N}$ with $|P| = m$, $(i_r, i'_r)_{r \geq 1} \in \mathcal{T}_n$, $A \in \mathcal{G}_n(\sigma)$, and $i \in [n]$. Note that from Lemma 5.2,

$$\partial_P(\hat{X}_n u^{[k_0+1]}(X)) = \frac{1}{\sqrt{n}} \sum_{r \in P} E_r \partial_{P \setminus \{r\}} u^{[k_0+1]}(X) + \hat{X}_n \partial_P u^{[k_0+1]}(X). \quad (5.5)$$

The first term can be controlled by

$$\begin{aligned} &\left(\mathbb{E} \left| \frac{1}{\sqrt{n}} \sum_{r \in P} E_r \partial_{P \setminus \{r\}} u^{[k_0+1]}(A)_i \right|^p \right)^{1/p} \\ &\leq \frac{1}{\sqrt{n}} \sum_{r \in P} (\mathbb{E} |\delta_{i, i_r} \partial_{P \setminus \{r\}} u^{[k_0+1]}(A)_{i'_r} + \delta_{i, i'_r} \partial_{P \setminus \{r\}} u^{[k_0+1]}(A)_{i_r}|^p)^{1/p} \\ &\leq \frac{1}{\sqrt{n}} \sum_{r \in P} \left((\mathbb{E} |\partial_{P \setminus \{r\}} u^{[k_0+1]}(A)_{i'_r}|^p)^{1/p} + (\mathbb{E} |\partial_{P \setminus \{r\}} u^{[k_0+1]}(A)_{i_r}|^p)^{1/p} \right) \\ &\leq \frac{2m \Gamma_{k_0+1, p, m-1}}{n^{m/2}}, \end{aligned} \quad (5.6)$$

where $\delta_{i, i'} = 1$ if $i = i'$ and it is zero if $i \neq i'$. As for the second term, note that for any $u \in \mathbb{R}^n$,

$$|(\hat{X}_n u)_i| \leq |(X_n u)_i| + \frac{1}{n} \left(\sum_{j=1}^n z_{ij}^2 \right)^{1/2} \|u\|_2.$$

This implies that

$$\begin{aligned} &(\mathbb{E} |(\hat{A}_n \partial_P u^{[k_0+1]}(A))_i|^p)^{1/p} \\ &\leq (\mathbb{E} |(A_n \partial_P u^{[k_0+1]}(A))_i|^p)^{1/p} + \frac{1}{n} \left(\mathbb{E} \left[\left(\sum_{j=1}^n z_{ij}^2 \right)^{p/2} \|\partial_P u^{[k_0+1]}(A)\|_2^p \right] \right)^{1/p} \\ &\leq (\mathbb{E} |(A_n \partial_P u^{[k_0+1]}(A))_i|^p)^{1/p} + \left(\mathbb{E} \left(\frac{\sum_{j=1}^n z_{ij}^2}{n^p} \right)^{1/2p} \left(\mathbb{E} \frac{\|\partial_P u^{[k_0+1]}(A)\|_2^{2p}}{n^p} \right)^{1/2p} \right). \end{aligned}$$

Here, from (2.1), $n^{-p} \mathbb{E} (\sum_{j=1}^n z_{ij}^2)^p$ is bounded above by a constant independent of n . From the validity of (5.3) for $k = k_0 + 1$, $p \geq 1$, and $m \geq 0$ that we established above, we also have that

$$\left(\mathbb{E} \frac{\|\partial_P u^{[k_0+1]}(A)\|_2^{2p}}{n^p} \right)^{1/2p} \leq \left(\frac{\mathbb{E} \sum_{i=1}^n |\partial_P u^{[k_0+1]}(A)_i|^{2p}}{n} \right)^{1/2p} \leq \frac{\Gamma_{k_0+1, 2p, m}}{n^{m/2}}$$

and from Lemma 5.3 below,

$$(\mathbb{E}|(A_n \partial_P u^{[k_0+1]}(A))_i|^p)^{1/p} \leq \frac{\Upsilon_{k_0+1,p,m}}{n^{m/2}},$$

where $\Upsilon_{k_0+1,p,m}$ is a universal constant independent of n . Therefore, we arrive at

$$(\mathbb{E}|(\hat{A}_n \partial_P u^{[k_0+1]}(A))_i|^p)^{1/p} \leq \frac{C}{n^{m/2}},$$

where C is a constant independent of n . Plugging this inequality and (5.6) into (5.5) yields (5.4) for $k = k_0 + 1$, $p \geq 1$, and $m \geq 0$. This completes our proof. \square

At the end of this subsection, we establish the following lemma used in the above proof.

Lemma 5.3. *Let $k \geq 1$. Assume that for any $p \geq 1$ and $m \geq 0$, there exists a constant $\Gamma_{k,p,m}$ such that*

$$\sup_{\mathcal{B}_n(k,p,m)} (\mathbb{E}|\partial_P u^{[k]}(A)_i|^p)^{1/p} \leq \frac{\Gamma_{k,p,m}}{n^{m/2}}, \quad \forall n \geq 2. \quad (5.7)$$

Then for any $p \geq 1$ and $m \geq 0$, there exists a constant $\Upsilon_{k,p,m}$ such that

$$\sup_{\mathcal{B}_n(k,p,m)} (\mathbb{E}|(A_n \partial_P u^{[k]}(A))_i|^p)^{1/p} \leq \frac{\Upsilon_{k,p,m}}{n^{m/2}}, \quad \forall n \geq 2.$$

Proof. Our idea is to use the Taylor expansion to track the dependence of

$$(A_n \partial_P u^{[k]}(A))^p$$

on each variable $a_{ii'}$ in each iteration. By Jensen's inequality, it suffices to assume that p is even. Let $m \geq 0$ be fixed. Let $P \subset \mathbb{N}$ with $|P| = m$, $(i_r, i'_r)_{r \geq 1} \in \mathcal{T}_n$, $A \in \mathcal{G}_n(\sigma)$, and $i \in [n]$. Denote

$$V(X) = \partial_P u^{[k]}(X).$$

For any $D \subseteq [p]$, let \mathcal{I}_D be the collection of all $I = (\iota_1, \dots, \iota_p) \in [n]^p$ such that ι_s are distinct for $s \in D$ and

$$\{\iota_s : s \in D^c\} \subseteq \{\iota_s : s \in D\}. \quad (5.8)$$

For $I \in \mathcal{I}_D$ and $X \in M_n(\mathbb{R})$, let $X^I \in M_n(\mathbb{R})$, in which each entry of X^I is equal to that of X except that it vanishes on the sites (i, ι_s) and (ι_s, i) for all $s \in D$. For $\iota \in [n]$ and $0 \leq t \leq 1$, define

$$f_\iota(t) = V_\iota(X^I(t))$$

for

$$X^I(t) := tX + (1-t)X^I.$$

Here, V_ι is the ι -th entry of V . Write by Taylor's theorem,

$$V_\iota(X) = f_\iota(1) = \sum_{a=0}^{p-1} \frac{f_\iota^{(a)}(0)}{a!} + \frac{1}{(p-1)!} \int_0^1 (1-t)^{p-1} f_\iota^{(p)}(t) dt =: M_\iota^I(X) + N_\iota^I(X).$$

Note that here

$$f_\iota^{(a)}(0) = \sum_{s_1, \dots, s_a \in D} \frac{\partial^a V_\iota(X^I)}{\partial x_{i\iota_{s_a}} \cdots \partial x_{i\iota_{s_1}}} x_{i\iota_{s_1}} \cdots x_{i\iota_{s_a}}.$$

Also, note that $\mathcal{I}_D \cap \mathcal{I}_{D'} = \emptyset$ for distinct D and D' and that $[n]^p = \cup_{D \subseteq [p]} \mathcal{I}_D$. Write

$$\begin{aligned}
\mathbb{E}(A_n V(A))_i^p &= \frac{1}{n^{p/2}} \sum_{D \subseteq [p]} \sum_{I \in \mathcal{I}_D} \mathbb{E} a_{i_{l_1}} \cdots a_{i_{l_p}} V_{l_1}(A) \cdots V_{l_p}(A) \\
&= \frac{1}{n^{p/2}} \sum_{D \subseteq [p]} \sum_{I \in \mathcal{I}_D} \sum_{S \subseteq [p]} \mathbb{E} \left(\prod_{l \in [p]} a_{i_{l_l}} \right) \left(\prod_{l \in S} M_{l_l}^I(A) \right) \left(\prod_{l \in S^c} N_{l_l}^I(A) \right) \\
&= \frac{1}{n^{p/2}} \sum_{D \subseteq [p]} \sum_{I \in \mathcal{I}_D} \sum_{S \subsetneq [p]} \mathbb{E} \left(\prod_{l \in [p]} a_{i_{l_l}} \right) \left(\prod_{l \in S} M_{l_l}^I(A) \right) \left(\prod_{l \in S^c} N_{l_l}^I(A) \right) \\
&\quad + \frac{1}{n^{p/2}} \sum_{D \subseteq [p]} \sum_{I \in \mathcal{I}_D} \mathbb{E} \left(\prod_{l \in [p]} a_{i_{l_l}} \right) \left(\prod_{l \in [p]} M_{l_l}^I(A) \right) \\
&=: \Delta_{n,1} + \Delta_{n,2}.
\end{aligned}$$

To control these two terms, note that from $x^{2p} \leq (2p)! \cosh x$ and the σ -subgaussianity, we have the bound

$$\sup_{i,j \in [n]} (\mathbb{E} |a_{ij}|^p)^{1/p} \leq \sup_{i,j \in [n]} (\mathbb{E} |a_{ij}|^{2p})^{1/2p} \leq \xi_p := (2p)! e^{\sigma^2/2}, \quad \forall p \geq 1.$$

First we handle $\Delta_{n,1}$. From the given assumption,

$$\begin{aligned}
&\left(\mathbb{E} \left| \frac{\partial^a V_{l_l}(A^I)}{\partial x_{i_{l_{s_a}}} \cdots \partial x_{i_{l_{s_1}}}} a_{i_{l_{s_1}}} \cdots a_{i_{l_{s_a}}} \right|^{2p} \right)^{1/2p} \\
&\leq \left(\mathbb{E} \left| \frac{\partial^a V_{l_l}(A^I)}{\partial x_{i_{l_{s_a}}} \cdots \partial x_{i_{l_{s_1}}}} \right|^{4p} \right)^{1/4p} \left(\mathbb{E} |a_{i_{l_{s_1}}} \cdots a_{i_{l_{s_a}}}|^{4p} \right)^{1/4p} \leq \frac{\Gamma_{k,4p,a+m} \xi_{4ap}^a}{n^{(a+m)/2}}.
\end{aligned}$$

Using the Minkowski inequality, this inequality, and (5.7) yields that after dropping $1/a!$,

$$\begin{aligned}
\prod_{l \in S} (\mathbb{E} |M_{l_l}^I(A)|^{2p})^{1/2p} &\leq \prod_{l \in S} \sum_{a=0}^{p-1} \sum_{s_1, \dots, s_a \in D} \frac{\Gamma_{k,4p,a+m} \xi_{4ap}^a}{n^{(a+m)/2}} \\
&\leq \frac{1}{n^{|S|m/2}} \left(\sum_{a=0}^{p-1} |D|^a \Gamma_{k,4p,a+m} \xi_{4ap}^a \right)^{|S|} =: \frac{C_{D,S}}{n^{|S|m/2}}.
\end{aligned}$$

Also, since

$$\mathbb{E} \left| \int_0^1 (1-t)^{p-1} f_t^{(p)}(t) dt \right|^{2p} \leq \mathbb{E} \left(\int_0^1 |f_t^{(p)}(t)| dt \right)^{2p} \leq \int_0^1 \mathbb{E} |f_t^{(p)}(t)|^{2p} dt,$$

we have, by dropping $1/(p-1)!$, that

$$\begin{aligned}
\prod_{l \in S^c} (\mathbb{E} |N_{l_l}^I(A)|^{2p})^{1/2p} &\leq \prod_{l \in S^c} \sum_{s_1, \dots, s_a \in D} \left(\int_0^1 \mathbb{E} \left| \frac{\partial^p V_{l_l}(A^I(t))}{\partial x_{i_{l_{s_p}}} \cdots \partial x_{i_{l_{s_1}}}} a_{i_{l_{s_1}}} \cdots a_{i_{l_{s_p}}} \right|^{2p} dt \right)^{1/2p} \\
&\leq \prod_{l \in S^c} \frac{|D|^p \Gamma_{k,4p,p+m} \xi_{4p^2}^p}{n^{(p+m)/2}} \\
&= \frac{1}{n^{(p+m)|S^c|/2}} \prod_{l \in S^c} |D|^p \Gamma_{k,4p,p+m} \xi_{4p^2}^p \\
&=: \frac{C'_{D,S}}{n^{(p+m)|S^c|/2}}.
\end{aligned}$$

Combining these together leads to

$$\begin{aligned}
|\Delta_{n,1}| &\leq \frac{1}{n^{p/2}} \sum_{D \subseteq [p]} \sum_{I \in \mathcal{I}_D} \sum_{S \subsetneq [p]} \prod_{l \in [p]} (\mathbb{E}|a_{i_l}|^{2p})^{1/2p} \prod_{l \in S} (\mathbb{E}|M_{i_l}^I(A)|^{2p})^{1/2p} \prod_{l \in S^c} (\mathbb{E}|N_{i_l}^I(A)|^{2p})^{1/2p} \\
&\leq \frac{1}{n^{p/2}} n^p \xi_{2p}^p \sum_{D \subseteq [p]} \sum_{S \subsetneq [p]} \frac{C_{D,S}}{n^{|S|m/2}} \frac{C'_{D,S}}{n^{(p+m)|S^c|/2}} \\
&= \xi_{2p}^p \sum_{D \subseteq [p]} \sum_{S \subsetneq [p]} \frac{C_{D,S} C'_{D,S}}{n^{(p(|S^c|-1)+pm)/2}} \\
&\leq \frac{\xi_{2p}^p}{n^{pm/2}} \sum_{D \subseteq [p]} \sum_{S \subsetneq [p]} C_{D,S} C'_{D,S},
\end{aligned}$$

where the last inequality used the fact that $|S^c| \geq 1$ since $S \subsetneq [p]$.

Next we turn to the control of $\Delta_{n,2}$, which requires more steps. Let $D \subseteq [n]$ and $I \in \mathcal{I}_D$. Write

$$\prod_{l \in [p]} M_{i_l}^I(A) = \sum_{\bar{a}} \frac{1}{\bar{a}!} \sum_{\bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p} \left(\prod_{r \in [p]} \prod_{b=1}^{a_r} a_{i_{s_b^r}} \right) \left(\prod_{r \in [p]} \frac{\partial^{a_r} V_{i_r}(A^I)}{\partial x_{i_{s_1^r}} \dots \partial x_{i_{s_{a_r}^r}}} \right),$$

where the first summation is over all $\bar{a} = (a_1, \dots, a_p) \in \{0, \dots, p-1\}^p$ and $\bar{a}! := a_1! \dots a_p!$, while the second summation is over all $\bar{s}_{a_r}^r = (s_1^r, \dots, s_{a_r}^r) \in D^{a_r}$ for $1 \leq r \leq p$. Set $\mathcal{S}_{\bar{a}} = D^{a_1} \times \dots \times D^{a_p}$. Note that from our construction of A^I , its entries at (i, i_s) and (i_s, i) are all zero for all $s \in D$ and consequently, (5.8) implies that A^I is independent of $a_{i_{\iota_1}}, \dots, a_{i_{\iota_p}}$. It follows that

$$\Delta_{n,2} = \frac{1}{n^{p/2}} \sum_{\bar{a}} \frac{1}{\bar{a}!} \sum_{D \subseteq [p]} \sum_{I \in \mathcal{I}_D} \sum_{\mathcal{S}_{\bar{a}}} \mathcal{AL}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p),$$

where

$$\begin{aligned}
\mathcal{A}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p) &:= \mathbb{E} \left(a_{i_{\iota_1}} \dots a_{i_{\iota_p}} \prod_{r \in [p]} \prod_{b=1}^{a_r} a_{i_{s_b^r}} \right), \\
\mathcal{L}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p) &:= \mathbb{E} \left(\prod_{r \in [p]} \frac{\partial^{a_r} V_{i_r}(A^I)}{\partial x_{i_{s_1^r}} \dots \partial x_{i_{s_{a_r}^r}}} \right), \\
\mathcal{AL}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p) &:= \mathcal{A}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p) \mathcal{L}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p).
\end{aligned}$$

To control the right-hand side, note that

$$|\mathcal{A}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p)| \leq \xi_{p+|\bar{a}|}^{p+|\bar{a}|}$$

and from (5.7),

$$\begin{aligned}
|\mathcal{L}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p)| &\leq \prod_{r \in [p]} \left(\mathbb{E} \left| \frac{\partial^{a_r} V_{i_r}(A^I)}{\partial x_{i_{s_1^r}} \dots \partial x_{i_{s_{a_r}^r}}} \right|^p \right)^{1/p} \\
&\leq \prod_{r \in [p]} \frac{\Gamma_{k,p,m+a_r}}{n^{(m+a_r)/2}} = \frac{1}{n^{(pm+|\bar{a}|)/2}} \prod_{r \in [p]} \Gamma_{k,p,m+a_r},
\end{aligned}$$

where $|\bar{a}| := a_1 + \dots + a_p$. For any fixed $D \subseteq [p]$ and $0 \leq b \leq |D|$, let $\mathcal{I}_{D,b}$ be the collection of all $I \in \mathcal{I}_D$ such that there are exactly b many entries that appear once in I . For any $I \in \mathcal{I}_{D,b}$, let $\mathcal{S}_{\bar{a},I,b}$

be the collection of all $(\bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p) \in \mathcal{S}_{\bar{a}}$ such that all entries that appear exactly once in I also appear in

$$\{\iota_{s_1^1}, \dots, \iota_{s_{a_1}^1}, \iota_{s_1^2}, \dots, \iota_{s_{a_2}^2}, \dots, \iota_{s_1^p}, \dots, \iota_{s_{a_p}^p}\}.$$

Note that when $b = 0$, every entry in $I \in \mathcal{I}_{D,0}$ must appear at least twice in I and hence, $\mathcal{S}_{\bar{a},I,b} = \mathcal{S}_{\bar{a}}$, or equivalently, $\mathcal{S}_{\bar{a},I,b}^c = \emptyset$. Also, note that $\mathcal{S}_{\bar{a},I,b}$ is nonempty only if $|\bar{a}| \geq b$. From these, we can write

$$\begin{aligned} \sum_{\bar{a}} \sum_{I \in \mathcal{I}_D} \sum_{\mathcal{S}_{\bar{a}}} &= \sum_{\bar{a}} \sum_{b=0}^{|D|} \sum_{I \in \mathcal{I}_{D,b}} \sum_{\mathcal{S}_{\bar{a},I,b}} + \sum_{\bar{a}} \sum_{b=0}^{|D|} \sum_{I \in \mathcal{I}_{D,b}} \sum_{\mathcal{S}_{\bar{a},I,b}^c} \\ &= \sum_{\bar{a}} \sum_{b=0}^{|D| \wedge |\bar{a}|} \sum_{I \in \mathcal{I}_{D,b}} \sum_{\mathcal{S}_{\bar{a},I,b}} + \sum_{\bar{a}} \sum_{b=1}^{|D|} \sum_{I \in \mathcal{I}_{D,b}} \sum_{\mathcal{S}_{\bar{a},I,b}^c}. \end{aligned}$$

To control the first summation, note that for any $I \in \mathcal{I}_{D,b}$, there are exactly b many entries in I that appear once and the other entries are repeated. This implies that

$$|\mathcal{I}_{D,b}| \leq C_{D,b} n^{b + \lfloor (p-b)/2 \rfloor},$$

where $C_{D,b}$ is a universal constant independent of n . Plugging this inequality to the above equation yields that

$$\begin{aligned} &\frac{1}{n^{p/2}} \sum_{\bar{a}} \sum_{b=0}^{|D| \wedge |\bar{a}|} \sum_{I \in \mathcal{I}_{D,b}} \sum_{\mathcal{S}_{\bar{a},I,b}} |\mathcal{AL}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p)| \\ &\leq \frac{1}{n^{p/2}} \sum_{\bar{a}} \sum_{b=0}^{|D| \wedge |\bar{a}|} \frac{|\mathcal{I}_{D,b}| |D|^{|\bar{a}|} \xi_{p+|\bar{a}|}^{p+|\bar{a}|}}{n^{(pm+|\bar{a}|)/2}} \prod_{r \in [p]} \Gamma_{k,p,m+a_r} \\ &\leq \sum_{\bar{a}} \sum_{b=0}^{|D| \wedge |\bar{a}|} \frac{1}{n^{pm/2 + |\bar{a}|/2 + p/2 - b - \lfloor (p-b)/2 \rfloor}} C_{D,b} |D|^{|\bar{a}|} \xi_{p+|\bar{a}|}^{p+|\bar{a}|} \prod_{r \in [p]} \Gamma_{k,p,m+a_r} \\ &\leq \sum_{\bar{a}} \sum_{b=0}^{|D| \wedge |\bar{a}|} \frac{1}{n^{pm/2 + (p-b)/2 - \lfloor (p-b)/2 \rfloor}} C_{D,b} |D|^{|\bar{a}|} \xi_{p+|\bar{a}|}^{p+|\bar{a}|} \prod_{r \in [p]} \Gamma_{k,p,m+a_r} \\ &\leq \frac{1}{n^{pm/2}} \sum_{\bar{a}} \sum_{b=0}^{|D| \wedge |\bar{a}|} C_{D,b} |D|^{|\bar{a}|} \xi_{p+|\bar{a}|}^{p+|\bar{a}|} \prod_{r \in [p]} \Gamma_{k,p,m+a_r}, \end{aligned} \tag{5.9}$$

where the third inequality used $|\bar{a}| \geq b$. As for the second summation, observe that for $1 \leq b \leq |D|$, if $I \in \mathcal{I}_{D,b}$ and $(\bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p) \in \mathcal{S}_{\bar{a},I,b}^c$, then there exists one entry, say ι_{ℓ} , in I that appears only once in I and it does not appear in the entries of $\bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p$. Hence, $a_{i\iota_{\ell}}$ is independent of

$$a_{i\iota_1} \cdots a_{i\iota_{\ell-1}} a_{i\iota_{\ell+1}} \cdots a_{i\iota_p} \prod_{r \in [p]} \prod_{b=1}^{a_r} a_{i\iota_{s_b^r}},$$

which results in $\mathbb{E}\mathcal{AL}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p) = 0$. Therefore,

$$\frac{1}{n^{p/2}} \sum_{\bar{a}} \sum_{b=1}^{|D|} \sum_{I \in \mathcal{I}_{D,b}} \sum_{\mathcal{S}_{\bar{a},I,b}^c} \mathcal{AL}(I, \bar{s}_{a_1}^1, \dots, \bar{s}_{a_p}^p) = 0. \tag{5.10}$$

Finally, combining (5.9) and (5.10) together and dropping $1/\bar{a}!$ lead to

$$|\Delta_{n,2}| \leq \frac{1}{n^{pm/2}} \sum_{D \subseteq [p]} \sum_{\bar{a}} \sum_{b=0}^{|D| \wedge |\bar{a}|} C_{D,b} |D|^{|\bar{a}|} \xi_{p+|\bar{a}|}^{p+|\bar{a}|} \prod_{r \in [p]} \Gamma_{k,p,m+a_r}.$$

This completes our proof. \square

5.3 Proof of Proposition 5.1

Let $m = |P|$. From Lemma 5.1, the Minkowski inequality, and the Cauchy-Schwarz inequality, we can compute the partial derivatives of $U(X)_i$ to see that $(\mathbb{E}|\partial_P U(A)_i|^p)^{1/p}$ is bounded above by a sum, in which each summand is of the form

$$(\mathbb{E}|\partial_{P_1} v_{j_1}(A) \cdots \partial_{P_r} v_{j_r}(A)|^p)^{1/p}$$

for some $1 \leq r \leq m$ and $(j_1, \dots, j_r) \in \mathcal{J}_r(2(k+1))$. More importantly, $\mathcal{P} = \{P_1, \dots, P_r\} \in \mathcal{P}_r(P)$ and each term $v_{j_s}(A)$ is equal to either $(\hat{X}_n u^{[\ell]}(A))_i$ or $u^{[\ell]}(A)_i$ for some $0 \leq \ell \leq k$. Now, using the Hölder inequality gives that

$$(\mathbb{E}|\partial_{P_1} v_{j_1}(A) \cdots \partial_{P_r} v_{j_r}(A)|^p)^{1/p} \leq (\mathbb{E}|\partial_{P_1} v_{j_1}(A)|^{rp})^{1/rp} \cdots (\mathbb{E}|\partial_{P_r} v_{j_r}(A)|^{rp})^{1/rp}.$$

Here, from Proposition 5.2, each term on the right-hand side is bounded above by a term of order $1/n^{|P_s|/2}$ and they together yield that a bound of order $1/n^{|P|/2}$ since $|P_1| + \dots + |P_r| = |P|$. This completes our proof.

6 Proof of Theorem 2.1

We establish universality for the generalized AMP in Definition 2.1. Recall that in Theorem 3.1, we use $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{E}}$ to denote the probability and expectation conditionally on u^0, Z . The following proposition shows that conditionally on u^0, Z , the first two moments of the AMP orbits between A and G asymptotically match each other.

Proposition 6.1. *Consider the AMP orbit in Definition 2.1 with the initialization $u^{[0]}(X) = u^0$ and assume that $(F_k)_{k \geq 0}$ satisfies (5.1). Let $k \geq 0$. Assume that $\phi \in C^\infty(\mathbb{R}^{k+1})$ has uniformly bounded partial derivatives of any nonzero orders. There exists a universal constant C independent of n such that for any $n \geq 2$,*

$$\mathbb{E}|\tilde{\mathbb{E}}\Phi_{k,n}(A) - \tilde{\mathbb{E}}\Phi_{k,n}(G)| \leq \frac{C}{\sqrt{n}} \quad (6.1)$$

and

$$\mathbb{E}|\tilde{\mathbb{E}}\Phi_{k,n}(A)^2 - \tilde{\mathbb{E}}\Phi_{k,n}(G)^2| \leq \frac{C}{\sqrt{n}}, \quad (6.2)$$

where $\Phi_{k,n}(X)$ is defined in (2.2).

The proof of Theorem 2.1 is argued as follows. From Proposition 4.1, it suffices to assume that the functions ϕ and F_k 's in Definition 2.1 have uniformly bounded partial derivatives of any nonzero orders. First of all, we claim that

$$\lim_{n \rightarrow \infty} \mathbb{E}(\Phi_{k,n}(A) - \tilde{\mathbb{E}}\Phi_{k,n}(A))^2 = \lim_{n \rightarrow \infty} \mathbb{E}(\Phi_{k,n}(G) - \tilde{\mathbb{E}}\Phi_{k,n}(G))^2 = 0.$$

Note that from Theorem 3.1,

$$\lim_{n \rightarrow \infty} \mathbb{E} |\Phi_{k,n}(G) - \tilde{\mathbb{E}}\Phi_{k,n}(G)|^2 = 0.$$

Now from Proposition 6.1,

$$\lim_{n \rightarrow \infty} |\mathbb{E}\Phi_{k,n}(A)^2 - \mathbb{E}\Phi_{k,n}(G)^2| \leq \lim_{n \rightarrow \infty} \mathbb{E} |\tilde{\mathbb{E}}\Phi_{k,n}(A)^2 - \tilde{\mathbb{E}}\Phi_{k,n}(G)^2| = 0.$$

Since

$$\begin{aligned} & |\mathbb{E}(\tilde{\mathbb{E}}\Phi_{k,n}(A))^2 - \mathbb{E}(\tilde{\mathbb{E}}\Phi_{k,n}(G))^2| \\ &= \mathbb{E}(|\tilde{\mathbb{E}}\Phi_{k,n}(A) - \tilde{\mathbb{E}}\Phi_{k,n}(G)| |\tilde{\mathbb{E}}\Phi_{k,n}(A) + \tilde{\mathbb{E}}\Phi_{k,n}(G)|) \\ &\leq (\mathbb{E}(|\tilde{\mathbb{E}}\Phi_{k,n}(A) - \tilde{\mathbb{E}}\Phi_{k,n}(G)|^2))^{1/2} (\mathbb{E}|\tilde{\mathbb{E}}\Phi_{k,n}(A) + \tilde{\mathbb{E}}\Phi_{k,n}(G)|^2)^{1/2} \\ &\leq (\mathbb{E}(|\tilde{\mathbb{E}}\Phi_{k,n}(A) - \tilde{\mathbb{E}}\Phi_{k,n}(G)|^2))^{1/2} ((\mathbb{E}\Phi_{k,n}(A)^2)^{1/2} + (\mathbb{E}\Phi_{k,n}(G)^2)^{1/2}), \end{aligned}$$

it follows from Proposition 6.1 and the moment control in (3.5),

$$\lim_{n \rightarrow \infty} |\mathbb{E}(\tilde{\mathbb{E}}\Phi_{k,n}(A))^2 - \mathbb{E}(\tilde{\mathbb{E}}\Phi_{k,n}(G))^2| = 0.$$

Putting these limits together yields the claim. Consequently, the proof of Theorem 2.1 follows by our claim and Proposition 6.1,

$$\begin{aligned} \mathbb{E}(\Phi_{k,n}(A) - \Phi_{k,n}(G))^2 &\leq 9\mathbb{E}(\Phi_{k,n}(A) - \tilde{\mathbb{E}}\Phi_{k,n}(A))^2 \\ &\quad + 9\mathbb{E}(\tilde{\mathbb{E}}\Phi_{k,n}(A) - \tilde{\mathbb{E}}\Phi_{k,n}(G))^2 \\ &\quad + 9\mathbb{E}(\Phi_{k,n}(G) - \tilde{\mathbb{E}}\Phi_{k,n}(G))^2 \rightarrow 0. \end{aligned}$$

For the rest of this section, we establish Proposition 6.1 in five subsections using the Gaussian interpolation and approximate Gaussian integration by parts. In doing these, Proposition 5.1 will be of great use in tracking the error terms. Subsection 6.1 shows that to prove Proposition 6.1, it suffices to assume that the main diagonals of A, G, Z are all equal to zero. The Gaussian interpolation between $\Phi_{k,n}(A)$ and $\Phi_{k,n}(G)$ is introduced in Subsection 6.2 and the control of its derivative is handled in Subsection 6.3. Finally, the proofs of (6.1) and (6.2) are established in Subsections 6.4 and 6.5, respectively.

6.1 Deletion of the main diagonal

By the virtue of Proposition 3.1, it suffices to assume that the main diagonals in A and Z are zero. To see this, recall $\Phi_{k,n}(X)$, $\Delta_k(X)$, and $\Theta_k(X)$ from Proposition 3.1. Assume that A' is equal to A except that the main diagonal vanishes. Note that from Lemma 3.3, $\|A_n\|_2$ and $\|A'_n\|_2$ are of order $O(1)$. On the other hand, since

$$\|A_n - A'_n\|_2 = \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |a_{ii}|$$

and

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |a_{ii}| \geq t\right) \leq \sum_{i \in [n]} P(|a_{ii}| \geq t\sqrt{n}) \leq ne^{-n\sigma^2 t^2/2},$$

these imply that in probability,

$$\lim_{n \rightarrow \infty} \|A_n - A'_n\|_2 = 0.$$

As a result, Proposition 3.1 implies that the AMP orbits corresponding to (A, Z) and (A', Z) satisfy that in probability,

$$\lim_{n \rightarrow \infty} |\Phi_{k,n}(A) - \Phi_{k,n}(A')| = 0,$$

which together with Lemma 3.4 gives that

$$\lim_{n \rightarrow \infty} \mathbb{E} |\Phi_{k,n}(A) - \Phi_{k,n}(A')| = 0.$$

Next one can prove by an almost identical argument to show that the AMP orbits correspond to (A', Z) and (A', Z') are also asymptotically the same under the $L^1(\mathbb{P})$ -distance, where Z' is the same as Z except that its main diagonal is zero. From this, in what follows, we assume that the main diagonals of A , G , and Z are all equal to zero.

6.2 Interpolation

Define the Gaussian interpolation between A and G by

$$A(t) = (a_{ij}(t))_{i,j \in [n]} = \sqrt{t}A + \sqrt{1-t}G, \quad 0 \leq t \leq 1.$$

Denote

$$A_n(t) = \frac{A(t)}{\sqrt{n}}$$

and

$$\hat{A}_n(t) = \frac{A(t)}{\sqrt{n}} + \frac{Z}{n}.$$

For $\phi \in C^\infty(\mathbb{R}^{k+1})$ with uniformly bounded partial derivatives of all nonzero orders, define

$$\Phi_{k,n}(t) = \tilde{\mathbb{E}} \Phi_{k,n}(A(t)).$$

Note that

$$\mathbb{E} |\tilde{\mathbb{E}} \Phi_{k,n}(A) - \tilde{\mathbb{E}} \Phi_{k,n}(G)| = \mathbb{E} |\Phi_{k,n}(1) - \Phi_{k,n}(0)|.$$

To show (6.1), our goal is to show that

$$\int_0^1 \mathbb{E} |\Phi'_{k,n}(t)| dt \leq \frac{C}{\sqrt{n}}$$

for some constant C independent of n . Note that $u^{[0]}(X) = u$. A direct differentiation gives

$$\Phi'_{k,n}(t) = \sum_{\ell=1}^k \frac{1}{n} \sum_{i=1}^n \tilde{\mathbb{E}} \left(\partial_{y_\ell} \phi(u^{[k]}(A(t)), \dots, u^{[0]}(A(t))) \circ \frac{d}{dt} u^{[\ell]}(A(t)) \right)_i, \quad 0 < t < 1. \quad (6.3)$$

Here and thereafter, if $v, v' \in \mathbb{R}^n$, we define $v \circ v' = (v_i v'_i)_{i \in [n]}$ as the Hadamard product between v and v' . Note that this operation is commutative.

To simplify our notation, denote

$$\dot{A}_n(t) = (\dot{a}_{ii'}(t))_{i,i' \in [n]},$$

for

$$\dot{a}_{ii'}(t) = \frac{1}{2} \left(\frac{a_{ii'}}{\sqrt{t}} - \frac{g_{ii'}}{\sqrt{1-t}} \right).$$

Also, denote

$$\begin{aligned} u^{[\ell]}(t) &= u^{[\ell]}(A(t)), \\ \partial_{y_r} F_\ell(t) &= \partial_{y_r} F_\ell(\hat{A}_n(t) u^{[\ell]}(t), u^{[\ell-1]}(t), \dots, u^{[0]}(t)). \end{aligned}$$

Observe that

$$\begin{aligned} \frac{d}{dt} u^{[1]}(t) &= \partial_{y_0} F_0(t) \circ (\dot{A}_n(t) u^{[0]}(t)), \\ \frac{d}{dt} u^{[2]}(t) &= \partial_{y_1} F_1(t) \circ (\dot{A}_n(t) u^{[1]}(t)) \\ &\quad + \partial_{y_1} F_1(t) \circ (\hat{A}_n(\partial_{y_0} F_0(t) \circ (\dot{A}_n(t) u^{[0]}(t)))), \end{aligned}$$

and

$$\begin{aligned} \frac{d}{dt} u^{[3]}(t) &= \partial_{y_2} F_2(t) \circ (\dot{A}_n(t) u^{[2]}(t)) \\ &\quad + \partial_{y_2} F_2(t) \circ (\hat{A}_n(t) (\partial_{y_1} F_1(t) \circ (\dot{A}_n(t) u^{[1]}(t)))) \\ &\quad + \partial_{y_2} F_2(t) \circ (\hat{A}_n(t) (\partial_{y_1} F_1(t) \circ (\hat{A}_n(t) (\partial_{y_0} F_0(t) \circ (\dot{A}_n(t) u^{[0]}(t)))))) \\ &\quad + \partial_{y_1} F_2(t) \circ \partial_{y_0} F_0(t) \circ (\dot{A}_n(t) u^{[0]}(t)). \end{aligned}$$

For general $1 \leq \ell \leq k$,

$$\frac{d}{dt} u^{[\ell]}(t) = \partial_{y_{\ell-1}} F_{\ell-1}(t) \left(\dot{A}_n(t) u^{[\ell-1]}(t) + \hat{A}_n(t) \frac{d}{dt} u^{[\ell-1]}(t) \right) + \sum_{s=1}^{\ell-2} \partial_{y_s} F_{\ell-1}(t) \frac{d}{dt} u^{[s]}(t).$$

From these equations, one readily sees that the vector

$$\partial_{y_\ell} \phi(u^{[k]}(t), \dots, u^{[0]}(t)) \circ \frac{d}{dt} u^{[\ell]}(A(t))$$

appearing in (6.3) can be written as a summation of column vectors, in which each summand is of the form $w^r(t) = (w_i^r(t))_{i \in [n]}$ for some $0 \leq r \leq \ell-1$ that is defined by an iterative procedure through some functions $L^0, L^1, \dots, L^{r+1} \in C^\infty(\mathbb{R}^{2^\ell})$, whose partial derivatives of any nonzero orders are uniformly bounded. More precisely, starting from

$$w^{r,[0]}(t) = U^{[1]}(A(t)) \circ (\dot{A}_n(t) U^{[0]}(A(t))),$$

define

$$w^{r,[s]}(t) = U^{[s+1]}(A(t)) \circ (\hat{A}_n(t) w^{r,[s-1]}(t)), \quad \forall 1 \leq s \leq r, \quad (6.4)$$

where

$$U^{[s]}(X) := L^s(\hat{X}_n u^{[\ell-1]}(X), \dots, \hat{X}_n u^{[0]}(X), u^{[\ell-1]}(X), \dots, u^{[0]}(X)), \quad \forall 0 \leq s \leq r \quad (6.5)$$

and

$$\begin{aligned} U^{[r+1]}(X) &:= \partial_{y_\ell} \phi(u^{[k]}(X), \dots, u^{[0]}(X)) \\ &\quad \circ L^{r+1}(\hat{X}_n u^{[\ell-1]}(X), \dots, \hat{X}_n u^{[1]}(X), u^{[\ell-1]}(X), \dots, u^{[0]}(X)). \end{aligned} \quad (6.6)$$

Finally, set $w^r(t) = w^{r,[r]}(t)$.

6.3 Bounding the derivative of the interpolation

For $r \geq 0$, from the iteration (6.4) and expanding the Hadamard product,

$$\frac{1}{n} \sum_{i=1}^n \tilde{\mathbb{E}} w_i^r(t) = \frac{1}{n^{1+(r+1)/2}} \sum_{I \in \mathcal{I}_r} \tilde{\mathbb{E}} V_I(t) \quad (6.7)$$

for

$$\begin{aligned} V_I(t) &:= \left(\prod_{l=1}^r U_{i_{l+1}}^{[l+1]}(A(t)) (a_{i_{l+1}, i_l}(t) + n^{-1/2} z_{i_{l+1}, i_l}) \right) U_{i_1}^{[1]}(A(t)) \dot{a}_{i_1, i_0}(t) U_{i_0}^{[0]}(A(t)) \\ &= \left(\prod_{l=0}^{r+1} U_{i_l}^{[l]}(A(t)) \right) \left(\prod_{l=1}^r (a_{i_{l+1}, i_l}(t) + n^{-1/2} z_{i_{l+1}, i_l}) \right) \dot{a}_{i_1, i_0}(t), \end{aligned} \quad (6.8)$$

where \mathcal{I}_r is the collection of all $I = (i_0, i_1, \dots, i_r, i_{r+1}) \in [n]^{r+2}$ with $i_0 \neq i_1 \neq \dots \neq i_r \neq i_{r+1}$. Here we view each I as a directed graph of length $r+1$ with vertices $(i_s)_{0 \leq s \leq r+1}$ and edges $e_I(l) = (i_l, i_{l+1})$ for $0 \leq l \leq r$. For any $I \in \mathcal{I}_r$, disregard the direction, let Λ_I^0 be the collection of all $0 \leq l \leq r$ with $e_I(l) = e_I(0)$. Let $\mathcal{I}_r(s)$ be the collection of all graphs in \mathcal{I}_r so that disregard the direction there are exactly s many edges that appear once.

Proposition 6.2. *For any $I \in \mathcal{I}_r(s)$, we have that for any $0 < t < 1$,*

$$\int_0^1 \mathbb{E} |\tilde{\mathbb{E}} V_I(t)| dt \leq \begin{cases} \frac{C_s}{n^{(s+1)/2}}, & \text{if } |\Lambda_I^0| \leq 2, \\ \frac{C_s}{n^{s/2}}, & \text{if } |\Lambda_I^0| \geq 3, \end{cases}$$

where $\tilde{\mathbb{E}}$ is the expectation conditionally on u^0, Z and C_s is a universal constant independent of n .

To prove this proposition, we first establish a key lemma. For any $0 \leq b \leq r+1$, let $\mathcal{I}_r(s, b)$ be the collection of all $I \in \mathcal{I}_r(s)$ with $|\Lambda_I^0| = b$. Note that when $b = 1$, the set $\mathcal{I}_r(s, b)$ is nonempty for all $1 \leq s \leq r+1$ and when $2 \leq b \leq r+1$, the set $\mathcal{I}_r(s, b)$ is nonempty only if $0 \leq s \leq r+1-b$.

Lemma 6.1. *If $b = 1$, then for any $1 \leq s \leq r+1$,*

$$|\mathcal{I}_r(s, 1)| \leq C_{r,1,s} n^{\lfloor \frac{r+1-s}{2} \rfloor + s + 1}. \quad (6.9)$$

If $2 \leq b \leq r+1$, then for any $0 \leq s \leq r+1-b$,

$$|\mathcal{I}_r(s, b)| \leq C_{r,b,s} n^{\lfloor \frac{r+1-b-s}{2} \rfloor + s + 2}. \quad (6.10)$$

Here, $C_{r,b,s}$'s are universal constants independent of n .

Proof. For any graph $I \in \mathcal{I}_r(s, 1)$, let I' be the graph, in which we disregard both multiplicities and directions of the edges. See Figure 1 for examples. Observe that there are at most $\lfloor (r+1-s)/2 \rfloor$ many edges in I' that appear at least twice in I and the total number of edges of I' is at most $\lfloor (r+1-s)/2 \rfloor + s$. This implies that the total number of vertices of I' should be at most $\lfloor (r+1-s)/2 \rfloor + s + 1$ so there are at most $n^{\lfloor \frac{r+1-s}{2} \rfloor + s + 1}$ many such I' .

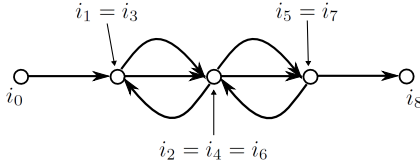
Since different I can correspond to the same I' (again we refer the reader to Figure 1 for examples), it remains to show that for each I' , there are at most a constant multiple (independent of n) of many different I that corresponds to I' . First fix such a possible I' and write $E(I')$ for the edge set of I' . Each edge of I' may correspond to an edge of multiplicity 1 in I or to an edge of multiplicity at least 2 in I (ignoring directions). We choose s edges in I' so that they correspond to

the multiplicity 1 edges in I . There are $\binom{|E(I')|}{s}$ ways to choose such s edges. Now, for the remaining $(|E(I')| - s)$ many edges, the multiplicities are at least 2 in I and they add up to $r + 1 - s$. To count how many possibilities there are, it is equivalent to find how many ways $r + 1 - s$ can be written as sum of $(|E(I')| - s)$ many integers which are at least 2, which in turn is bounded above by the number of ways to partition the integer $r + 1 - s$ as sum of $(|E(I')| - s)$ many positive integers, and it is well-known that the number of ways is $\binom{r-s}{|E(I')|-s-1}$. Moreover, each edge has two possible directions in I . Finally, each vertex in I' can correspond to several i_t ($0 \leq t \leq r + 1$) in I (see Figure 1). As there are at most $\lfloor (r + 1 - s)/2 \rfloor + s + 1$ vertices in I' , and each vertex can correspond to at most $r + 1$ many i_t 's, there are at most $(r + 1)^{\lfloor (r + 1 - s)/2 \rfloor + s + 1}$ many such correspondence. Therefore, the total number of I that an I' can correspond to is bounded above by

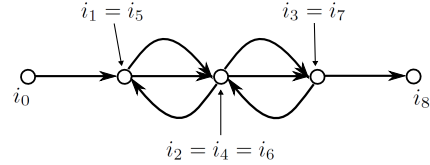
$$(r + 1)^{\lfloor (r + 1 - s)/2 \rfloor + s + 1} 2^{\binom{|E(I')|}{s} \cdot \binom{r-s}{|E(I')|-s-1}} \leq (r + 1)^{\lfloor (r + 1 - s)/2 \rfloor + s + 1} 2^{\lfloor (r + 1 - s)/2 \rfloor \cdot 2^{r-s}} =: C_{r,1,s}.$$

This proves (6.9).

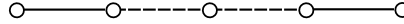
To prove (6.10), note that in this case, the edge $e_I(0)$ has multiplicity b , and hence there are at most $\lfloor (r + 1 - b - s)/2 \rfloor + 1$ many edges in I' that appear at least twice in I . Here, the latest $+1$ comes from $e_I(0)$. The remaining of the proof is similar to that of (6.9), and we omit the detail. \square



(a) One example of $I \in \mathcal{I}_r(s, 1)$



(b) Another example of $I \in \mathcal{I}_r(s, 1)$



(c) The corresponding undirected graph I'

Figure 1: (a) and (b) are two different directed graphs in $\mathcal{I}_r(s, 1)$ for $r = 7$ and $s = 2$. After we disregard the multiplicities and directions, they correspond to the same I' as shown in (c), where the solid edges correspond to the edges in I that appear only once and the dashed edges correspond to those in I with multiplicity ≥ 2 .

Proof of Proposition 6.2. Let $I \in \mathcal{I}_r(s)$ be fixed. Disregard the direction, let Λ_I^1 be the collection of all $l \notin \Lambda_I^0$ so that $e_I(l)$ appears exactly once in I and Λ_I^2 be the collection of all $l \notin \Lambda_I^0$ so that $e_I(l)$ appears more than once in I . In addition, for any $R \subseteq [r]$, set $\Lambda_I^0(R) = \Lambda_I^0 \cap R$, $\Lambda_I^1(R) = \Lambda_I^1 \cap R$, and $\Lambda_I^2(R) = \Lambda_I^2 \cap R$. Let $\Lambda_I(R) := \Lambda_I^1(R) \cup \{0\}$. From these and (6.8), after expanding

$$\prod_{l=1}^r (a_{i_{l+1}, i_l}(t) + n^{-1/2} z_{i_{l+1}, i_l}),$$

we can write

$$\tilde{\mathbb{E}}V_I(t) = \sum_{R \subseteq [r]} \frac{Z_{I,R}}{n^{|R^c|/2}} \tilde{\mathbb{E}} \left[U_I(A(t)) A_{I,R}(t) \left(a_{e_I(0)}(t)^{|\Lambda_I^0(R)|} \dot{a}_{e_I(0)}(t) \right) \left(\prod_{l \in \Lambda_I^1(R)} a_{e_I(l)}(t) \right) \right]$$

for

$$U_I(X) := \prod_{l=0}^{r+1} U_{i_l}^{[l]}(X), \quad A_{I,R}(t) := \prod_{l \in \Lambda_I^2(R)} a_{e_I(l)}(t), \quad Z_{I,R} := \prod_{l \in R^c} z_{i_{l+1}, i_l}.$$

Note that inside the expectation, the two parentheses are independent of $A_{I,R}(t)$, and each term in the second parentheses appears only once in I . From these, we can apply Proposition 5.1 and Lemmas A.1, A.2, and A.3 in Appendix to control $V_I(t)$. To see this, note that for any $0 < t < 1$ and $p \geq 1$,

$$(\mathbb{E}|a_{e_I(0)}(t)|^p)^{1/p} \leq C_p \quad (6.11)$$

and

$$(\mathbb{E}|\dot{a}_{e_I(0)}(t)|^p)^{1/p} \leq C_p \left(\frac{1}{\sqrt{t}} + \frac{1}{\sqrt{1-t}} \right)$$

for some universal constant C_p independent of t . Let $\mathbb{E}_{I,R}$ be the expectation only with respect to $a_{e_I(l)}(t)$ for all $l \in \Lambda_I(R)$. Using these bounds and Lemmas A.1, A.2, and A.3, we get that

$$\left| \mathbb{E}_{I,R} \left[U_I(A(t)) \left(a_{e_I(0)}(t)^{|\Lambda_I^0(R)|} \dot{a}_{e_I(0)}(t) \right) \left(\prod_{l \in \Lambda_I^1(R)} a_{e_I(l)}(t) \right) \right] \right|$$

is bounded above, up to an absolute constant independent of I and n , by

$$\left(\frac{1}{\sqrt{t}} + \frac{1}{\sqrt{1-t}} \right) \sum_{|\alpha|=s_0} \left(\int_0^1 \mathbb{E}_{I,R} [|\partial_{I,R}^\alpha U_I(A(t, \xi))|^2] d\xi \right)^{1/2},$$

where $A(t, \xi) = (a_{ii'}(t, \xi))_{i, i' \in [n]}$ is defined as $a_{ii'}(t, \xi) = \xi a_{ii'}(t)$ for all $i, i' \in [n]$ satisfying $(i, i') = e_I(l)$ or $(i', i) = e_I(l)$ for some $l \in \Lambda_I(R)$ and $a_{ii'}(t, \xi) = a_{ii'}(t)$ otherwise. Here,

$$s_0 = \begin{cases} |\Lambda_I^1(R)| + 2, & \text{if } |\Lambda_I^0(R)| = 0 \text{ by (A.4),} \\ |\Lambda_I^1(R)| + 1, & \text{if } |\Lambda_I^0(R)| = 1 \text{ by (A.1),} \\ |\Lambda_I^1(R)|, & \text{if } |\Lambda_I^0(R)| \geq 2 \text{ by (A.3).} \end{cases} \quad (6.12)$$

The summand in the above bound is over all $\alpha := (\alpha_l)_{l \in \Lambda_I(R)} \in (\{0\} \cup \mathbb{N})^{s_0}$, $|\alpha| := \sum_{l \in \Lambda_I(R)} \alpha_l$, and $\partial_{I,R}^\alpha$ is the partial derivative with respect to $x_{e_I(l)}$ of order α_l for all $l \in \Lambda_I(R)$. From this inequality, it follows that

$$\begin{aligned} & \int_0^1 \left| \mathbb{E} \left[U_I(A(t)) A_{I,R}(t) \left(a_{e_I(0)}(t)^{|\Lambda_I^0(R)|} \dot{a}_{e_I(0)}(t) \right) \left(\prod_{l \in \Lambda_I^1(R)} a_{e_I(l)}(t) \right) \right] \right| dt \\ & \leq C \sum_{|\alpha|=s_0} \int_0^1 \mathbb{E} \left[|A_{I,R}(t)| \left(\int_0^1 \mathbb{E}_{I,R} [|\partial_{I,R}^\alpha U_I(A(t, \xi))|^2] d\xi \right)^{1/2} \right] dt \\ & \leq C \int_0^1 \left(\mathbb{E} A_{I,R}(t)^2 \right)^{1/2} \sum_{|\alpha|=s_0} \left(\int_0^1 \mathbb{E} |\partial_{I,R}^\alpha U_I(A(t, \xi))|^2 d\xi \right)^{1/2} dt \\ & \leq C' \int_0^1 \sum_{|\alpha|=s_0} \left(\int_0^1 \mathbb{E} |\partial_{I,R}^\alpha U_I(A(t, \xi))|^2 d\xi \right)^{1/2} dt, \end{aligned} \quad (6.13)$$

for some constants C and C' independent of I and n , where the second inequality used the Cauchy-Schwarz inequality, while the third inequality used the Hölder inequality and the bounds (2.1) and (6.11). The last inequality can further be controlled as follows. From the product rule,

$$\partial_{I,R}^\alpha U_I = \sum_{\beta_l: l \in \Lambda_I(R)} \left(\prod_{l \in \Lambda_I(R)} \binom{\alpha_l}{\beta_l} \right) \left(\prod_{\ell=1}^{r+1} \partial_I^{\beta(\ell)} U_{i_\ell}^{[\ell]} \right),$$

where the summand is over all $\beta_l \in (\{0\} \cup \mathbb{N})^{r+2}$ satisfying that $\sum_{\ell=0}^{r+1} \beta_{l,\ell} = \alpha_l$ for $l \in \Lambda_I(R)$ and $\partial_I^{\beta(\ell)}$ is the partial derivative $\partial_{x_{e_I(l)}}^{\beta_{l,\ell}}$ for all $l \in \Lambda_I(R)$. Now, using the Minkowski and Hölder inequalities leads to

$$\begin{aligned} & \left(\mathbb{E} |\partial_{I,R}^\alpha U_I(A(t, \xi))|^2 \right)^{1/2} \\ & \leq \sum_{\beta_l: l \in \Lambda_I(R)} \left(\prod_{l \in \Lambda_I(R)} \binom{\alpha_l}{\beta_l} \right) \prod_{\ell=1}^{r+1} \left(\mathbb{E} |\partial_{I,R}^{\beta(\ell)} U_{i_\ell}^{[\ell]}(A(t, \xi))|^{2(r+2)} \right)^{1/2(r+2)}. \end{aligned}$$

From (6.5) and (6.6), note that any nonzero-order partial derivatives of $U^{[\ell]}$'s are uniformly bounded. From Proposition 5.1, each term on the right-hand side is bounded by

$$\left(\mathbb{E} |\partial_{I,R}^{\beta(\ell)} U_{i_\ell}^{[\ell]}(A(t, \xi))|^{2(r+2)} \right)^{1/2(r+2)} \leq \frac{\Gamma_\ell}{n^{\sum_{l \in \Lambda_I(R)} \beta_{l,\ell}/2}},$$

where Γ_ℓ is a constant independent of n . Consequently,

$$\begin{aligned} \left(\mathbb{E} |\partial_{I,R}^\alpha U_I(A(t, \xi))|^2 \right)^{1/2} & \leq \sum_{\beta_l: l \in \Lambda_I(R)} \left(\prod_{l \in \Lambda_I(R)} \binom{\alpha_l}{\beta_l} \right) \frac{\prod_{\ell=0}^{r+1} \Gamma_\ell}{n^{\sum_{\ell=0}^{r+1} \sum_{l \in \Lambda_I(R)} \beta_{l,\ell}/2}} \\ & = \sum_{\beta_l: l \in \Lambda_I(R)} \left(\prod_{l \in \Lambda_I(R)} \binom{\alpha_l}{\beta_l} \right) \frac{\prod_{\ell=0}^{r+1} \Gamma_\ell}{n^{\sum_{l \in \Lambda_I(R)} \alpha_l/2}} \\ & = \frac{1}{n^{s_0/2}} \sum_{\beta_l: l \in \Lambda_I(R)} \left(\prod_{l \in \Lambda_I(R)} \binom{\alpha_l}{\beta_l} \right) \prod_{\ell=0}^{r+1} \Gamma_\ell. \end{aligned}$$

Plugging this inequality into (6.13) and noting that Z is independent of A, G together with the bound (2.1) yield that

$$\frac{1}{n^{|R^c|/2}} \int_0^1 \left| \mathbb{E} \left[U_I(A(t)) Z_{I,R} A_{I,R}(t) \left(a_{e_I(0)}(t)^{|\Lambda_I^0(R)|} \dot{a}_{e_I(0)}(t) \right) \left(\prod_{l \in \Lambda_I^1(R)} a_{e_I(l)}(t) \right) \right] \right| dt \leq \frac{C''}{n^{|R^c|/2 + s_0/2}}.$$

Here, note that

$$\begin{aligned} |R^c| + |\Lambda_I^1(R)| &= (|\Lambda_I^0(R^c)| + |\Lambda_I^1(R^c)| + |\Lambda_I^2(R^c)|) + |\Lambda_I^1(R)| \\ &= |\Lambda_I^0(R^c)| + |\Lambda_I^1([r])| + |\Lambda_I^2(R^c)| \\ &\geq |\Lambda_I^1([r])|. \end{aligned}$$

Also, note that s is the number of edges in I that are crossed once disregard the direction. This implies that $|\Lambda_I^1([r])| \geq s - 1$ and that $|\Lambda_I^1([r])| = s$ if $|\Lambda_I^0(R)| \geq 1$ since $|\Lambda_I^0| \geq 1 + |\Lambda_I^0(R)| \geq 2$. Recall (6.12). If $|\Lambda_I^0(R)| = 0$, then

$$|R^c| + s_0 = |R^c| + |\Lambda_I^1(R)| + 2 \geq (s - 1) + 2 = s + 1;$$

if $|\Lambda_I^0(R)| = 1$, then

$$|R^c| + s_0 = |R^c| + |\Lambda_I^1(R)| + 1 \geq s + 1;$$

if $|\Lambda_I^0(R)| \geq 2$, then

$$|R^c| + s_0 = |R^c| + |\Lambda_I^1(R)| \geq s.$$

From these, if $|\Lambda_I^0| \leq 2$, then $|\Lambda_I^0(R)|$ can only be 0 or 1 for any $R \subseteq [r]$ and this implies that

$$\int_0^1 \mathbb{E}|\tilde{\mathbb{E}}V_I(t)|dt \leq \frac{C_s}{n^{(s+1)/2}}$$

and if $|\Lambda_I^0| \geq 3$, then $|\Lambda_I^0(R)|$ could be larger than 2 for some $R \subseteq [r]$ and hence,

$$\int_0^1 \mathbb{E}|\tilde{\mathbb{E}}V_I(t)|dt \leq \frac{C_s}{n^{s/2}}$$

for some constant $C_s > 0$. This completes our proof. \square

6.4 Proof of Proposition 6.1: first moment

Recall (6.7). Our proof will be completed once we establish that

$$\frac{1}{n^{1+(r+1)/2}} \sum_{I \in \mathcal{I}_r} \int_0^1 \mathbb{E}|\tilde{\mathbb{E}}V_I(t)|dt \leq \frac{C}{n^{1/2}}, \quad (6.14)$$

where C is an absolute constant independent of n . Recall from the definition of $\mathcal{T}_r(s, b)$ that when $b = 1$, $\mathcal{I}_r(0, b) = \emptyset$ and $\mathcal{I}_r(s, b) \neq \emptyset$ for all $1 \leq s \leq r + 1$ and that the set $\mathcal{I}_r(s, b)$ is nonempty only if $0 \leq s \leq r + 1 - b$. From these,

$$\begin{aligned} \sum_{I \in \mathcal{I}_r} \tilde{\mathbb{E}}V_I(t) &= \sum_{s=0}^{r+1} \sum_{I \in \mathcal{I}_r(s)} \tilde{\mathbb{E}}V_I(t) \\ &= \sum_{s=1}^{r+1} \sum_{I \in \mathcal{I}_r(s,1)} \tilde{\mathbb{E}}V_I(t) + \sum_{s=1}^{r-1} \sum_{I \in \mathcal{I}_r(s,2)} \tilde{\mathbb{E}}V_I(t) + \sum_{b=3}^{r+1} \sum_{s=0}^{r+1-b} \sum_{I \in \mathcal{I}_r(s,b)} \tilde{\mathbb{E}}V_I(t). \end{aligned} \quad (6.15)$$

In what follows, we let $C_{r,b,s}$ and $C'_{r,b,s}$ be absolute constants independent of n . Here, from the first case of Proposition 6.2 and Lemma 6.1, the first two summations can be controlled by

$$\begin{aligned} \frac{1}{n^{1+(r+1)/2}} \sum_{I \in \mathcal{I}_r(s,1)} \int_0^1 \mathbb{E}|\tilde{\mathbb{E}}V_I(t)|dt &\leq \frac{1}{n^{1+(r+1)/2}} \cdot C_{r,1,s} n^{\lfloor \frac{r+1-s}{2} \rfloor + s + 1} \cdot \frac{C'_{r,1,s}}{n^{(s+1)/2}} \\ &= C_{r,1,s} C'_{r,1,s} n^{\lfloor \frac{r+1-s}{2} \rfloor - \frac{(r+1-s)}{2} - \frac{1}{2}} \\ &\leq C_{r,1,s} C'_{r,1,s} n^{-1/2} \end{aligned}$$

for $1 \leq s \leq r + 1$ and

$$\begin{aligned} \frac{1}{n^{1+(r+1)/2}} \sum_{I \in \mathcal{I}_r(s,2)} \int_0^1 \mathbb{E}|\tilde{\mathbb{E}}V_I(t)|dt &\leq \frac{1}{n^{1+(r+1)/2}} \cdot C_{r,2,s} n^{\lfloor \frac{r-1-s}{2} \rfloor + s + 2} \cdot \frac{C'_{r,2,s}}{n^{(s+1)/2}} \\ &= C_{r,2,s} C'_{r,2,s} n^{\lfloor \frac{r-1-s}{2} \rfloor - \frac{(r-1-s)}{2} - \frac{1}{2}} \\ &\leq C_{r,2,s} C'_{r,2,s} n^{-1/2} \end{aligned}$$

for $0 \leq s \leq r-1$. To control the second summation, from the second case of Proposition 6.2 and Lemma 6.1,

$$\begin{aligned} \frac{1}{n^{1+(r+1)/2}} \sum_{I \in \mathcal{I}_r(s,b)} \int_0^1 \mathbb{E} |\tilde{\mathbb{E}} V_I(t)| dt &\leq \frac{1}{n^{1+(r+1)/2}} \cdot C_{r,b,s} n^{\lfloor \frac{r+1-b-s}{2} \rfloor + s + 2} \cdot \frac{C'_{r,b,s}}{n^{s/2}} \\ &= C_{r,b,s} C'_{r,b,s} n^{\lfloor \frac{r+1-b-s}{2} \rfloor - \frac{r+1-b-s}{2} - (\frac{b}{2}-1)} \\ &\leq C_{r,b,s} C'_{r,b,s} n^{-(\frac{b}{2}-1)} \end{aligned}$$

for $0 \leq s \leq r+1-b$. Plugging these into (6.15) yields (6.14) and this completes our proof.

6.5 Proof of Proposition 6.1: second moment

Our approach is the same as that for the first moment. Set

$$\Psi_{k,n}(t) = \tilde{\mathbb{E}} \Phi_{k,n}(A(t))^2, \quad 0 < t < 1.$$

Note that

$$\mathbb{E} |\tilde{\mathbb{E}} \Phi_{k,n}(A)^2 - \tilde{\mathbb{E}} \Phi_{k,n}(G)^2| = \mathbb{E} |\Psi_{k,n}(1) - \Psi_{k,n}(0)|.$$

To control this expectation, we again consider the derivative

$$\frac{d}{dt} \Psi_{k,n}(t) = 2 \tilde{\mathbb{E}} \Phi_{k,n}(A(t)) \frac{d}{dt} \Phi_{k,n}(A(t)).$$

Again, our goal would be to show that

$$\int_0^1 \mathbb{E} |\tilde{\mathbb{E}} \Psi'_{k,n}(t)| dt$$

is bounded above, up to an absolute constant, by $n^{-1/2}$. To see this, recall from Section 6.2 that $\Phi'_{k,n}(t)$ can be written as a summation, in which each summand is of the form

$$\frac{1}{n} \sum_{i=1}^n w_i^r(t)$$

for some $0 \leq r \leq k-1$. In a similar manner, from (6.7) and (6.8), $\mathbb{E} \Psi'_{k,n}(t)$ can also be written as a sum, in which each term is equal to

$$\tilde{\mathbb{E}} \Phi_{k,n}(A(t)) \left(\frac{1}{n} \sum_{i=1}^n w_i^r(t) \right) = \frac{1}{n^{2+(r+1)/2}} \sum_{i=1}^n \sum_{I \in \mathcal{I}_r} \tilde{\mathbb{E}} \bar{V}_{i,I}(t),$$

where

$$\bar{V}_{i,I}(t) := \tilde{\mathbb{E}} U_i^{[r+2]}(A(t)) \left(\prod_{l=0}^{r+1} U_i^{[l]}(A(t)) \right) \left(\prod_{l=1}^r (a_{i_{l+1}, i_l}(t) + n^{-1/2} z_{i_{l+1}, i_l}(t)) \right) \dot{a}_{i_1, i_0}(t)$$

for

$$U_i^{[r+2]}(A(t)) := \phi(u_i^{[k]}(t), \dots, u_i^{[0]}(t)), \quad \forall i \in [n].$$

Here, $\bar{V}_{i,I}(t)$ is essentially the same as $V_I(t)$ (see (6.8)) except that it contains one extra term $U_i^{[r+2]}(A(t))$. In view of the proof of Proposition 6.2, an identical argument implies that for any $i \in [n]$ and $I \in \mathcal{I}_r(s)$,

$$\int_0^1 \mathbb{E} |\tilde{\mathbb{E}} \bar{V}_{i,I}(t)| dt \leq \begin{cases} \frac{C'_s}{n^{(s+1)/2}}, & \text{if } |\Lambda_I^0| \leq 2, \\ \frac{C'_s}{n^{s/2}}, & \text{if } |\Lambda_I^0| \geq 3, \end{cases}$$

where C'_s is a universal constant independent of n and i . Consequently, as in the proof of (6.14), it follows that

$$\frac{1}{n^{1+(r+1)/2}} \sum_{I \in \mathcal{I}_r} \int_0^1 \mathbb{E} |\tilde{\mathbb{E}} \bar{V}_{i,I}(t)| dt \leq \frac{C}{n^{1/2}}$$

for some constant C independent of n and i and the same inequality remains valid after taking $n^{-1} \sum_{i=1}^n$,

$$\frac{2}{n^{1+(r+1)/2}} \sum_{i=1}^n \sum_{I \in \mathcal{I}_r} \int_0^1 \mathbb{E} |\tilde{\mathbb{E}} \bar{V}_{i,I}(t)| dt \leq \frac{C}{n^{1/2}},$$

which completes our proof.

7 Proof of Theorem 2.2

We establish the proof of Theorem 2.2. For notational convenience, if a_n and b_n are two random variables, we denote

$$a_n \asymp b_n$$

if $|a_n - b_n| \rightarrow 0$ in probability; if they are n -dimensional random vectors, then this notation means that in probability,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \|a_n - b_n\|_2^2 = 0.$$

To begin with, recall from Lemma 3.4 and Theorem 3.1 that due to the Lipschitz property of the functions $(F_k)_{k \geq 0}$, the AMP orbit defined in Definition 2.1 is uniformly square-integrable and the average along the Gaussian AMP orbit is concentrated with respect to $\tilde{\mathbb{E}}$. Here, in the setting of Definition 2.2, since both f_ℓ and its first-order derivatives are Lipschitz, an identical argument also allows to show that $(v^{[k]})_{k \geq 0}$ is uniformly squared-integrable and when $X = G$, its average along the orbit is self-averaged with respect to $\tilde{\mathbb{E}}$. More precisely, for any Lipschitz $\phi \in C(\mathbb{R}^{k+1})$,

$$\frac{1}{n} \sum_{i \in [n]} \phi(v^{[k]}(G), \dots, v^{[0]}(G)) \asymp \frac{1}{n} \tilde{\mathbb{E}} \sum_{i \in [n]} \phi(v^{[k]}(G), \dots, v^{[0]}(G)). \quad (7.1)$$

Denote by $(v^{G,[k]})_{k \geq 0}$ the AMP orbit in Definition 2.2 with the replacement of $b_{k,j}$ by

$$b_{k,j}^G := \frac{1}{n} \tilde{\mathbb{E}} \sum_{i \in [n]} \frac{\partial f_k}{\partial u_i^{[j]}}(v^{[k]}(G), \dots, v^{[0]}(G))$$

and initialization $v^{G,[0]} = u^0$.

Lemma 7.1. For any Lipschitz $\phi \in C(\mathbb{R}^{k+1})$,

$$\frac{1}{n} \sum_{i \in [n]} \phi(v_i^{G,[k]}(G), \dots, v_i^{G,[0]}(G)) \asymp \frac{1}{n} \sum_{i \in [n]} \phi(v_i^{G,[k]}(A), \dots, v_i^{G,[0]}(A)).$$

Proof. Consider the initialization

$$u^{[0]}(X) = v^{[0]}(X) = u^0.$$

Set

$$\begin{aligned} u^{[1]}(X) &= 0, \\ u^{[2]}(X) &= f_0(u^{[0]}(X)), \\ u^{[3]}(X) &= \hat{X}_n u^{[2]}(X) = \hat{X}_n f_0(u^{[0]}(X)) \end{aligned}$$

and for $\ell \geq 1$, set

$$\begin{aligned} u^{[3\ell+1]}(X) &= 0, \\ u^{[3\ell+2]}(X) &= f_\ell(u^{[3\ell]}(X), u^{[3(\ell-1)]}(X), \dots, u^{[3]}(X), u^{[0]}(X)), \\ u^{[3\ell+3]}(X) &= \hat{X}_n u^{[3\ell+2]}(X) - \sum_{j=1}^{\ell} b_{\ell,j}^G u^{[3j-1]}(X). \end{aligned}$$

The main feature of this construction is that $u^{[3\ell]}(X) = v^{G,[\ell]}(X)$ for all $\ell \geq 0$. Note that $b_{\ell,j}^G$ depends only on u^0, Z and it is uniformly bounded. Although Definition 2.1 assumes that F_k 's are nonrandom, with no essential changes to the proof, Theorem 2.1 indeed extends to randomized F_k 's that are dependent only on u^0, Z and the Lipschitz constants of F_k 's are bounded by some constants independent of u^0, Z . Hence, the assertion follows by applying Theorem 2.1 to $(u^{[\ell]})_{\ell \geq 0}$ and $\phi(u^{[3k]}, u^{[3(k-1)]}, \dots, u^{[0]})$. \square

Lemma 7.2. For any $k \geq 0$,

$$v^{[k]}(G) \asymp v^{G,[k]}(G).$$

Proof. We argue by induction. Obviously the assertion is valid for $k = 0$. Assume that there exists some $k' \geq 0$ such that it is also valid for all $0 \leq k \leq k'$. From the triangle inequality,

$$\begin{aligned} & \|v^{[k'+1]}(G) - v^{G,[k'+1]}(G)\|_2 \\ & \leq \|\hat{G}_n\|_2 \|f_{k'}(v^{[k']} (G), \dots, v^{[0]}(G)) - f_k(v^{G,[k']} (G), \dots, v^{G,[0]}(G))\|_2 \\ & \quad + \sum_{j=1}^{k'} |b_{k',j}(G)| \|f_{j-1}(v^{[j-1]}(G), \dots, v^{[0]}(G)) - f_{j-1}(v^{G,[j-1]}(G), \dots, v^{G,[0]}(G))\|_2 \\ & \quad + \sum_{j=1}^{k'} |b_{k',j}(G) - b_{k',j}^G| \|f_{j-1}(v^{G,[j-1]}(G), \dots, v^{G,[0]}(G))\|_2. \end{aligned}$$

By induction hypothesis, Lemma 3.3, and noting that the first-order partial derivatives of f_j 's are uniformly bounded, the first two terms after dividing by \sqrt{n} converge to zero in probability. As for the last term, note that (7.1) implies that $b_{k',j}^G \asymp b_{k',j}(G)$ for all $1 \leq j \leq k'$. Also, note that from the relation, $u^{[3\ell]} = v^{G,[\ell]}$, in the proof of Lemma 7.1, the Lipschitz property of f_{j-1} and Lemmas

3.3 and 3.4 (here, again Lemma 3.4 is valid despite of the fact that F_k 's are dependent on Z and u^0) imply

$$\sup_{n \geq 1} \frac{1}{\sqrt{n}} \mathbb{E} \|f_{j-1}(v^{G,[j-1]}(G), \dots, v^{G,[0]}(G))\|_2 < \infty. \quad (7.2)$$

Hence, the third term also vanishes in probability and this validates the announced result. \square

We are ready to prove Theorem 2.2, namely, for any $k \geq 0$ and Lipschitz $\phi \in C(\mathbb{R}^{k+1})$,

$$\frac{1}{n} \sum_{i \in [n]} \phi(v_i^{[k]}(G), \dots, v_i^{[0]}(G)) \asymp \frac{1}{n} \sum_{i \in [n]} \phi(v_i^{[k]}(A), \dots, v_i^{[0]}(A)). \quad (7.3)$$

We argue by induction. Evidently this is valid for $k = 0$. Assume that there exists some $k' \geq 0$ such that it is also valid for all $0 \leq k \leq k'$. From Lemmas 7.1 and 7.2,

$$\begin{aligned} \frac{1}{n} \sum_{i \in [n]} \phi(v_i^{[k'+1]}(G), \dots, v_i^{[0]}(G)) &\asymp \frac{1}{n} \sum_{i \in [n]} \phi(v_i^{G,[k'+1]}(G), \dots, v_i^{G,[0]}(G)) \\ &\asymp \frac{1}{n} \sum_{i \in [n]} \phi(v_i^{G,[k'+1]}(A), \dots, v_i^{G,[0]}(A)). \end{aligned}$$

We claim that

$$\frac{1}{n} \sum_{i \in [n]} \phi(v_i^{G,[k'+1]}(A), \dots, v_i^{G,[0]}(A)) \asymp \frac{1}{n} \sum_{i \in [n]} \phi(v_i^{[k'+1]}(A), \dots, v_i^{[0]}(A)).$$

If this is valid, then (7.3) is also true for $k + 1$ and this would complete our proof. It suffices to show that

$$v^{G,[k]}(A) \asymp v^{[k]}(A), \quad \forall 0 \leq k \leq k' + 1.$$

Easy to see that this is valid if $k = 0$. Assume that there exists some $0 \leq k'' \leq k'$ such that this equation holds for all $0 \leq k \leq k''$. Write

$$\begin{aligned} &\|v^{[k''+1]}(A) - v^{G,[k''+1]}(A)\|_2 \\ &\leq \|\hat{A}_n\|_2 \|f_{k''}(v^{[k'']}(A), \dots, v^{[0]}(A)) - f_{k''}(v^{G,[k'']}(A), \dots, v^{G,[0]}(A))\| \\ &+ \sum_{j=1}^{k''} |b_{k'',j}| \|f_{j-1}(v^{[j-1]}(A), \dots, v^{[0]}(A)) - f_{j-1}(v^{G,[j-1]}(A), \dots, v^{G,[0]}(A))\|_2 \\ &+ \sum_{j=1}^{k''} |b_{k'',j}(A) - b_{k'',j}^G| \|f_{j-1}(v^{G,[j-1]}(A), \dots, v^{G,[0]}(A))\|_2. \end{aligned}$$

Here, from the induction hypothesis, after dividing by \sqrt{n} , the first two lines vanish in probability by using the fact that $b_{k'',j}$ is uniformly bounded and Lemma 3.3. As for the last one, write

$$b_{k'',j}(A) - b_{k'',j}^G = (b_{k'',j}(A) - b_{k'',j}(G)) + (b_{k'',j}(G) - b_{k'',j}^G).$$

Note that (7.1) implies $b_{k'',j}^G \asymp b_{k'',j}(G)$, while the induction hypothesis of (7.3) implies that $b_{k'',j}^G \asymp b_{k'',j}^A$. Hence, $b_{k'',j}(A) \asymp b_{k'',j}^G$. This and (7.2) imply that the third line also vanishes in probability. Hence, $v^{[k''+1]}(A) \asymp v^{G,[k''+1]}(A)$ and this completes the proof of our claim.

8 Proof of Theorem 2.3

We establish the proof of Theorem 2.3. Our strategy is to approximate the principal eigenvector by the power method. In view of this, it is essentially a special case of the generalized AMP in Definition 2.1. Once this is done, universality would follow by an analogous argument as that for Theorem 2.3. Again, we adapt the notation $a_n \asymp b_n$ from Section 7.

8.1 Power method

The well-known power method states that if the principal eigenvalue stays a gap away from the other eigenvalues, then one can generate the principal eigenvector via an iteration procedure.

Lemma 8.1 (Power method). *Let $Y \in M_n(\mathbb{R})$ and $y \in \mathbb{R}^n$ with $\|y\|_2 = 1$. Let $\lambda_1 \geq \dots \geq \lambda_n$ be the eigenvalues of Y satisfying $\lambda_1 \geq \max_{2 \leq r \leq n} |\lambda_r|$ and y^1 be the normalized eigenvector associated to λ_1 . If $\lambda_1 \neq 0$ and $y \not\propto y^1$, then for any $d \geq 1$,*

$$\left\| \frac{Y^d y}{\|Y^d y\|_2} - \text{sign}(\langle y^1, y \rangle) y^1 \right\|_2 \leq \frac{1}{|\langle y^1, y \rangle|} \max_{2 \leq r \leq n} \left| \frac{\lambda_r}{\lambda_1} \right|^d. \quad (8.1)$$

Proof. Let y^1, \dots, y^n be the orthonormal eigenvectors associated to $\lambda_1, \dots, \lambda_n$. Write

$$y = c_1 y^1 + \dots + c_n y^n,$$

where $c = (c_1, \dots, c_n) \in \mathbb{R}^n$ satisfies $\|c\|_2 = \|y\|_2 = 1$. Note that

$$Y^d y = c_1 \lambda_1^d y^1 + \dots + c_n \lambda_n^d y^n$$

and

$$\|Y^d y\|_2 = (c_1^2 \lambda_1^{2d} + \dots + c_n^2 \lambda_n^{2d})^{1/2}. \quad (8.2)$$

From these,

$$\begin{aligned} \frac{Y^d y}{\|Y^d y\|_2} &= \frac{c_1 \lambda_1^d y^1 + \dots + c_n \lambda_n^d y^n}{(c_1^2 \lambda_1^{2d} + \dots + c_n^2 \lambda_n^{2d})^{1/2}} \\ &= \text{sign}(c_1) \frac{|c_1 \lambda_1^d| y^1 + \text{sign}(c_1) \sum_{r=2}^n c_r \lambda_r^d y^r}{(c_1^2 \lambda_1^{2d} + \dots + c_n^2 \lambda_n^{2d})^{1/2}} \\ &= \text{sign}(c_1) \frac{y^1 + \text{sign}(c_1) \sum_{r=2}^n \frac{c_r \lambda_r^d}{|c_1 \lambda_1^d|} y^r}{\left(1 + \sum_{r=2}^n \frac{c_r^2}{c_1^2} \left(\frac{\lambda_r}{\lambda_1}\right)^{2d}\right)^{1/2}}. \end{aligned}$$

If we denote

$$\Pi = \sum_{r=2}^n \frac{c_r^2}{c_1^2} \left(\frac{\lambda_r}{\lambda_1}\right)^{2d},$$

then

$$\left\| \frac{Y^d y}{\|Y^d y\|_2} - \text{sign}(c_1) y^1 \right\|_2 = \left(\frac{(1 - \sqrt{1 + \Pi})^2 + \Pi}{1 + \Pi} \right)^{1/2} \leq \left(\frac{\Pi^2 + \Pi}{1 + \Pi} \right)^{1/2} = \Pi^{1/2},$$

where we used that $\sqrt{1+x} - 1 \leq x$ for $x \geq 0$. Now, the assertion follows by

$$\Pi \leq \frac{1}{c_1^2} \max_{2 \leq r \leq n} \left| \frac{\lambda_r}{\lambda_1} \right|^{2d} \sum_{r=2}^n c_r^2 \leq \frac{1}{|\langle y, y^1 \rangle|^2} \max_{2 \leq r \leq n} \left| \frac{\lambda_r}{\lambda_1} \right|^{2d}.$$

□

We continue to show that the AMP orbits in Definition 2.1 initialized by the principal eigenvector and the power method can be as close as we want by increasing the power iteration d . Let $\varepsilon \in (0, 1)$. For any $d \geq 1$, let

$$u^{A,\varepsilon,d}(X) = \frac{\sqrt{n}\hat{X}_n^d u^0}{\tilde{\mathbb{E}}\|\hat{A}_n^d u^0\|_2 + \sqrt{n}\varepsilon}$$

and

$$u^{G,\varepsilon,d}(X) = \frac{\sqrt{n}\hat{X}_n^d u^0}{\tilde{\mathbb{E}}\|\hat{G}_n^d u^0\|_2 + \sqrt{n}\varepsilon}.$$

Let $\lambda_1(\hat{X}_n) \geq \dots \geq \lambda_n(\hat{X}_n)$ be the eigenvalues of \hat{X}_n . Let $\psi^1(\hat{X}_n)$ be the principal eigenvector of \hat{X}_n with $\|\psi^1(\hat{X}_n)\|_2 = \sqrt{n}$. Recall the vector ψ defined through (2.4). Let $v^{[k]}$, $v^{A,\varepsilon,d,[k]}$, and $v^{G,\varepsilon,d,[k]}$ be three AMP orbits that are defined via Definition 2.1 associated to the initializations ψ , $u^{A,\varepsilon,d}$, and $u^{G,\varepsilon,d}$, respectively. For any Lipschitz $\phi \in C(\mathbb{R}^{k+1})$, denote by $\phi_{k,n}^\psi$, $\phi_{k,n}^{u^{A,\varepsilon,d}}$, and $\phi_{k,n}^{u^{G,\varepsilon,d}}$ the averages of ϕ over these AMP orbits, respectively.

Lemma 8.2. *Assume that (2.5) and (2.6) are valid. Suppose that $\phi \in C(\mathbb{R}^{k+1})$ is Lipschitz. We have that in probability,*

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} \lim_{d \rightarrow \infty} \lim_{n \rightarrow \infty} |\phi_{k,n}^\psi(A) - \phi_{k,n}^{u^{A,\varepsilon,d}}(A)| &= 0, \\ \lim_{\varepsilon \downarrow 0} \lim_{d \rightarrow \infty} \lim_{n \rightarrow \infty} |\phi_{k,n}^\psi(G) - \phi_{k,n}^{u^{G,\varepsilon,d}}(G)| &= 0. \end{aligned}$$

Proof. We only need to establish the first equality. First of all, we claim that in probability

$$\lim_{\varepsilon \downarrow 0} \lim_{d \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \|\psi(A) - u^{A,\varepsilon,d}(A)\|_2^2 = 0.$$

Define

$$u^{\varepsilon,d}(X) := \frac{\sqrt{n}\hat{X}_n^d u^0}{\|\hat{X}_n^d u^0\|_2 + \varepsilon\sqrt{n}}.$$

From (2.5), (2.6), and (8.2), there exist $0 < \delta < 1$ and $\delta', \delta'' > 0$ such that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \max_{2 \leq r \leq n} \left| \frac{\lambda_r(\hat{A}_n)}{\lambda_1(\hat{A}_n)} \right| &\leq \delta, \\ \liminf_{n \rightarrow \infty} \frac{1}{n} |\langle \psi^1(\hat{A}_n), u^0 \rangle| &\geq \delta', \end{aligned}$$

and

$$\liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|\hat{A}_n^d u^0\|_2 \geq \liminf_{n \rightarrow \infty} \frac{1}{n} |\langle \psi^1(\hat{A}_n), u^0 \rangle| \lambda_1(\hat{A}_n)^d \geq \delta'(1 + \delta'')^d.$$

Note that (8.1) implies that

$$\left\| \frac{u^{\varepsilon,d}(A)}{\sqrt{n}} - \frac{\psi(A)}{\sqrt{n}} \right\|_2 \leq \frac{n}{|\langle \psi^1(\hat{A}_n), u^0 \rangle|} \max_{2 \leq r \leq n} \left| \frac{\lambda_r(\hat{A}_n)}{\lambda_1(\hat{A}_n)} \right|^d + \frac{\varepsilon}{\frac{\|\hat{A}_n^d u^0\|_2}{\sqrt{n}} + \varepsilon}.$$

From these inequalities, as long as d is large enough such that

$$\frac{\delta^d}{\delta'} + \frac{\varepsilon}{\delta'(1 + \delta'')^d + \varepsilon} < \varepsilon,$$

we have

$$\limsup_{n \rightarrow \infty} \left\| \frac{u^{\varepsilon,d}(A)}{\sqrt{n}} - \frac{\psi(A)}{\sqrt{n}} \right\|_2 \leq \varepsilon. \quad (8.3)$$

Next, note that

$$\begin{aligned} \frac{1}{\sqrt{n}} \|u^{\varepsilon,d}(A) - u^{A,\varepsilon,d}(A)\|_2 &= \frac{\|\hat{A}_n^d u^0\|_2}{\sqrt{n}} \left| \frac{\|\hat{A}_n^d u^0\|_2 / \sqrt{n} - \tilde{\mathbb{E}} \|\hat{A}_n^d u^0\|_2 / \sqrt{n}}{(\|\hat{A}_n^d u^0\|_2 / \sqrt{n} + \varepsilon)(\tilde{\mathbb{E}} \|\hat{A}_n^d u^0\|_2 / \sqrt{n} + \varepsilon)} \right| \\ &\leq \frac{\|\hat{A}_n^d u^0\|_2}{\varepsilon^2 \sqrt{n}} \left| \frac{\|\hat{A}_n^d u^0\|_2}{\sqrt{n}} - \frac{\tilde{\mathbb{E}} \|\hat{A}_n^d u^0\|_2}{\sqrt{n}} \right|. \end{aligned}$$

From Lemmas 3.3 and 8.3 (established below), in probability,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \|u^{\varepsilon,d}(A) - u^{A,\varepsilon,d}(A)\|_2^2 = 0.$$

Combining this with (8.3), our claim follows.

Now to establish the first assertion, note that similar to Proposition 3.1, the assumption that both f_k and its first-order partial derivative are Lipschitz ensures that the average $\phi_{k,n}(X)$ of ϕ along the AMP orbit $(v^{[k]})_{k \geq 1}$ is Lipschitz with respect to its initialization. The proof of this fact follows directly from the same proof as that of Proposition 3.1. Hence, it suffices to show that in probability,

$$\lim_{\varepsilon \downarrow 0} \lim_{d \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n} \|v^{[k]}(A) - v^{A,\varepsilon,d,[k]}(A)\|_2 = 0, \quad \forall k \geq 0.$$

When $k = 0$, this is valid by our claim. Assume that this is also valid for all $0 \leq k \leq k'$ for some k' . We prove that this is also valid for $k' + 1$. To see this, we use triangle inequality to write

$$\begin{aligned} &\|v^{[k'+1]}(A) - v^{A,\varepsilon,d,[k'+1]}(A)\|_2 \\ &\leq \|\hat{A}_n\|_2 \|f_{k'}(v^{[k']}(A), \dots, v^{[0]}(A)) - f_{k'}(v^{A,\varepsilon,d,[k']}(A), \dots, v^{A,\varepsilon,d,[0]}(A))\| \\ &+ \sum_{j=1}^{k'} |b_{k',j}(A)| \|f_{j-1}(v^{[j-1]}(A), \dots, v^{[0]}(A)) - f_{j-1}(v^{A,\varepsilon,d,[j-1]}(A), \dots, v^{A,\varepsilon,d,[0]}(A))\|_2 \\ &+ \sum_{j=1}^{k'} |b_{k',j}(A) - b_{k',j}^{A,\varepsilon,d}(A)| \|f_{j-1}(v^{A,\varepsilon,d,[j-1]}(A), \dots, v^{A,\varepsilon,d,[0]}(A))\|_2, \end{aligned}$$

where

$$b_{k',j}^{A,\varepsilon,d}(A) := \frac{1}{n} \sum_{i \in [n]} \frac{\partial f_{k'}}{\partial v_i^{[j]}}(v_i^{A,\varepsilon,d,[k']}(A), \dots, v_i^{A,\varepsilon,d,[0]}(A)).$$

Here by Lemma 3.3, Lemma 3.4, and the Lipschitz property of f_k and its first-order derivative, these terms vanish in probability from the induction hypothesis. \square

At the end of this subsection, we establish the following lemma, which was used in the above proof.

Lemma 8.3.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \|\hat{A}_n^d u^0\|_2 - \|\hat{G}_n^d u^0\|_2^2 = 0, \quad (8.4)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \|\hat{G}_n^d u^0\|_2 - \tilde{\mathbb{E}} \|\hat{G}_n^d u^0\|_2^2 = 0. \quad (8.5)$$

Proof. We establish (8.4) first. Consider the AMP orbit

$$\begin{aligned} u^{[0]}(X) &= u^0, \\ u^{[k+1]}(X) &= \hat{X}_n u^{[k]}(X) = \hat{X}_n^k u^0, \quad k \geq 0. \end{aligned}$$

Note that $u^{[d+1]}(G) = \hat{G}_n^d u^0$ and $u^{[d+1]}(A) = \hat{A}_n^d u^0$. Recall that Theorem 2.1 implies that if $\phi \in C(\mathbb{R}^{k+1})$ is Lipschitz, then $\lim_{n \rightarrow \infty} |\phi_{k,n}(A) - \phi_{k,n}(G)| = 0$ in probability. However, we can not apply this result directly to $\|u^{[d+1]}(X)\|_2^2$ since each term inside the summation is not Lipschitz. To this end, we adapt a truncation argument. For any $M > 1$, let $\rho \in C^1(\mathbb{R})$ be uniformly bounded by $2M^2$ and satisfy that $\rho(x) = x^2$ on $[-M, M]$ and $\rho(x) = (M+1)^2$ for $x \notin [-(M+1), M+1]$. Note that for any $x \in \mathbb{R}$,

$$x^2 - \rho(x) = (x^2 - \rho(x))1_{\{|x| \in [M, M+1]\}} + (x^2 - (M+1)^2)1_{\{|x| \in (M+1, \infty)\}},$$

which implies that

$$|x^2 - \rho(x)|^2 \leq 4(x^4 + 9M^4)1_{\{|x| \geq M\}}.$$

It follows that from the Jensen and Cauchy-Schwarz inequalities,

$$\begin{aligned} & \mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n (u_i^{[d+1]}(G)^2 - \rho(u_i^{[d+1]}(G))) \right|^2 \\ & \leq \frac{4}{n} \mathbb{E} \sum_{i=1}^n ((\mathbb{E}(u_i^{[d+1]}(G))^8)^{1/2} \mathbb{P}(|u_i^{[d+1]}(G)| \geq M)^{1/2} + 9M^4 \mathbb{P}(|u_i^{[d+1]}(G)| \geq M)). \end{aligned}$$

Here, from Proposition 5.2, there exists a constant $C > 0$ independent of n, i , and $M > 1$,

$$\mathbb{E}|u_i^{[d+1]}(G)|^8 \leq C$$

and

$$\mathbb{P}(|u_i^{[d+1]}(G)| \geq M) \leq \frac{\mathbb{E}|u_i^{[d+1]}(G)|^8}{M^8} \leq \frac{C}{M^8}.$$

Consequently,

$$\mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n (u_i^{[d+1]}(G)^2 - \rho(u_i^{[d+1]}(G))) \right|^2 \leq \frac{40C}{M^4}.$$

Similarly, the same inequality is valid for A . Now from Theorem 2.1 and the dominated convergence theorem,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n (\rho(u_i^{[d+1]}(G)) - \rho(u_i^{[d+1]}(A))) \right|^2 = 0.$$

Hence, we arrive at

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left| \|u^{[d+1]}(G)\|_2^2 - \|u^{[d+1]}(A)\|_2^2 \right|^2 \leq \frac{640C}{M^4}.$$

Since this is valid for all $M > 1$, this limit is indeed equal to zero. From this, since both $n^{-1/2}\|u^{[d+1]}(G)\|_2$ and $n^{-1/2}\|u^{[d+1]}(A)\|_2$ are uniformly square-integrable by Lemma 3.4, the assertion (8.4) follows. The proof of (8.5) can be established by using Theorem 3.1 and an identical truncation argument. As this part of the proof does not involve additional complications, we omit the details here. \square

8.2 Main argument

We are ready to establish the proof of Theorem 2.3. Recall the AMP orbits $(v^{A,\varepsilon,d,[k]})_{k \geq 0}$ and $(v^{G,\varepsilon,d,[k]})_{k \geq 0}$ from last subsection. Define

$$v^{A,\varepsilon,d,[-d]}(X) = \frac{\sqrt{n}u^0}{\tilde{\mathbb{E}}\|\hat{A}_n^d u^0\|_2 + \sqrt{n}\varepsilon},$$

$$v^{G,\varepsilon,d,[-d]}(X) = \frac{\sqrt{n}u^0}{\tilde{\mathbb{E}}\|\hat{G}_n^d u^0\|_2 + \sqrt{n}\varepsilon}.$$

For $-d \leq k \leq -1$, set

$$v^{A,\varepsilon,d,[k+1]}(X) = \hat{X}_n v^{A,\varepsilon,d,[k]}(X),$$

$$v^{G,\varepsilon,d,[k+1]}(X) = \hat{X}_n v^{G,\varepsilon,d,[k]}(X).$$

Note that $v^{A,\varepsilon,d,[0]}(X) = u^{A,\varepsilon,d}(X)$ and $v^{G,\varepsilon,d,[0]}(X) = u^{G,\varepsilon,d}(X)$. This implies that $(v^{A,\varepsilon,d,[k]})_{k \geq -d}$ and $(v^{G,\varepsilon,d,[k]})_{k \geq -d}$ are again AMP orbits with the initializations $v^{A,\varepsilon,d,[-d]}(X)$ and $v^{G,\varepsilon,d,[-d]}(X)$. The key feature of this construction is that the initializations are independent of X and its norm is bounded above by $\|u^0\|_2/\varepsilon$. Hence, the assumption (2.1) is satisfied with possibly a larger σ . From an identical argument as that of Theorem 2.2, we see that the AMP orbit $(v^{A,\varepsilon,d,[k]})_{k \geq -d}$ satisfies universality. In particular, for any Lipschitz $\phi \in C(\mathbb{R}^{k+1})$, this implies that

$$\phi_{k,n}^{u^{A,\varepsilon,d}}(A) \asymp \phi_{k,n}^{u^{A,\varepsilon,d}}(G).$$

Finally, since Lemma 8.3 implies that the initialization satisfies

$$v^{A,\varepsilon,d,[-d]}(G) = u^{A,\varepsilon,d}(G) \asymp u^{G,\varepsilon,d}(G) = v^{G,\varepsilon,d,[-d]}(G),$$

the argument in the second half of the proof of Lemma 8.2 applies to the present setting and it yields that

$$v^{A,\varepsilon,d,[k]}(G) \asymp v^{G,\varepsilon,d,[k]}(G)$$

for all $k \geq -d$. Consequently,

$$\phi_{k,n}^{u^{A,\varepsilon,d}}(G) \asymp \phi_{k,n}^{u^{G,\varepsilon,d}}(G).$$

From this and Lemma 8.2, the announced result follows.

A Approximate Gaussian integration by parts

This appendix gathers three inequalities of approximate Gaussian integration by parts. Let $s \geq 1$ be fixed. Let a_1, \dots, a_s be independent random variables with zero mean and unit variance. Suppose that g_1, \dots, g_s are i.i.d. standard standard normal and are independent of a_1, \dots, a_s . Set

$$a_j(t) = \sqrt{t}a_j + \sqrt{1-t}g_j$$

for $0 \leq t \leq 1$ and $1 \leq j \leq s$. Set $a(t) = (a_1(t), \dots, a_s(t))$. In what follows, we denote $\alpha = (\alpha_1, \dots, \alpha_s) \in (\{0\} \cup \mathbb{N})^s$, $\alpha! = \alpha_1! \cdots \alpha_s!$, and $|\alpha| = \alpha_1 + \cdots + \alpha_s$. Also, $\partial^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_s}^{\alpha_s}$ and $x^\alpha = x_1^{\alpha_1} \cdots x_s^{\alpha_s}$.

Lemma A.1. Let $f \in C^s(\mathbb{R}^s)$. For any $0 < t < 1$, we have that

$$\begin{aligned} & |\mathbb{E}f(a(t))a_1(t) \cdots a_s(t)\dot{a}_1(t)| \\ & \leq \frac{\mathbb{E}[\dot{a}_1(t)^4]^{1/4}}{(s-1)!} \sum_{|\alpha|=s} \prod_{j=1}^s \mathbb{E}[a_j(t)^{4(\alpha_j+1)}]^{1/4} \left(\int_0^1 \mathbb{E}[|\partial^\alpha f(\xi a(t))|^2] d\xi \right)^{1/2}. \end{aligned} \quad (\text{A.1})$$

Proof. From Taylor's theorem, for any $x \in \mathbb{R}^s$,

$$f(x) = \sum_{|\alpha| \leq s-1} \frac{\partial^\alpha f(0)}{\alpha!} x^\alpha + \frac{1}{(s-1)!} \sum_{|\alpha|=s} x^\alpha \int_0^1 \partial^\alpha f(\xi x) d\xi.$$

From these, for any α satisfying that $|\alpha| \leq s-1$, we can write

$$\mathbb{E}a(t)^\alpha a_1(t) \cdots a_n(t)\dot{a}_1(t) = \mathbb{E}a_1(t)^{\alpha_1+1} \dot{a}_1(t) \cdot \mathbb{E}a_2(t)^{\alpha_2+1} \cdots \mathbb{E}a_s(t)^{\alpha_s+1}(t).$$

If $\alpha_j = 0$ for some $2 \leq j \leq s$, then this expectation vanishes. If $\alpha_j \neq 0$ for all $2 \leq j \leq s$, then the condition $|\alpha| \leq s-1$ forces that $\alpha_1 = 0$ and $\alpha_2 = \cdots = \alpha_s = 1$ so that

$$\mathbb{E}a_1(t)^{\alpha_1+1} \dot{a}_1(t) \cdot \mathbb{E}a_2(t)^{\alpha_2+1} \cdots \mathbb{E}a_s(t)^{\alpha_s+1}(t) = \mathbb{E}a_1(t)\dot{a}_1(t) \cdot \mathbb{E}a_2(t)^2 \cdots \mathbb{E}a_s(t)^2 = 0$$

since

$$\begin{aligned} \mathbb{E}a_1(t)\dot{a}_1(t) &= \mathbb{E}(\sqrt{t}a_1 + \sqrt{1-t}g_1) \left(\frac{a_1}{\sqrt{t}} - \frac{g_1}{\sqrt{1-t}} \right) \\ &= \mathbb{E}a_1^2 - \frac{\sqrt{t}}{\sqrt{1-t}} \mathbb{E}a_1g_1 + \frac{\sqrt{1-t}}{\sqrt{t}} \mathbb{E}a_1g_1 - \mathbb{E}g_1^2 = 0. \end{aligned} \quad (\text{A.2})$$

From these,

$$\begin{aligned} & |\mathbb{E}f(a(t))a_1(t) \cdots a_s(t)\dot{a}_1(t)| \\ &= \left| \frac{1}{(s-1)!} \sum_{|\alpha|=s} \int_0^1 \mathbb{E}[a_1(t) \cdots a_s(t)\dot{a}_1(t)a(t)^\alpha \partial^\alpha f(\xi a(t))] d\xi \right| \\ &\leq \frac{1}{(s-1)!} \sum_{|\alpha|=s} \mathbb{E}[(a_1(t) \cdots a_s(t)\dot{a}_1(t)a(t)^\alpha)^2]^{1/2} \left(\int_0^1 \mathbb{E}[|\partial^\alpha f(\xi a(t))|^2] d\xi \right)^{1/2}. \end{aligned}$$

Finally applying the Cauchy-Schwarz inequality and the independence of $a_1(t), \dots, a_s(t)$ to the first expectation in the last line ends our proof. \square

Lemma A.2. Let $f \in C^{s-1}(\mathbb{R}^s)$. For any $0 < t < 1$ and $r \geq 2$, we have that

$$\begin{aligned} & |\mathbb{E}f(a(t))a_1(t)^r a_2(t) \cdots a_s(t)\dot{a}_1(t)| \\ & \leq \frac{\mathbb{E}[\dot{a}_1(t)^4]^{1/4}}{(s-2)!} \sum_{|\alpha|=s-1} \mathbb{E}[a_1(t)^{4(\alpha_1+r)}]^{1/4} \prod_{j=2}^s \mathbb{E}[a_j(t)^{4(\alpha_j+1)}]^{1/4} \left(\int_0^1 \mathbb{E}[|\partial^\alpha f(\xi a(t))|^2] d\xi \right)^{1/2}. \end{aligned} \quad (\text{A.3})$$

Proof. Consider Taylor's expansion,

$$f(x) = \sum_{|\alpha| \leq s-2} \frac{\partial^\alpha f(0)}{\alpha!} x^\alpha + \frac{1}{(s-2)!} \sum_{|\alpha|=s-1} x^\alpha \int_0^1 \partial^\alpha f(\xi x) d\xi.$$

For any α satisfying $|\alpha| \leq s - 2$, we can write

$$\mathbb{E}a(t)^\alpha a_1(t)^r \cdots a_s(t) \dot{a}(t) = \mathbb{E}a_1(t)^{\alpha_1+r} \dot{a}(t) \cdot \mathbb{E}a_2(t)^{\alpha_2+1} \cdots \mathbb{E}a_s(t)^{\alpha_s+1}(t).$$

Observe that if $\alpha_j \geq 1$ for all $2 \leq i \leq s$, then $|\alpha| \geq s - 1$, which is not possible. Thus, one of the $\alpha_2, \dots, \alpha_s$ must be zero so that this expectation vanishes. Consequently,

$$\begin{aligned} & |\mathbb{E}f(a(t))a_1(t)^r a_2(t) \cdots a_s(t) \dot{a}_1(t)| \\ &= \left| \frac{1}{(s-2)!} \sum_{|\alpha|=s-1} \int_0^1 \mathbb{E}[a_1(t)^r a_2(t) \cdots a_s(t) \dot{a}_1(t) a(t)^\alpha \partial^\alpha f(\xi a(t))] d\xi \right| \\ &\leq \frac{1}{(s-2)!} \sum_{|\alpha|=s-1} \mathbb{E}[(a_1(t)^r \cdots a_s(t) \dot{a}_1(t) a(t)^\alpha)^2]^{1/2} \left(\int_0^1 \mathbb{E}[|\partial^\alpha f(\xi a(t))|^2] d\xi \right)^{1/2}, \end{aligned}$$

which leads to the assertion by applying the Cauchy-Schwarz inequality to the first expectation in the last line. \square

Lemma A.3. *Let $f \in C^{s+1}(\mathbb{R}^s)$. For any $0 < t < 1$, we have that*

$$\begin{aligned} & |\mathbb{E}f(a(t))a_2(t) \cdots a_s(t) \dot{a}_1(t)| \\ &\leq \frac{\mathbb{E}[\dot{a}_1(t)^4]^{1/4}}{s!} \sum_{|\alpha|=s+1} \mathbb{E}[a_1(t)^{4\alpha_1}]^{1/4} \prod_{j=2}^s \mathbb{E}[a_j(t)^{4(1+\alpha_j)}]^{1/4} \left(\int_0^1 \mathbb{E}[|\partial^\alpha f(\xi a(t))|^2] d\xi \right)^{1/2}. \end{aligned} \quad (\text{A.4})$$

Proof. Write

$$f(x) = \sum_{|\alpha| \leq s} \frac{\partial^\alpha f(0)}{\alpha!} x^\alpha + \frac{1}{s!} \sum_{|\alpha|=s+1} x^\alpha \int_0^1 \partial^\alpha f(\xi x) d\xi,$$

For α satisfying $|\alpha| \leq s$, write

$$\mathbb{E}a(t)^\alpha a_2(t) \cdots a_s(t) \dot{a}_1(t) = \mathbb{E}a_1(t)^{\alpha_1} \dot{a}_1(t) \mathbb{E}a_2(t)^{\alpha_2+1} \cdots \mathbb{E}a_s(t)^{\alpha_s+1}.$$

If $\alpha_j = 0$ for some $2 \leq j \leq s$, then this expectation is equal to zero. If $\alpha_j > 0$ for all $2 \leq j \leq s$, then $\alpha_1 = 0$ or $\alpha_1 = 1$. In the former case, the expectation vanishes; in the latter case, this expectation is also equal to zero since $\mathbb{E}a_1(t) \dot{a}(t) = 0$ due to (A.2). Consequently,

$$\begin{aligned} & |\mathbb{E}f(a(t))a_2(t) \cdots a_s(t) \dot{a}_1(t)| \\ &= \left| \frac{1}{s!} \sum_{|\alpha|=s+1} \int_0^1 \mathbb{E}[a_2(t) \cdots a_s(t) \dot{a}_1(t) a(t)^\alpha \partial^\alpha f(\xi a(t))] d\xi \right| \\ &\leq \frac{1}{s!} \sum_{|\alpha|=s+1} \mathbb{E}[(a_2(t) \cdots a_s(t) \dot{a}_1(t) a(t)^\alpha)^2]^{1/2} \left(\int_0^1 \mathbb{E}[|\partial^\alpha f(\xi a(t))|^2] d\xi \right)^{1/2}. \end{aligned}$$

The rest of the proof follows by using the Cauchy-Schwarz inequality and the independence of $a_1(t), \dots, a_s(t)$ to the first expectation of the last line. \square

References

- [1] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

- [2] J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 97(6):1382–1408, 2006.
- [3] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Advances in Neural Information Processing Systems*, 29:424–432, 2016.
- [4] M. Bayati, M. Lelarge, and A. Montanari. Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.*, 25(2):753–822, 2015.
- [5] M. Bayati and A. Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory*, 57(2):764–785, 2011.
- [6] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, 227(1):494–521, 2011.
- [7] F. Benaych-Georges and R. R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.*, 111:120–135, 2012.
- [8] R. Berthier, A. Montanari, and P.-M. Nguyen. State evolution for approximate message passing with non-separable functions. *arXiv preprint arXiv:1708.03950*, 2017.
- [9] M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009.
- [10] Y. Deshpande, E. Abbe, and A. Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Inf. Inference*, 6(2):125–170, 2017.
- [11] Y. Deshpande and A. Montanari. Information-theoretically optimal sparse PCA. *IEEE International Symposium on Information Theory*, pages 2197–2201, 2014.
- [12] Y. Deshpande and A. Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Found. Comput. Math.*, 15(4):1069–1128, 2015.
- [13] D. L. Donoho, A. Javanmard, and A. Montanari. Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE transactions on information theory*, 59(11):7434–7464, 2013.
- [14] D. L. Donoho, I. Johnstone, and A. Montanari. Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. *IEEE transactions on information theory*, 59(6):3396–3433, 2013.
- [15] D. L. Donoho, A. Maleki, and A. Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.
- [16] D. L. Donoho, A. Maleki, and A. Montanari. Message passing algorithms for compressed sensing: I. motivation and construction. In *2010 IEEE information theory workshop on information theory (ITW 2010, Cairo)*, pages 1–5. IEEE, 2010.
- [17] A. El Alaoui, A. Montanari, and M. Sellke. Optimization of mean-field spin glasses. *arXiv preprint arXiv:2001.00904*, 2020.
- [18] D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.*, 272(1):185–228, 2007.
- [19] A. Javanmard and A. Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference*, 2(2):115–144, 2013.
- [20] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.
- [21] Y. Kabashima, F. Krzakala, M. Mézard, A. Sakata, and L. Zdeborová. Phase transitions and sample complexity in Bayes-optimal matrix factorization. *IEEE Trans. Inform. Theory*, 62(7):4228–4265, 2016.
- [22] A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66(11):1663–1750, 2013.
- [23] F. Krzakala, J. Xu, and L. Zdeborová. Mutual information in rank-one matrix estimation. *IEEE Information Theory Workshop (ITW)*, pages 71–75, 2016.
- [24] M. Lelarge and L. Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Proceedings of the 2017 Conference on Learning Theory*, PMLR 65:1297–1301, 2017.
- [25] T. Lesieur, F. Krzakala, and L. Zdeborová. Constrained low-rank matrix estimation: phase transitions, approximate message passing and applications. *J. Stat. Mech. Theory Exp.*, (7):073403, 86, 2017.
- [26] L. Miolane. Fundamental limits of symmetric low-rank matrix estimation: the non-symmetric case. *ArXiv e-prints*, 2017.

- [27] A. Montanari. Optimization of the sherrington-kirkpatrick hamiltonian. *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1417–1433, 2019.
- [28] A. Montanari and E. Richard. Non-negative principal component analysis: message passing algorithms and sharp asymptotics. *IEEE Trans. Inform. Theory*, 62(3):1458–1484, 2016.
- [29] A. Montanari and R. Venkataramanan. Estimation of low-rank matrices via approximate message passing. *arXiv preprint arXiv:1711.01682*, 2017.
- [30] J. T. Parker, P. Schniter, and V. Cevher. Bilinear generalized approximate message passing—Part I: Derivation. *IEEE Trans. Signal Process.*, 62(22):5839–5853, 2014.
- [31] J. T. Parker, P. Schniter, and V. Cevher. Bilinear generalized approximate message passing—Part II: Applications. *IEEE Trans. Signal Process.*, 62(22):5854–5867, 2014.
- [32] D. Paul. Asymptotics of the leading sample eigenvalues for a spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.
- [33] S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields*, 134(1):127–173, 2006.
- [34] S. Rangan and A. K. Fletcher. Iterative estimation of constrained rank-one matrices in noise. *IEEE International Symposium on Information Theory (ISIT)*, pages 1246–1250, 2012.
- [35] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35:1792–1796, Dec 1975.
- [36] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer.
- [37] J. Vila, P. Schniter, and J. Meola. Hyperspectral unmixing via turbo bilinear approximate message passing. *IEEE Trans. Comput. Imaging*, 1(3):143–158, 2015.