The SK model with a sparse variance profile: free energy and AMP algorithm for TAP equations at high temperature

Walid Hachem

CNRS, Laboratoire d'informatique Gaspard Monge (LIGM / UMR 8049),
Université Gustave Eiffel, ESIEE Paris, France
walid.hachem@univ-eiffel.fr

April 28, 2026

Abstract

A generalization of the Sherrington–Kirkpatrick (SK) model for spin glasses is considered, in which the interaction matrix is endowed with a variance profile that has no particular structure and may be sparse. In the first part of the paper, an asymptotic equivalent of the free energy is derived at sufficiently high temperatures, regardless of the signature of the variance profile matrix. In the second part, the mean of the spin vector under the Gibbs measure is estimated using an Approximate Message Passing algorithm based on the Thouless–Anderson–Palmer equations. The dynamical approach of Adhikari *et al.* (J. Stat. Phys., 2021), originally developed for the classical SK model, is adapted to the present setting to obtain these results.

1 Model, problem and the results

For each integer $n > 0$, let $S^{(n)} = \left[s_{ij}^{(n)}\right]_{i,j=1}^{n}$ be a deterministic symmetric matrix with elements $s_{ij}^{(n)} \geqslant 0$ and with a zero diagonal. Let $W^{(n)} = \left[W_{ij}^{(n)}\right]_{i,j=1}^{n}$ be a real symmetric $n \times n$ random matrix such that the random variables $\{W_{ij}^{(n)}\}_{1 \leqslant i < j \leqslant n}$ are independent, and such that $W_{ij}^{(n)} \sim \mathcal{N}(0, ts_{ij}^{(n)})$ for all $i, j \in [n]$ and for some $t > 0$.

Let $\Sigma_n = \{-1, +1\}^n$ be the space of vectors of Ising spins with size n. Define the random $\mathcal{P}(\Sigma_n)$–valued Gibbs measure $G^{(n)}$ as follows. The measure of a singleton $\{\sigma\} \subset \Sigma_n$ by $G^{(n)}$ is $G^{(n)}(\sigma) = \exp(H^{(n)}(\sigma))/Z^{(n)}$, where $H^{(n)}$ is the Hamiltonian defined as

$$H^{(n)}(\sigma) = \frac{1}{2}\sigma^\top W^{(n)}\sigma + h\,(\sigma \cdot 1_n)\,,$$

$h \in \mathbb{R}$ is the amplitude of an external field, and $Z^{(n)} = \sum_{\sigma\in\Sigma_n} \exp(H^{(n)}(\sigma))$ is the partition function. In the particular case where $s_{ij}^{(n)} = 1/n$ for $i \neq j$, this model for the Gibbs measure boils down to the classical Sherrington-Kirkpatrick (SK) model that has been studied at length in the statistical physics literature, as detailed in the treatises [25, 26, 23]. Since the interactions between our spins are subjected to the more general variance profile represented by the matrix $S^{(n)}$, we term our model a SK model with a variance profile.

Let us state our conditions on this matrix. Considering a sequence of positive numbers (K_n) such that $K_n \leqslant n$ and $K_n \to \infty$, we assume the following:

Assumption 1. It holds that

- There exists a constant $\boldsymbol{C}_s > 0$ such that $s_{ij}^{(n)} \leqslant \boldsymbol{C}_s K_n^{-1}$ for all n and all $i, j \in [n]$.
- The number $\boldsymbol{C}_{\text{row}} = \sup_n \left|\!\left|\!\left| S^{(n)} \right|\!\right|\!\right|$ where $|\!|\!|\cdot|\!|\!|$ is the max row sum norm is finite.

One case of interest covered by Assumption 1 and that can be useful in the fields of statistical physics, graph inference, and large-dimensional signal estimation among others is the case where $K_n \ll n$, and where there are at most $\boldsymbol{C}_{\text{card}} K_n$ non-zero $s_{ij}^{(n)}$'s in each row and column of $S^{(n)}$ for some constant $\boldsymbol{C}_{\text{card}} > 0$. We refer to these models as the "sparse" ones.

The first aim of this paper is to study the large n asymptotics of the free energy

$$F_n = \frac{1}{n} \mathbb{E} \log Z^{(n)}$$

in a "high temperature" regime represented by the following assumption:

Assumption 2. $t < \dfrac{\log 2}{\boldsymbol{C}_{\text{row}}}$.

Denoting as usual as $\langle \cdot \rangle$ the mean operator with respect to the measure $(G^{(n)})^{\otimes \infty}$, the second aim of this paper is to provide a large–n approximation of the mean vector $m^{(n)} = \langle \sigma \rangle$. Returning to the classical SK model, it is well-known that this vector can be approximated in the high temperature regime by an Approximate Message Passing (AMP) algorithm by building on the so-called Thouless-Anderson-Palmer (TAP) equations [9]. The second purpose of this paper is to generalize this result to our SK model with a variance profile in the temperature regime specified by Assumption 2.

The conditions on the variance profile that can be found in the mathematical physics literature are usually much more restrictive than those provided by Assumption 1. Regarding the free energy computation, this literature is mainly limited to the so-called multi-species model, where it is assumed that $S^{(n)}$ consists in a finite number of blocks which dimensions scale with n (implying that $K_n = n$ in our setting), and typically, that $S^{(n)}$ is a non-negative matrix in the semi-definite positive ordering. These two assumptions are made in, *e.g.*, [5, 24] which deal, on the other hand, with the free energy problem at all temperatures. More recent contributions dealing with the multi-species model with the non-negativity assumption include [6, 2, 11, 20]. Cases where this matrix can be indefinite were discussed in [14, 10, 27, 7]. Diluted variants of the multi-species model are considered in [21, 3]. Considering the potential applications, graph max κ–cut problems and low-rank matrix estimation problems related with the multi-species model were considered in [19] and in [17] respectively. It would be useful to extend these results to more flexible and more general variance profiles such as the ones considered in this paper.

Denoting in all this paper as ξ a standard Gaussian random variable, and writing $\text{Tanh}(x) = \tanh(x + h)$, define the scalar function $g : \mathbb{R}_+ \to \mathbb{R}_+$ as

$$g(x) = \mathbb{E}\, \text{Tanh}\left(\sqrt{x}\xi\right)^2, \tag{1}$$

Consider the system of equations defined in $q^{(n)} \in \mathbb{R}_+^n$ as

$$q^{(n)} = tS^{(n)} g(q^{(n)}) \tag{2}$$

where $g(q^{(n)}) \in \mathbb{R}_+^n$ is the vector obtained by an element-wise application of the function g to the vector $q^{(n)}$ (this notational convention regarding scalar functions applied to vectors will be used all along this paper).

Lemma 1. Under Assumptions 1 and 2, Equation (2) admits an unique solution $q^{(n)} \in \mathbb{R}_+^n$. Given any vector $q^{(n),0} \in \mathbb{R}_+^n$, the iterative algorithm $q^{(n),l+1} = tS^{(n)} g(q^{(n),l})$ converges to this solution. Moreover, $\sup_n \|q^{(n)}\|_\infty < \log 2$ where $\|\cdot\|_\infty$ is the max–norm.

This simple lemma is proven in Appendix A.1 for completeness. The large–n behavior of the free energy is specified by the following theorem:

Theorem 2. Let Assumptions 1 and 2 hold true. Defining the function

$$\boldsymbol{F}_n = \log 2 + \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\log\cosh\left(\sqrt{q_i^{(n)}}\xi + h\right) + \frac{t}{4n}\left(1_n - g(q^{(n)})\right)^\top S^{(n)}\left(1_n - g(q^{(n)})\right),$$

where $q^{(n)} = \left[q_i^{(n)}\right]_{i=1}^{n}$ is the solution of (2), it holds that $\boldsymbol{F}_n$ is bounded, and moreover, that

$$F_n - \boldsymbol{F}_n \xrightarrow[n\to\infty]{} 0.$$

It is seen here that the signature of $S^{(n)}$ has no impact on the form of the large-n approximation of the free energy.

It is worth considering the particular case of this theorem where the matrix $S^{(n)}$ is doubly stochastic. In this case, the solution of Equation (2) is reduced to $q^{(n)} = q1_n$ where the scalar q is the unique solution to $q = tg(q)$. Therefore, as long as the degree of sparsity satisfies $K_n \to \infty$, we recover the expression of the asymptotic free energy of the classical SK model at high temperature [25]:

Corollary 3 (the doubly stochastic case)**.** Assume that $S^{(n)}$ is a doubly stochastic matrix. Then, it holds under Assumptions 1 and 2 (which reads $t < \log 2$) that

$$F_n \xrightarrow[n\to\infty]{} \log 2 + \mathbb{E}\log\cosh\left(\sqrt{q}\xi + h\right) + \frac{t}{4}(1 - q/t)^2, \tag{3}$$

where q is the unique solution to the scalar equation $q = tg(q)$ on $\mathbb{R}_+$.

A particular case that can be useful in the field of statistical physics is the case where $S^{(n)}$ is a Toeplitz banded matrix with a bandwidth of order $K_n \to \infty$ with $K_n \ll n$. Here, K_n represents the range of the interactions among the spins. Within this range, the random interactions are furthermore subjected to a variance profile that depends on the distance between the sites as shown in the statement of the following corollary:

Corollary 4 (banded Toeplitz interaction profile)**.** Let $\psi^{(n)} : \{0, \ldots, n-1\} \to \mathbb{R}_+$ be a function that satisfies the following assumptions: $\psi^{(n)}(i) \leqslant \boldsymbol{C}_s/K_n$, $\sum_i \psi^{(n)}(i) = 1/2$, and the support of $\psi^{(n)}$ is included into $\{1, \ldots, K_n\}$. Let $s_{ij}^{(n)} = \psi^{(n)}(|i-j|)$, and assume that $K_n \to \infty$ with $K_n/n \to 0$. Then, for $t < \log 2$, the convergence (3) holds true.

The proof is provided in Appendix A.2.

We now tackle the approximation problem of $m^{(n)} = \langle\sigma\rangle$ with the help of an AMP algorithm. To this end, we need to strengthen a bit Assumption 1. Keeping our sequence $K_n \to \infty$ with $K_n \leqslant n$, we set:

Assumption 3. The following facts hold true.

- There exists a constant $\boldsymbol{C}_s > 0$ such that $s_{ij}^{(n)} \leqslant \boldsymbol{C}_s K_n^{-1}$ for all n and all $i, j \in [n]$.
- There exists a constant $\boldsymbol{C}_{\mathrm{card}} > 0$ such that

  $$\forall n,\ \forall i \in [n],\ \left|\left\{j \in [n] \,:\, s_{ij}^{(n)} > 0\right\}\right| \leqslant \boldsymbol{C}_{\mathrm{card}} K_n,$$

  where $|\cdot|$ is the cardinality of a set.

Of course, Assumption 3 implies Assumption 1 with $\boldsymbol{C}_{\mathrm{row}} = \sup_n \left|\!\left|\!\left|S^{(n)}\right|\!\right|\!\right|$ satisfying $\boldsymbol{C}_{\mathrm{row}} \leqslant \boldsymbol{C}_s\boldsymbol{C}_{\mathrm{card}}$. If $K_n \ll n$, Assumption 3 models the sparse cases alluded to above.

In the remainder, we denote as $\|\cdot\|$ the Euclidean norm of a vector or the spectral norm of a matrix. We also write $\|\cdot\|_n = \|\cdot\|/\sqrt{n}$.

Theorem 5. Let Assumptions 3 and 2 hold true. Assume that $K_n \geqslant \log n$. Consider the iterates $(q^{(n),l})_{l\in\mathbb{N}}$ defined in the statement of Lemma 1 starting with $q^{(n),0} = 0_n$. For each $l = 0, 1, \ldots$, let $X^{(n),l} \sim \mathcal{N}(0, \operatorname{diag}(q^{(n),l}))$. Starting with $x^{(n),0} = 0$ and $x^{(n),1} = W^{(n)} \operatorname{Tanh}(0)$, consider the following iterative AMP algorithm in $l = 1, \ldots$

$$\begin{aligned} x^{(n),l+1} &= W^{(n)} \operatorname{Tanh}\left(x^{(n),l}\right) - \operatorname{diag}\left(tS^{(n)}1_n - q^{(n),l+1}\right) \operatorname{Tanh}\left(x^{(n),l-1}\right) \\ &= W^{(n)} \operatorname{Tanh}\left(x^{(n),l}\right) - \operatorname{diag}\left(tS^{(n)}\mathbb{E}\operatorname{Tanh}'\left(X^{(n),l}\right)\right) \operatorname{Tanh}\left(x^{(n),l-1}\right). \end{aligned} \tag{4}$$

Then, the vector $m^{(n)} = \langle \sigma \rangle$ satisfies

$$\lim_{k\to\infty} \limsup_n \mathbb{E} \left\| m^{(n)} - \operatorname{Tanh}\left(x^{(n),k}\right) \right\|_n^2 = 0.$$

2 Proofs

In the remainder, we omit the superscript $^{(n)}$ from the notations unless when useful. We denote as $C > 0$ a constant that depends on $\boldsymbol{C}_{\text{row}}$, $\boldsymbol{C}_s$, $\boldsymbol{C}_{\text{card}}$ and t at most, and that can change from a display to another.

We shall approximate the free energy at high temperature via the so-called Guerra's interpolation of the Hamiltonian. In our setting, this gives the following scheme. Let

$$\eta = [\eta_i]_{i=1}^n \sim \mathcal{N}(0, \operatorname{diag}(q))$$

be independent with W, where we recall that q is the unique solution of the system $q = tSg(q)$ as shown by Lemma 1. Given $u \in [0, 1]$, define the Hamiltonian on the space of spins Σ_n as

$$H_u(\sigma) = \frac{\sqrt{u}}{2} \sigma^\top W \sigma + h(\sigma \cdot 1) + \sqrt{1-u}\,(\sigma \cdot \eta),$$

and consider the Gibbs measure G_u defined on Σ_n as $G_u(\sigma) = \exp(H_u(\sigma))/Z_u$ where $Z_u = \sum_\sigma \exp(H_u(\sigma))$ is the partition function. Our Gibbs measure of interest is of course G_1, and its free energy is $F_n = n^{-1}\mathbb{E}\log Z_1$. The Hamiltonian H_u is an interpolation between the Hamiltonian of interest H_1 and the Hamiltonian H_0 which free energy has a tractable expression.

To pursue, we introduce some notations. Given a set of indices $A \subset [n]$, we write $A^{\text{c}} = [n]\backslash A$. Given a vector $x = [x_i] \in \mathbb{R}^n$ we denote respectively $x_A \in \mathbb{R}^n$ the vector x which elements x_i are set to zero when $i \in A^{\text{c}}$.

We denote as $H_{u,(A)}$ the reduced Hamiltonian obtained by removing the spins $\{\sigma_i\}_{i\in A}$ from the system. More precisely, this Hamiltonian is written as

$$H_{u,(A)}(\sigma_{A^{\text{c}}}) = \frac{\sqrt{u}}{2} \sigma_{A^{\text{c}}}^\top W \sigma_{A^{\text{c}}} + h(\sigma_{A^{\text{c}}} \cdot 1_n) + \sqrt{1-u}\,(\sigma_{A^{\text{c}}} \cdot \eta_{A^{\text{c}}}),$$

and is considered as a Hamiltonian on $\{-1, +1\}^{|A^{\text{c}}|}$. We denote as $G_{u,(A)}(\sigma_{A^{\text{c}}}) \propto \exp(H_{u,(A)}(\sigma_{A^{\text{c}}}))$ the Gibbs probability measure on $\{-1, +1\}^{|A^{\text{c}}|}$ which Hamiltonian is $H_{u,(A)}(\sigma_{A^{\text{c}}})$. For $i \in A^{\text{c}}$, we also denote as $m_{(A),i}$ the mean of the spin σ_i with respect to $G_{u,(A)}$. Conventionally, we write $m_{(A),i} = 0$ when $i \in A$ so that we can define the vector $m_{(A)} = [m_{(A),i}]_{i\in[n]} \in \mathbb{R}^n$. When $A = \{i_1, \ldots, i_k\}$, we sometimes write $m_{(i_1,\ldots,i_k)} = [m_{(i_1,\ldots,i_k),i}]_i$ for the vector $m_{(A)} = [m_{(A),i}\,]_i$. Of course, $m_{(\varnothing)} = m = [m_i]$.

Proof idea. Our starting point will be the approach of Adikhari *et.al.* in [1], where the SK case with $u = 1$ was considered. This approach falls within a research axis that dates back up to our knowledge to the work of Comets and Neveu [13], and that considers the SK model from a stochastic calculus perspective. Denoting as $\langle\cdot\rangle_u$ the mean with respect to the measure $G_u^{\otimes\infty}$, the first step is to show that the covariances $m_{ij} = \left\langle(\sigma_i - \langle\sigma_i\rangle_u)(\sigma_j - \langle\sigma_j\rangle_u)\right\rangle_u$ for $i \neq j$ satisfy

$\mathbb{E}m_{ij}^2 \sim 1/K_n$. This is shown in Lemma 6 below, which is a straightforward adaptation of [1] to the variance profile case of interest in this paper.

Using this result, the "pre-TAP" bound

$$\mathbb{E}\left(m_i - \mathrm{Tanh}\left(\sqrt{u}\sum_k W_{ik}m_{(i),k} + \sqrt{1-u}\eta_i\right)\right)^2 \leqslant \frac{C}{K_n}, \tag{5}$$

as well as the bound

$$\mathbb{E}\left(m_l - m_{(i),l}\right)^2 \leqslant \frac{C}{K_n}, \quad l \neq i$$

can be obtained. These bounds are generalizations to our model of quite well-known results in the SK literature, see, *e.g.*, [25, Lemma 1.7.4]. Here, they will serve two purposes.

First, defining the random vector

$$R_{12} = \left[R_{12}(i)\right]_{i=1}^n = tS(\sigma^1\sigma^2),$$

where $\sigma^1\sigma^2$ is the vector obtained by an elementwise product of the elements of the replicas σ^1 and σ^2, it can be deduced from these bounds that $\mathbb{E}\left\langle\|R_{12} - q\|_n^2\right\rangle_u \leqslant CK_n^{-1/2}$ (Lemma 8 below). Employing the usual Guerra's interpolation trick in order to compute the free energy of G_1, we can see that thanks to the bound $\mathbb{E}\left\langle\|R_{12} - q\|_n^2\right\rangle_u \leqslant CK_n^{-1/2}$, the "annoying" term obtained through this interpolation is negligible, which leads to Theorem 2. We note here that in the references cited above which deal with the multi-species model, it is assumed that S is a non-negative matrix specifically to force this term to be non-positive.

Second, the bound (5) taken for $u = 1$ leads to the construction (13) for approximating m_i, as was done in Chen and Tang in [12] for the SK model. Starting from this construction, we shall devise a series of approximations of the vector $m = [m_i]$ that will ultimately lead to the AMP approximation given by Theorem 5. These approximations will be based on the approach of Bayati *et.al.* in [8], devoted to the classical AMP algorithm, which was generalized to the AMP algorithm with an interaction matrix with a variance profile in [18].

2.1 Adapting the approach of [1] to the SK model with a variance profile

We need to introduce some more notations. For a set $A \subset [n]$, we need to work on the conditional interpolated Gibbs measure $G_u(\cdot|\sigma_A)$ given σ_A, which is the measure on $\{-1,+1\}^{|A^c|}$ with the Hamiltonian $\sigma_{A^c} \mapsto H_u^{[A]}(\sigma_{A^c})(\sigma_A)$ given as

$$H_u^{[A]}(\sigma_{A^c})(\sigma_A) = \frac{\sqrt{u}}{2}\sigma_{A^c}^\top W\sigma_{A^c} + h\left(\sigma_{A^c}\cdot 1\right) + \sqrt{1-u}\left(\sigma_{A^c}\cdot\eta_{A^c}\right) + \sqrt{u}\sigma_{A^c}^\top W\sigma_A.$$

Given $i, j \in A^c$, we denote as $m_i^{[A]}(\sigma_A)$ and $m_{ij}^{[A]}(\sigma_A)$ the mean of σ_i and the covariance of σ_i and σ_j with respect to the conditional probability $G_u(\cdot|\sigma_A)$. We also write $m_i^{[i_1,\ldots,i_k]}$ and $m_{ij}^{[i_1,\ldots,i_k]}$ for $m_i^{[A]}$ and $m_{ij}^{[A]}$ respectively when $A = \{i_1,\ldots,i_k\}$. For $i \in A$ and a real function $f(\sigma_A)$, we also define the functions $\sigma_{A\backslash\{i\}} \mapsto \delta_i f(\sigma_{A\backslash\{i\}})$ and $\sigma_{A\backslash\{i\}} \mapsto \varepsilon_i f(\sigma_{A\backslash\{i\}})$ as

$$\delta_i f(\sigma_{A\backslash\{i\}}) = \frac{1}{2}\left(f(\sigma_A)_{|\sigma_i=1} - f(\sigma_A)_{|\sigma_i=-1}\right),$$

and

$$\varepsilon_i f(\sigma_{A\backslash\{i\}}) = \frac{1}{2}\left(f(\sigma_A)_{|\sigma_i=1} + f(\sigma_A)_{|\sigma_i=-1}\right).$$

The following key identity can be obtained by direct calculation and is provided in [1, Eq. (3.1)]:

$$m_{ij}^{[A]} = \left(1 - \left(m_i^{[A]}\right)^2\right)\delta_i m_j^{[A\cup\{i\}]} \tag{6}$$

for $i, j \in A^c$ with $i \neq j$. Following [1], we consider $\sqrt{u}W_{ij}$ as the value at tu of the process $\sqrt{s_{ij}}B_{ij}(v)$ where B_{ij} is a standard Brownian Motion. Given a set $A \subset [n]$ and an index $i \in A^c$, it is possible to obtain a characterization of $\delta_i m_j^{[A\cup\{i\}]}(\sigma_A)$ with the help of Itô's lemma. Similarly to [1, Eq. (3.3)], we obtain by this lemma

$$
\begin{aligned}
\delta_i m_j^{[A\cup\{i\}]}(\sigma_A) = &\sum_{k\notin A\cup\{i\}} \sqrt{s_{ik}} \int_0^{tu} \varepsilon_i m_{kj}^{[A\cup\{i\}]}(\sigma_A)(v) dB_{ik}(v) \\
&- \sum_{k\notin A\cup\{i\}} s_{ik} \int_0^{tu} \delta_i \left(m_k^{[A\cup\{i\}]} m_{kj}^{[A\cup\{i\}]} \right) (\sigma_A)(v) dv.
\end{aligned}
$$

(here, $\varepsilon_i m_{kj}^{[A\cup\{i\}]}(\sigma_A)(v)$ is of course the value of $\varepsilon_i m_{kj}^{[A\cup\{i\}]}(\sigma_A)$ for which $\sqrt{u}W_{ij}$ in the Hamiltonian is replaced with $\sqrt{s_{ij}}B_{ij}(v)$, and similarly for the second integrand).

This Itô characterization of $\delta_i m_j^{[A\cup\{i\}]}$ together with the identity (6) lie at the basis of proof of the following result, which is an adaptation of [1, Lemma 3.1] to our situation. For completeness, we provide this proof in Appendix A.3.

Lemma 6. For $t \in [0, \log 2/\boldsymbol{C}_{\mathrm{row}})$, there exists a constant C such that

$$
\mathbb{E} m_{ij}^2 \leqslant \frac{C}{K_n} \quad \text{for all } i \neq j.
$$

With the help of the previous lemma, we obtain the following result by a straightforward adaptation of the proof of [1, Lemma 4.1]:

Lemma 7. For $t \in [0, \log 2/\boldsymbol{C}_{\mathrm{row}})$, there exists a constant C such that

$$
\mathbb{E} \left(m_i - \mathrm{Tanh} \left(\sqrt{u} \sum_{k\neq i} W_{ik} m_{(i),k} + \sqrt{1-u}\eta_i \right) \right)^2 \leqslant \frac{C}{K_n},
$$

and

$$
\mathbb{E} \left(m_l - m_{(i),l} \right)^2 \leqslant \frac{C}{K_n}
$$

for each $i \neq l \in [n]$.

Building the random vector $R_{12} = tS(\sigma^2\sigma^2)$ from two i.i.d. vectors σ^1 and σ^2 under G_u (the so-called replicas), we can now use this result to show that $\mathbb{E}\left\langle \|R_{12} - q\|_n^2 \right\rangle_u$ converges to zero uniformly in $u \in [0,1]$:

Lemma 8. For $t \in [0, \log 2/\boldsymbol{C}_{\mathrm{row}})$, there exists a constant C such that

$$
\mathbb{E} \left\langle \|R_{12} - q\|_n^2 \right\rangle_u \leqslant \frac{C}{\sqrt{K_n}}.
$$

Proof. Writing $R_{12} = [R_{12}(i)]_{i=1}^n$, we first show that

$$
\mathbb{E} \left\langle (R_{12}(i) - \langle R_{12}(i) \rangle_u)^2 \right\rangle_u \leqslant C/\sqrt{K_n}. \tag{7}
$$

We write

$$
\begin{aligned}
\mathbb{E} \left\langle (R_{12}(i) - \langle R_{12}(i) \rangle_u)^2 \right\rangle_u &= \mathbb{E} \left\langle R_{12}(i)^2 \right\rangle_u - \mathbb{E} \left\langle R_{12}(i) \right\rangle_u^2 \\
&= \mathbb{E} \sum_{j,\ell} t^2 s_{ij} s_{i\ell} \left\langle \sigma_j^1 \sigma_j^2 \sigma_\ell^1 \sigma_\ell^2 \right\rangle_u - \mathbb{E} \left(\sum_j t s_{ij} \left\langle \sigma_j \right\rangle_u^2 \right)^2 \\
&= \sum_{j\ell} t^2 s_{ij} s_{i\ell} \mathbb{E} \left[\left\langle \sigma_j \sigma_\ell \right\rangle_u^2 - \left\langle \sigma_j \right\rangle_u^2 \left\langle \sigma_\ell \right\rangle_u^2 \right],
\end{aligned}
$$

where $\sigma^1 = [\sigma_i^1]$ and $\sigma^2 = [\sigma_i^2]$. The contribution of the terms $j = \ell$ is bounded by C/K_n. Regarding the terms $j \neq \ell$, we note that $\langle \sigma_j \sigma_\ell \rangle_u^2 - \langle \sigma_j \rangle_u^2 \langle \sigma_\ell \rangle_u^2 = m_{j\ell}(m_{j\ell} + 2m_j m_\ell)$ and we use Lemma 6 to obtain (7), which leads to the inequality

$$\mathbb{E} \left\langle \| R_{12} - \langle R_{12} \rangle_u \|_n^2 \right\rangle_u \leqslant C/\sqrt{K_n}.$$

To obtain the result of the lemma, it remains to prove that

$$\mathbb{E} \, \| \langle R_{12} \rangle_u - q \|_n^2 \leqslant C/\sqrt{K_n}. \tag{8}$$

The proof of this result is just an adaptation of the proof of [1, Proposition 1.2] to our context. The main modifications are related with the fact that Equation (2) is no more a scalar equation.

We write $p = [p_i]_{i=1}^n = \langle R_{12} \rangle_u = tSm^2$. Given a set $A = \{i_1, \ldots, i_k\} \subset [n]$ and an index $i \in A$, we write for brevity

$$\mathrm{Tanh}_{(i_1,\ldots,i_k)}(i) = \mathrm{Tanh} \left(\sum_l \sqrt{u} W_{il} m_{(i_1,\ldots,i_k),l} + \sqrt{1-u}\eta_i \right).$$

Recall that $m_{(i_1,\ldots,i_k),l} = 0$ if $l \in A$, and notice that $\{W_{il}\}_{l \notin A}$ and $\{m_{(i_1,\ldots,i_k),l}\}_l$ are independent, a fact that we shall used repeatidly in the proof without further mention. We also write

$$p_{(i_1,\ldots,i_k),i} = \sum_l t s_{il} m^2_{(i_1,\ldots,i_k),l} = \left[tS m^2_{(i_1,\ldots,i_k)} \right]_i,$$

to be compared with $p_i = \sum_l t s_{il} m_l^2$.

Our first purpose is to show that

$$|\mathbb{E} p_i - [tS\mathbb{E} g(up + (1-u)q)]_i| \leqslant C/\sqrt{K_n}. \tag{9}$$

Using Lemma 7, we have that

$$\mathbb{E} \left| m_i^2 - \mathrm{Tanh}^2_{(i),i} \right| \leqslant 2\mathbb{E} \left| m_i - \mathrm{Tanh}_{(i),i} \right| \leqslant C/\sqrt{K_n}.$$

Observing that the conditional distribution of the random variable $\sum_l \sqrt{u} W_{il} m_{(i),l} + \sqrt{1-u}\eta_i$ with respect to the σ–field $\mathscr{F}_{-i}$ generated by $\{W_{k\ell} \, : \, k, l \neq i, k < l, \; \eta_j \, : \, j \neq i\}$ is $\mathcal{N}(0, up_{(i),i} + (1-u)q_i)$, we have

$$\mathbb{E}\, \mathrm{Tanh}^2_{(i),i} = \mathbb{E} \left[\mathbb{E} \left[\mathrm{Tanh}^2_{(i),i} \mid \mathscr{F}_{-i} \right] \right] = \mathbb{E} g(up_{(i),i} + (1-u)q_i).$$

By Lemma 7, we know that $\mathbb{E}(m_l - m_{(i),l})^2 \leqslant C/K_n$ for $l \neq i$. Therefore, $\mathbb{E}|m_l^2 - \mathbb{E} m^2_{(i),l}| \leqslant C/\sqrt{K_n}$, and then, $\mathbb{E}|p_i - p_{(i),i}| \leqslant C/\sqrt{K_n}$. Since g is Lipschitz (see the proof of Lemma 1), we obtain that $|\mathbb{E} g\left(up_i + (1-u)q_i\right) - \mathbb{E} g\left(up_{(i),i} + (1-u)q_i\right)| \leqslant C/\sqrt{K_n}$, and we deduce from these bounds that

$$\left| \mathbb{E} m_i^2 - \mathbb{E} g\left(up_i + (1-u)q_i\right) \right| \leqslant C/\sqrt{K_n},$$

and the bound (9) follows.

Next, we show that

$$\left| \mathbb{E} p_i^2 - \mathbb{E}[tSg(up + (1-u)q)]_i^2 \right| \leqslant C/\sqrt{K_n} \tag{10}$$

along the same principle. For $i \neq j$, we have by Lemma 7 again that

$$\mathbb{E}|m_i^2 m_j^2 - \mathrm{Tanh}_{(i)}(i)^2 \, \mathrm{Tanh}_{(j)}(j)^2| \leqslant C/\sqrt{K_n}.$$

We now need to replace $\mathrm{Tanh}_{(i)}(i)$ and $\mathrm{Tanh}_{(j)}(j)$ with $\mathrm{Tanh}_{(i,j)}(i)$ and $\mathrm{Tanh}_{(i,j)}(j)$ respectively. We have

$$\begin{aligned}\mathbb{E}\left(\mathrm{Tanh}_{(i),i} - \mathrm{Tanh}_{(i,j),i}\right)^2 &\leqslant u\mathbb{E}\left(\sum_r W_{ir} m_{(i),r} - \sum_r W_{ir} m_{(i,j),r}\right)^2 \\ &\leqslant 2u \sum_{r \neq j} \mathbb{E}W_{ir}^2 \mathbb{E}(m_{(i),r} - m_{(i,j),r})^2 + 2u\mathbb{E}W_{ij}^2 m_{(i),j}^2 \\ &\leqslant C/K_n.\end{aligned}$$

This implies that $\mathbb{E}|m_i^2 m_j^2 - \mathbb{E}\,\mathrm{Tanh}_{(i,j)}(i)^2\,\mathrm{Tanh}_{(i,j)}(j)^2| \leqslant C/\sqrt{K_n}$. Denoting as $\mathscr{F}_{-(i,j)}$ the σ–field generated by $\{W_{kl} \,:\, k,l \notin \{i,j\},\ k < l,\ \eta_r \,:\, r \notin \{i,j\}\}$, and writing as $\mathscr{L}(\cdot \mid \mathscr{F}_{-(i,j)})$ the conditional distribution with respect to this σ–field, we have

$$\begin{aligned}&\mathscr{L}\left(\begin{bmatrix}\sum_r \sqrt{u} W_{ir} m_{(i,j),r} + \sqrt{1-u}\eta_i \\ \sum_r \sqrt{u} W_{jr} m_{(i,j),r} + \sqrt{1-u}\eta_j\end{bmatrix} \mid \mathscr{F}_{-(i,j)}\right) \\ &\qquad\qquad = \mathcal{N}\left(0, \begin{bmatrix} up_{(i,j),i} + (1-u)q_i & 0 \\ 0 & up_{(i,j),j} + (1-u)q_j\end{bmatrix}\right),\end{aligned}$$

which shows that

$$\mathbb{E}\,\mathrm{Tanh}_{(i,j)}(i)^2\,\mathrm{Tanh}_{(i,j)}(j)^2 = \mathbb{E}g(up_{(i,j)}(i) + (1-u)q_i)g(up_{(i,j)}(j) + (1-u)q_j).$$

Similarly to above, we also have $\mathbb{E}|p_{(i,j)}(i) - p_i| + \mathbb{E}|p_{(i,j)}(j) - p_j| \leqslant C/\sqrt{K_n}$, thus,

$$\left|\mathbb{E}\,\mathrm{Tanh}_{(i,j)}(i)^2\,\mathrm{Tanh}_{(i,j)}(j)^2 - \mathbb{E}g(up_i + (1-u)q_i)g(up_j + (1-u)q_j)\right| \leqslant C/\sqrt{K_n}$$

since g is Lipschitz and bounded. Gathering these bounds, we obtain that

$$\left|\mathbb{E}m_i^2 m_j^2 - \mathbb{E}g(up_i + (1-u)q_i)g(up_j + (1-u)q_j)\right| \leqslant C/\sqrt{K_n},$$

and since $p_i^2 = \sum_{k,\ell} s_{ik}s_{i\ell}m_k^2 m_\ell^2$, the bound (10) follows.

From Inequalities (9) and (10), we have

$$\|\mathbb{E}p - \mathbb{E}tSg(X_p)\|_n \leqslant C/\sqrt{K_n} \quad \text{and} \quad \left|\mathbb{E}\|p\|_n^2 - \mathbb{E}\|tSg(X_p)\|_n^2\right| \leqslant C/\sqrt{K_n},$$

where we wrote $X_p = up + (1-u)q$ for notational simplicity.

Since the spectral norm $\|S\|$ satisfies $\|S\| \leqslant |||S|||$, it holds by Assumptions 1 and 2 that the function $x \in \mathbb{R}_+^n \mapsto tSg(x)$ is Lipschitz for the Euclidean norm with a Lipschitz constant $\alpha < 1$ independent of n. From what precedes, we therefore have

$$\begin{aligned}\mathbb{E}\|p - \mathbb{E}p\|_n^2 &\leqslant \mathbb{E}\|p - tSg(\mathbb{E}X_p)\|_n^2 \\ &= \mathbb{E}\|p\|_n^2 + \|tSg(\mathbb{E}X_p)\|_n^2 - 2n^{-1}\left(\mathbb{E}p \cdot tSg(\mathbb{E}X_p)\right) \\ &= \mathbb{E}\|tSg(X_p)\|_n^2 + \|tSg(\mathbb{E}X_p)\|_n^2 - 2n^{-1}\left(\mathbb{E}tSg(X_p) \cdot tSg(\mathbb{E}X_p)\right) + \mathcal{O}(1/\sqrt{K_n}) \\ &= \mathbb{E}\|tSg(X_p) - tSg(\mathbb{E}X_p)\|_n^2 + \mathcal{O}(1/\sqrt{K_n}) \\ &\leqslant \alpha^2 \mathbb{E}\|p - \mathbb{E}p\|_n^2 + \mathcal{O}(1/\sqrt{K_n})\end{aligned}$$

which implies that

$$\mathbb{E}\|p - \mathbb{E}p\|_n^2 \leqslant C/\sqrt{K_n}. \tag{11}$$

From this result, we have

$$\|\mathbb{E}p - tSg(\mathbb{E}X_p)\|_n \leqslant \|\mathbb{E}tSg(X_p) - tSg(\mathbb{E}X_p)\|_n + C/\sqrt{K_n} \leqslant \mathbb{E}\|p - \mathbb{E}p\|_n + C/\sqrt{K_n} \leqslant C/K_n^{1/4}.$$

Finally,

$$\|\mathbb{E}p - q\|_n \leqslant \|tSg(u\mathbb{E}p + (1-u)q) - tSg(uq + (1-u)q)\|_n + C/K_n^{1/4} \leqslant \alpha\|\mathbb{E}p - q\|_n + C/K_n^{1/4},$$

which leads to

$$\|\mathbb{E}p - q\|_n \leqslant C/K_n^{1/4}.$$

Together with (11), we obtain the bound (8), and the lemma is proven. □

2.2 Proof of Theorem 2

Since g is a bounded function, the quadratic form in the expression of $\boldsymbol{F}_n$ is bounded. By Lemma 1, $\|q\|_\infty$ is bounded, thus, the second term in the expression of $\boldsymbol{F}_n$ is bounded, hence the boundedness of $\boldsymbol{F}_n$.

To establish Theorem 2, we use the classical Guerra's approach based on the interpolated Hamiltonian H_u. Defining the function φ on $[0,1]$ as

$$\varphi(u) = \frac{1}{n}\mathbb{E}\log Z_u^{(n)} = \frac{1}{n}\mathbb{E}\log \sum_{\sigma\in\Sigma_n} e^{H_u(\sigma)},$$

we have

$$\begin{aligned}\varphi(0) &= \log 2 + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\log\cosh\left(\sqrt{q_i}\xi + h\right), \quad \text{and} \\ \varphi(1) &= F_n.\end{aligned}$$

We need to compute the derivative $\varphi'(u) = n^{-1}\langle \partial_u H_u(\sigma)\rangle_u$. Writing

$$U(\sigma^1,\sigma^2) = \mathbb{E}(\partial_u H_u(\sigma^1))H_u(\sigma^2) = \frac{t}{4}(\sigma^1\sigma^2)^\top S(\sigma^1\sigma^2) - 2\left((\sigma^1\sigma^2)\cdot Sg(q)\right),$$

we know by the well-known Gaussian integration by parts formula, see, *e.g.*, [23, Lemma 1.1], that

$$\begin{aligned}\varphi'(u) &= \frac{1}{n}\mathbb{E}\left\langle U(\sigma^1,\sigma^1) - U(\sigma^1,\sigma^2)\right\rangle_u \\ &= \frac{t}{4n}(1-g(q))^\top S(1-g(q)) - \frac{t}{4n}\mathbb{E}\left\langle(\sigma^1\sigma^2 - g(q))^\top S(\sigma^1\sigma^2 - g(q))\right\rangle_u.\end{aligned}$$

Noticing that $\|\sigma^2\sigma^2 - g(q)\|_\infty \leqslant 2$ and using Lemma 8, we have

$$\frac{t}{4n}\left|\mathbb{E}\left\langle(\sigma^1\sigma^2 - g(q))^\top S(\sigma^1\sigma^2 - g(q))\right\rangle_u\right| \leqslant \frac{1}{2}\mathbb{E}\left\langle\|tS(\sigma^1\sigma^2 - g(q))\|_n\right\rangle_u = \frac{1}{2}\mathbb{E}\left\langle\|R_{12} - q\|_n\right\rangle_u \leqslant \frac{C}{K_n^{1/4}}$$

where we recall that the constant C does not depend on u. Writing

$$F_n = \varphi(1) = \varphi(0) + \int_0^1 \varphi'(u)du,$$

and using the last bound, we obtain the result of Theorem 2.

2.3 Proof of Theorem 5

In all the remainder of the paper, we shall work on the original Hamiltonian defined in the introduction, and the use of Lemma 7 will be restricted to $u = 1$.

The first bound obtained for $u = 1$ in the statement of Lemma 7 can be rewritten as

$$\max_{i\in[n]}\mathbb{E}(m_i - \operatorname{Tanh}([Wm_{(i)}]_i))^2 \leqslant C/K_n.$$

By a straightforward adaptation of this lemma, we obtain that for a fixed integer $M > 0$, it holds that

$$\max_{A\subset[n],|A|=M}\max_{i\in A^c}\mathbb{E}\left(m_{(A),i} - \operatorname{Tanh}\left(\left[Wm_{(A\cup\{i\})}\right]_i\right)\right)^2 \leqslant \frac{C}{K_n}, \tag{12}$$

This bound will be at the basis of our proof. Let us fix an integer $k > 0$. Given indices $i_1, i_2, \ldots, i_k \in [n]$ which are all different, we obtain thanks to the bound (12) that

$$\begin{aligned} m_{i_1} &= \operatorname{Tanh}\left(\left[Wm_{(i_1)}\right]_{i_1}\right) + e_1 \qquad (13)\\ m_{(i_1),i_2} &= \operatorname{Tanh}\left(\left[Wm_{(i_1,i_2)}\right]_{i_2}\right) + e_2 \\ &\cdots \\ m_{(i_1,\ldots,i_{k-1}),i_k} &= \operatorname{Tanh}\left(\left[Wm_{(i_1,\ldots,i_k)}\right]_{i_k}\right) + e_k \end{aligned}$$

with $\mathbb{E}e_l^2 \leqslant C/K_n$ for each $l \in [k]$.

This construction will be at the basis of a series of approximations of the vector m ending with the one provided by the AMP algorithm of Theorem 5. We first define the sequence of families of $\mathbb{R}^n$–valued vectors $\{y^1_{(A_1)}\}_{A_1 \subset [n], |A_1|=k-1}$, $\{y^2_{(A_2)}\}_{A_2 \subset [n], |A_2|=k-2}$, ..., $\{y^k\}$ as follows. We write $y^l_{(A_l)} = [y^l_{(A_l),i}]_{i \in [n]}$, and we set $y^l_{(A_l),i} = -h$ if $i \in A_l$ in such a way that $\operatorname{Tanh}(y^l_{(A_l),i}) = 0$ if $i \in A_l$. With this convention, given indices $i_1, i_2, \ldots, i_k \in [n]$ which are all different, we set

$$\begin{aligned}
y^1_{(i_1,\ldots,i_{k-1}),i_k} &= \left[W m_{(i_1,\ldots,i_k)} \right]_{i_k} \\
y^2_{(i_1,\ldots,i_{k-2}),i_{k-1}} &= \left[W \operatorname{Tanh}\left(y^1_{(i_1,\ldots,i_{k-1})} \right) \right]_{i_{k-1}} \\
y^3_{(i_1,\ldots,i_{k-3}),i_{k-2}} &= \left[W \operatorname{Tanh}\left(y^2_{(i_1,\ldots,i_{k-2})} \right) \right]_{i_{k-2}} \\
&\cdots \\
y^k_{i_1} &= \left[W \operatorname{Tanh}\left(y^{k-1}_{(i_1)} \right) \right]_{i_1} .
\end{aligned}$$

The same kind of construction is provided in [12]. To make things clearer to the reader, let us assume that $k = 3$. Then we have

$$y^3_{i_1} = \sum_{i_2 \notin \{i_1\}} W_{i_1 i_2} \operatorname{Tanh}\left(\sum_{i_3 \notin \{i_1, i_2\}} W_{i_2 i_3} \operatorname{Tanh}\left(\sum_{i_4 \notin \{i_1, i_2, i_3\}} W_{i_3 i_4} m_{(i_1,i_2,i_3),i_4} \right) \right).$$

We notice here that the family $\{W_{i_1 i_2}\}_{i_2}$ is independent of the family of random variables Tanh(...) that follow these terms in the first summand, the family $\{W_{i_2 i_3}\}_{i_3 \notin \{i_1, i_2\}}$ is independent from what follows, and so on. More formally, by writing $y^0_{(i_1,\ldots,i_k)} = \operatorname{Tanh}^{-1}(m_{(i_1,\ldots,i_k)})$, we have

$$y^l_{(i_1,\ldots,i_{k-l}),i_{k-l+1}} = \sum_{j \notin \{i_1,\ldots,i_{k-l+1}\}} W_{i_{k-l+1},j} \operatorname{Tanh}(y^{l-1}_{(i_1,\ldots,i_{k-l+1}),j}), \quad l \in [k],$$

and we observe that the families $\{W_{i_{k-l+1},j}\}_{j \notin \{i_1,\ldots,i_{k-l+1}\}}$ and $\{\operatorname{Tanh}(y^{l-1}_{(i_1,\ldots,i_{k-l+1}),j}\}_{j \notin \{i_1,\ldots,i_{k-l+1}\}}$ are independent for each $l \in [k]$. The same kind of remark will hold for the next two algorithms. We call this phenomenon the "independence along a path of indices".

The next algorithm is similar to the previous one except for the fact that the initial value is Tanh(0) instead of being $m_{(i_1,\ldots,i_k)}$. Namely, we define the $\mathbb{R}^n$–valued vectors $\{\tilde{x}^1_{(A_1)}\}_{A_1 \subset [n], |A_1|=k-1}$, $\{\tilde{x}^2_{(A_2)}\}_{A_2 \subset [n], |A_2|=k-2}$, ..., $\{\tilde{x}^k\}$ with $\tilde{x}^l_{(A_l)} = [\tilde{x}^l_{(A_l),i}]_{i \in [n]}$ as follows (as above, $\tilde{x}^l_{(A_l),i} = -h$ if $i \in A_l$): for indices $i_1, \ldots, i_k$ which are all different, we set

$$\begin{aligned}
\tilde{x}^1_{(i_1,\ldots,i_{k-1}),i_k} &= \left[W_{(i_1,\ldots,i_k)} \operatorname{Tanh}(0) \right]_{i_k} \\
\tilde{x}^2_{(i_1,\ldots,i_{k-2}),i_{k-1}} &= \left[W \operatorname{Tanh}\left(\tilde{x}^1_{(i_1,\ldots,i_{k-1})} \right) \right]_{i_{k-1}} \\
&\cdots \\
\tilde{x}^k_{i_1} &= \left[W \operatorname{Tanh}\left(\tilde{x}^{k-1}_{(i_1)} \right) \right]_{i_1} ,
\end{aligned}$$

where $W_{(i_1,\ldots,i_k)}$ is the matrix W in which the rows $i_1, \ldots, i_k$ and the columns $i_1, \ldots, i_k$ are set to zero.

Our next step consists in replacing the function Tanh with a polynomial. Writing

$$f(x) = \sum_{\ell=0}^{d} \alpha_\ell x^\ell$$

as a degree–d polynomial, we define the iterates $\check{x}^1_{(i_1,\dots,i_{k-1})}$, $\check{x}^2_{(i_1,\dots,i_{k-2})}$, ..., $\check{x}^k$ with the same notational conventions as above as

$$\begin{aligned}
\check{x}^1_{(i_1,\dots,i_{k-1}),i_k} &= \left[W_{(i_1,\dots,i_k)} f(0)\right]_{i_k} \\
\check{x}^2_{(i_1,\dots,i_{k-2}),i_{k-1}} &= \left[W f(\check{x}^1_{(i_1,\dots,i_{k-1})})\right]_{i_{k-1}} \\
&\cdots \\
\check{x}^k_{i_1} &= \left[W f(\check{x}^{k-1}_{(i_1)})\right]_{i_1},
\end{aligned}$$

by setting $f(\check{x}^l_{(A_l),i}) = 0$ if $i \in A_l$. The iterations for these two last algorithms can be rewritten as follows for later use. Writing $A_l = \{i_1, \dots, i_{k-l}\}$ for $l = 0, \dots, k$, we have for $l \in [k]$

$$\tilde{x}^l_{(A_l),i_{k-l+1}} = \sum_{r \notin A_{l-1}} W_{i_{k-l+1},r} \operatorname{Tanh}(\tilde{x}^{l-1}_{(A_{l-1}),r}), \quad \text{and} \tag{14}$$

$$\check{x}^l_{(A_l),i_{k-l+1}} = \sum_{r \notin A_{l-1}} W_{i_{k-l+1},r} f(\check{x}^{l-1}_{(A_{l-1}),r}), \tag{15}$$

starting with $\check{x}^0_{(A_0),r} = \tilde{x}^0_{(A_0),r} = 0$.

As said above, the three preceding algorithms share the property of the independence along a path of indices. In order to be able to use the approach of [8] and [18] as announced at the beginning of this section, we need to introduce another kind of dependence, namely the one based on the so-called Non-Backtraking (NB) iterations. At every iteration l, our next algorithm produces a family of vectors $\check{z}^l_{(j)} = [\check{z}^l_{(j),i}]_{i\in[n]}$ as follows. We initialize the algorithm with $\check{z}^0_{(j),i} = 0$, and write

$$\check{z}^{\ell+1}_{(j),i} = \sum_{r \neq j} W_{ir} f(\check{z}^{\ell}_{(i),r}).$$

Furthermore, stopping at Iteration k, we set

$$\check{z}^k_i = \sum_r W_{ir} f(\check{z}^{k-1}_{(i),r}).$$

Our last intermediate is the following AMP algorithm with the polynomial activation function f. Starting with $z^0 = 0$ and $z^1 = W f(0)$, it reads

$$z^{l+1} = W f\left(z^l\right) - \operatorname{diag}\left((W \odot W) f'(z^l)\right) f\left(z^{l-1}\right). \tag{16}$$

For a better readability, we summarize the main features of these five algorithms along with the AMP algorithm (4) in the following table:

	Activation function	Algorithm structure	Initialization
$y^l_{(\cdots)}$	Tanh	independence along a path of indices	$m_{(\cdots)}$
$\tilde{x}^l_{(\cdots)}$	Tanh	independence along a path of indices	Tanh(0)
$\check{x}^l_{(\cdots)}$	polynomial f	independence along a path of indices	$f(0)$
$\check{z}^l_{(\cdot)}$	polynomial f	NB	$f(0)$
z^l	polynomial f	AMP	$z^0 = 0, z^1 = Wf(0)$
x^l	Tanh	AMP	$x^0 = 0, x^1 = W\,\mathrm{Tanh}(0)$

We shall develop below a sequence of approximation results starting with the vector m and ending with x^k. Before we begin, some new notations and preliminary results are necessary.

We say that a real continuous function φ belongs to the set PL of pseudo-Lipschitz functions if there exists $C > 0$ and an integer $a > 0$ such that

$$|\varphi(x) - \varphi(y)| \leqslant C|x - y| \left(1 + |x|^a + |y|^a\right).$$

It is easy to show that each polynomial belongs to PL, and so is the case of the functions Tanh and Tanh$'$. Pseudo-Lipschitz functions are conveniently used as test functions in the AMP literature, see, *e.g.*, [16].

Given $i, j \in [n]$ with $i \neq j$, we also denote as $\mathscr{F}_{-i}$ and $\mathscr{F}_{-(i,j)}$ the σ–fields generated by the random variables $\{W_{kl} \ : \ k < l \text{ and } k, l \in [n]\backslash\{i\}\}$ and $\{W_{kl} \ : \ k < l \text{ and } k, l \in [n]\backslash\{i, j\}\}$ respectively. The notation $A_l^{(n)} = A_l$ will always refer to a set of indices $A_l \subset [n]$ such that $|A_l| = k - l$.

We begin with an approximation result related with the function Tanh and its derivative:

Lemma 9. Let $C > 0$. For each $e > 0$, there exists a polynomial p_e such that $p_e(0) = \mathrm{Tanh}(0)$,

$$\max_{\alpha\in[0,C]} \mathbb{E}\left(p_e(\alpha\xi) - \mathrm{Tanh}(\alpha\xi)\right)^2 \leqslant e \quad \text{and} \quad \max_{\alpha\in[0,C]} \mathbb{E}\left(p_e'(\alpha\xi) - \mathrm{Tanh}'(\alpha\xi)\right)^2 \leqslant e.$$

Proof. Given a small $\delta > 0$, it is known, see [22, Th. 1] or [15], that there exists a polynomial u on $\mathbb{R}$ such that

$$\forall x \in \mathbb{R}, \ \left|u(x) - \mathrm{Tanh}'(x)\right| \leqslant \delta \exp(\delta x^2).$$

Defining the polynomial U as

$$U(x) = \mathrm{Tanh}(0) + \int_0^x u(s)\, ds$$

we obtain that $|U(x) - \mathrm{Tanh}(x)| = |\int_0^x (u(s) - \mathrm{Tanh}'(s))\, ds| \leqslant \delta |x| \exp(\delta x^2)$. Therefore, given $\alpha > 0$ not too large, we have after a simple derivation that

$$\mathbb{E}(u(\alpha\xi) - \mathrm{Tanh}'(\alpha\xi))^2 \leqslant \delta^2/\sqrt{1 - \delta\alpha^2}, \quad \text{and}$$
$$\mathbb{E}(U(\alpha\xi) - \mathrm{Tanh}(\alpha\xi))^2 \leqslant \delta\alpha^2/(1 - \delta\alpha^2)^{3/2}.$$

By assumption, $0 \leqslant \alpha^2 \leqslant C^2$. Thus, by setting δ small enough, we can take $p_e = U$. $\square$

Given a polynomial f, we need to introduce the sequence $(\check{q}^l)_{l\in\mathbb{N}}$ of $\mathbb{R}_+^n$–valued vectors defined recursively as

$$\check{q}^0 = 0, \quad \check{q}^{l+1} = tS\mathbb{E}f\left(\check{X}^l\right), \tag{17}$$

where $\check{X}^l \sim \mathcal{N}(0, \mathrm{diag}(\check{q}^l))$.

Lemma 10. The iterations q^l defined in the statement of Theorem 5 satisfy $\sup_l \|q^l\|_\infty \leqslant \log 2$. Let e be a positive number such that $\sqrt{e} \leqslant (1 - \log 2)/10$. Let p_e be a polynomial such that $p_e(0) = \mathrm{Tanh}(0)$ and

$$\max_{\alpha\in[0,1]} \mathbb{E}\left(p_e(\alpha\xi) - \mathrm{Tanh}(\alpha\xi)\right)^2 \leqslant e,$$

which existence is guaranteed by Lemma 9. Consider the iterates $\check{q}^l$ provided by equations (17). with $f = p_e$. Then, $\sup_l \|\check{q}^l\|_\infty \leqslant 1$, and $\sup_l \|\check{q}^l - q^l\|_\infty \leqslant 10\sqrt{e}$.

Proof. Recall the expression of g in (1). The bound on $\|q^l\|_\infty$ follows from $q^0 = 0$, $0 \leqslant g(q) < 1$ and $|||tS||| < \log 2$.

We now show that $\|\check{q}^l\|_\infty \leqslant 1$ and $\|\check{q}^l - q^l\|_\infty \leqslant 10\sqrt{e}$ by recurrence on l. This is trivial for $l = 0$ since $\check{q}^0 = q^0$. Assume that this is true for l. Since $x^2 - y^2 = 2y(x - y) + (x - y)^2$, we have for each $a \in [0, 1]$ that

$$|\mathbb{E}p_e(a\xi)^2 - \mathbb{E}\,\mathrm{Tanh}(a\xi)^2| \leqslant 2\mathbb{E}|p_e(a\xi) - \mathrm{Tanh}(a\xi)| + \mathbb{E}(p_e(a\xi) - \mathrm{Tanh}(a\xi))^2 \leqslant 3\sqrt{e}.$$

Remembering that g is 1–Lipschitz (see the proof of Lemma 1), we obtain from what precedes that

$$\begin{aligned}\|\check{q}^{l+1} - q^{l+1}\|_\infty &\leqslant (\log 2)\|\mathbb{E}p_e(\check{X}^l)^2 - g(q^l)\|_\infty \leqslant (\log 2)\left(\|\check{q}^l - q^l\|_\infty + \|\mathbb{E}p_e(\check{X}^l)^2 - g(\check{q}^l)\|_\infty\right) \\ &\leqslant 13(\log 2)\sqrt{e} \leqslant 10\sqrt{e},\end{aligned}$$

and furthermore, $\|\check{q}^{l+1}\|_\infty \leqslant \log 2 + 10\sqrt{e} \leqslant 1$. $\square$

We now turn to the approximation results alluded to above.

Lemma 11. It holds that $\mathbb{E}(\mathrm{Tanh}(y_i^k) - m_i)^2 \leqslant C/K_n$.

Proof. It is clear from the bound (12) that $\mathbb{E}\left(\mathrm{Tanh}(y^1_{(i_1,\ldots,i_{k-1}),i_k}) - m_{(i_1,\ldots,i_{k-1}),i_k}\right)^2 \leqslant C/K_n$. Given $l \geqslant 2$, assume that $\mathbb{E}\left(\mathrm{Tanh}(y^l_{(i_1,\ldots,i_{k-l}),i_{k-l+1}}) - m_{(i_1,\ldots,i_{k-l}),i_{k-l+1}}\right)^2 \leqslant C/K_n$. Since Tanh is 1-Lipschitz, we have thanks to the independence along a path of indices property that

$$\begin{aligned}
&\mathbb{E}\left(\mathrm{Tanh}(y^{l+1}_{(i_1,\ldots,i_{k-l-1}),i_{k-l}}) - m_{(i_1,\ldots,i_{k-l-1}),i_{k-l}}\right)^2 \\
&\qquad = \mathbb{E}\left(\mathrm{Tanh}\left(\left[W\,\mathrm{Tanh}(y^l_{(i_1,\ldots,i_{k-l})})\right]_{i_{k-l}}\right) - \mathrm{Tanh}\left(\left[Wm_{(i_1,\ldots,i_{k-l})}\right]_{i_{k-l}}\right) - e_l\right)^2
\end{aligned}$$

where $\mathbb{E}e_l^2 \leqslant C/K_n$. Therefore,

$$\begin{aligned}
&\mathbb{E}\left(\mathrm{Tanh}(y^{l+1}_{(i_1,\ldots,i_{k-l-1}),i_{k-l}}) - m_{(i_1,\ldots,i_{k-l-1}),i_{k-l}}\right)^2 \\
&\leqslant 2\mathbb{E}\left(\left[W\,\mathrm{Tanh}(y^l_{(i_1,\ldots,i_{k-l})})\right]_{i_{k-l}} - \left[Wm_{(i_1,\ldots,i_{k-l})}\right]_{i_{k-l}}\right)^2 + C/K_n \\
&\leqslant 2\sum_r ts_{i_{k-l},r}\mathbb{E}\left(\mathrm{Tanh}(y^l_{(i_1,\ldots,i_{k-l}),r}) - m_{(i_1,\ldots,i_{k-l}),r}\right)^2 + C/K_n \\
&\leqslant C/K_n.
\end{aligned}$$

□

Lemma 12. It holds that $\mathbb{E}(\tilde{x}_i^k - y_i^k)^2 \leqslant 4(\log 2)^k$.

Proof. By recurrence. For different indices $i_1, \ldots, i_k$, we have

$$\mathbb{E}(\tilde{x}^1_{(i_1,\ldots,i_{k-1}),i_k} - y^1_{(i_1,\ldots,i_{k-1}),i_k})^2 \leqslant t\sum_r s_{i_k,r}\mathbb{E}(\mathrm{Tanh}(0) - m_{(i_1,\ldots,i_k),r})^2 \leqslant 4\log 2.$$

For $l \geqslant 1$, assume that $\mathbb{E}(\tilde{x}^l_{(i_1,\ldots,i_{k-l}),i_{k-l+1}} - y^l_{(i_1,\ldots,i_{k-l}),i_{k-l+1}})^2 \leqslant 4(\log 2)^l$. Then, by using the independence along a path of indices property and doing the same calculation as above, we get that $\mathbb{E}(\tilde{x}^{l+1}_{(i_1,\ldots,i_{k-l-1}),i_{k-l}} - y^{l+1}_{(i_1,\ldots,i_{k-l-1}),i_{k-l}})^2 \leqslant 4(\log 2)^{l+1}$. □

Proposition 13. For each $b > 0$ and each $l \in [k]$, it holds that

$$\sup_n \max_{A_l^{(n)}} \max_{i\in[n]} \mathbb{E}\left|\check{x}^{(n),l}_{(A_l^{(n)}),i}\right|^b < \infty. \tag{18}$$

For a function $\varphi \in \mathrm{PL}$, an integer $l \in [k]$, a sequence of sets $\mathcal{S}^{(n)} \in [n]$ with $|\mathcal{S}^{(n)}| \to \infty$, a sequence of $|\mathcal{S}^{(n)}|$–tuples $(\beta_i^{(n)})_{i\in\mathcal{S}^{(n)}}$ such that $|\beta_i^{(n)}| \leqslant 1$, a sequence of sets of the type $A_l^{(n)}$, and a number $b > 0$ which are all arbitrary, it holds that

$$\mathbb{E}\left|\frac{1}{|\mathcal{S}^{(n)}|}\sum_{i\in\mathcal{S}^{(n)}} \beta_i^{(n)}\varphi(\check{x}^{(n),l}_{(A_l^{(n)}),i}) - \beta_i^{(n)}\mathbb{E}\varphi(\check{X}_i^{(n),l})\right|^b \xrightarrow[n\to\infty]{} 0. \tag{19}$$

Proof. In the expression (15) of $\check{x}^l_{(A_l),i_{k-l+1}}$, we have set $|A_l| = k - l$. We need to extend a bit this expression to include a set of indices $B \subset [n]$ that might be larger than A_l by writing for $i \notin B$:

$$\check{x}^l_{(B),i} = \sum_{r\notin B} W_{ir} f(\check{x}^{l-1}_{(B\cup\{i\}),r}),$$

which provides consistent iterations one we set $\check{x}^0_{(\cdot),\cdot} = 0$. We also write

$$(\varsigma^l_{(B),i})^2 = t \sum_{r \notin B} s_{ir} f(\check{x}^{l-1}_{(B \cup \{i\}),r})^2.$$

We notice that $(\varsigma^l_{(B),i})^2$ is the conditional variance of $\check{x}^l_{(B),i}$ given $\mathscr{F}_{-i}$, a fact that we shall use repeatidly in the proof.

The moment bound (18) can be proven by recurrence on l. For $l = 1$, consider a set A_1 and an index $i \notin A_1$. Since $\check{x}^1_{(A_1),i} = f(0)\sum_{r \notin A_1} W_{ir}$, it is clear that the bound (18) holds true for $l = 1$. Assuming (18) is true for l, let $i \in A_l$ and $A_{l+1} = A_l \backslash \{i\}$. We have here $\check{x}^{l+1}_{(A_{l+1}),i} = \sum_{r \notin A_{l+1}} W_{ir} f(\check{x}^l_{(A_l),r})$, and thus,

$$\mathbb{E}\left|\check{x}^{l+1}_{(A_{l+1}),i}\right|^b = \mathbb{E}\mathbb{E}\left[\left|\check{x}^{l+1}_{(A_{l+1}),i}\right|^b \mid \mathscr{F}_{-i}\right] = \mathbb{E}(\varsigma^{l+1}_{(A_{l+1}),i})^b \mathbb{E}|\xi|^b$$

which is bounded by the recurrence assumption.

We now prove by recurrence on l the convergence (19) as well as

$$\mathbb{E}(\check{x}^l_{(A_l),r} - \check{x}^l_{(A_l \cup \{i\}),r})^2 \to 0 \tag{20}$$

for all sequences $(A^{(n)}_l)$, (i_n) with $i_n \notin A^{(n)}_l$, and (r_n) with $r_n \notin A^{(n)}_l \cup \{i_n\}$.

We first notice that at the left hand side of (19), the terms for which $i \in A_l$ have a negligible contribution. Thus, in all the remainder of the proof, we can assume without generality loss that $\mathcal{S} \cap A_l = \varnothing$ for each $l \in [k]$.

Let us start our recurrence with $l = 1$. As above, consider a set A_1 and an index $i \notin A_1$. We have

$$\mathbb{E}\varphi(\check{x}^1_{(A_1),i}) = \mathbb{E}\varphi\left(f(0)\sum_{r \notin A_1} W_{ir}\right) = \mathbb{E}\varphi\left(\varsigma^1_{(A_1),i}\xi\right).$$

We also have $|\varphi(x) - \varphi(y)| \leqslant C|x - y|(1 + |x|^a + |y|^a)$ for some $C, a > 0$. Recalling that $\check{X}^l = [\check{X}^l_i] \sim \mathcal{N}(0, \mathrm{diag}(\check{q}^l))$, we can write that $\check{X}^1_i = \sqrt{\check{q}^1_i}\xi$. Note also that $\check{q}^1_i = tf(0)^2 \sum_r s_{ir}$ and that $(\varsigma^1_{(A_1),i})^2 = tf(0)^2 \sum_{r \notin A_l} s_{ir}$ with $|A_1| = k - 1$. With this, we have

$$\left|\mathbb{E}\varphi(\check{x}^1_{(A_1),i}) - \mathbb{E}\varphi(\check{X}^1_i)\right| \leqslant C|\varsigma^1_{(A_1),i} - \sqrt{\check{q}^1_i}|\mathbb{E}|\xi|(1 + (\varsigma^1_{(A_1),i})^a|\xi|^a + (\check{q}^1_i)^{a/2}|\xi|^a)$$

which converges to zero for each sequence of sets $(A^{(n)}_1)$ and sequence of indices (i_n). We now show that

$$\frac{1}{|\mathcal{S}|^2}\mathbb{E}\left(\sum_{i \in \mathcal{S}} \beta_i \varphi(\check{x}^1_{(A_1),i}) - \beta_i \mathbb{E}\varphi(\check{X}^1_i)\right)^2 \to 0. \tag{21}$$

By developing the square, we obtain a sum $\sum_{i,j \in \mathcal{S}} \cdots$. The diagonal $i = j$ is easily shown to be $\mathcal{O}(1/|\mathcal{S}|)$. Let us show that for $i \neq j$, it holds that

$$\mathbb{E}\left(\varphi(\check{x}^1_{(A_1),i}) - \mathbb{E}\varphi(\check{X}^1_i)\right)\left(\varphi(\check{x}^1_{(A_1),j}) - \mathbb{E}\varphi(\check{X}^1_j)\right) \to 0. \tag{22}$$

Since we showed that $\mathbb{E}\varphi(\check{x}^1_{(A_1),i}) - \mathbb{E}\varphi(\check{X}^1_i) \to 0$, it remains to show that $\mathbb{E}\varphi(\check{x}^1_{(A_1),i})\varphi(\check{x}^1_{(A_1),j}) - \mathbb{E}\varphi(\check{X}^1_i)\mathbb{E}\varphi(\check{X}^1_j) \to 0$ to obtain (22). We first have

$$\mathbb{E}(\check{x}^1_{(A_1),i} - \check{x}^1_{(A_1 \cup \{j\}),i})^2 = f(0)^2 \mathbb{E}W^2_{ij} \leqslant C/K_n$$

which establishes in passing (20) for $l = 1$. Also, by using the pseudo-Lipschitz property of φ, we have (details for obtaining the terms $o_n(1)$ below omitted):

$$\begin{aligned}
\mathbb{E}\varphi(\check{x}^1_{(A_1),i})\varphi(\check{x}^1_{(A_1),j}) &= \mathbb{E}\varphi(\check{x}^1_{(A_1\cup\{j\}),i})\varphi(\check{x}^1_{(A_1\cup\{i\}),j}) + o_n(1) \\
&= \mathbb{E}\varphi\left(f(0)\sum_{r\notin A_1\cup\{j\}} W_{ir}\right)\varphi\left(f(0)\sum_{r\notin A_1\cup\{i\}} W_{jr}\right) + o_n(1) \\
&= \mathbb{E}\varphi(\varsigma^1_{(A_1\cup\{j\}),i}\xi)\mathbb{E}\varphi(\varsigma^1_{(A_1\cup\{i\}),j}\xi) + o_n(1) \\
&= \mathbb{E}\varphi(\check{X}^1_i)\mathbb{E}\varphi(\check{X}^1_j) + o_n(1),
\end{aligned}$$

hence (22). By the moment bound (18) and dominated convergence, the convergence (21) holds true.

Given any moment $b > 2$, we also have

$$\sup \frac{1}{|\mathcal{S}|}\left(\mathbb{E}\left|\sum_{i\in\mathcal{S}}\beta_i\varphi(\check{x}^1_{(A_1),i}) - \beta_i\mathbb{E}\varphi(\check{X}^1_i)\right|^b\right)^{1/b} \leqslant \sup\frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}}\left(\mathbb{E}\left|\varphi(\check{x}^1_{(A_1),i}) - \mathbb{E}\varphi(\check{X}^1_i)\right|^b\right)^{1/b} < \infty$$

thanks to (18), where the sup is taken on $\mathcal{S}^{(n)}$, $(\beta_i^{(n)})_{i\in\mathcal{S}^{(n)}}$, and $A_1^{(n)}$. The convergence (19) follows for $l = 1$.

Assume the recurrence assumption is true for l. Letting $i \in A_l$ and $A_{l+1} = A_l\backslash\{i\}$, we have

$$\mathbb{E}\varphi(\check{x}^{l+1}_{(A_{l+1}),i}) = \mathbb{E}\mathbb{E}\left[\varphi\left(\sum_{r\notin A_l} W_{ir} f(\check{x}^l_{(A_l),r})\right) \mid \mathscr{F}_i\right] = \mathbb{E}\varphi(\varsigma^{l+1}_{(A_{l+1}),i}\xi)$$

Recalling that $\check{q}_i^{l+1} = t\sum_r s_{ir}\mathbb{E}f(\check{X}^l_r)^2$, we have by using the recurrence assumption and the fact that $|A_l| = k - l$ is fixed that

$$\mathbb{E}\left((\varsigma^{l+1}_{(A_{l+1}),i})^2 - \check{q}_i^{l+1}\right)^2 \to 0.$$

Writing $\check{X}^{l+1}_i = \sqrt{\check{q}_i^{l+1}}\xi$, we have

$$|\mathbb{E}\varphi(\varsigma^{l+1}_{(A_{l+1}),i}\xi) - \mathbb{E}\varphi(\check{X}^{l+1}_i)| \leqslant C\mathbb{E}|(\varsigma^{l+1}_{(A_{l+1}),i} - \sqrt{\check{q}_i^{l+1}}|)|\xi|(1 + (\varsigma^{l+1}_{(A_{l+1}),i})^a|\xi|^a + (\check{q}_i^{l+1})^{a/2}|\xi|^a),$$

we then have that

$$\mathbb{E}\varphi(\check{x}^{l+1}_{(A_{l+1}),i}) - \mathbb{E}\varphi(\check{X}^{l+1}_i) \to 0,$$

by Cauchy-Schwarz and the bound (18).

We now show that

$$\frac{1}{|\mathcal{S}|^2}\mathbb{E}\left(\sum_{i\in\mathcal{S}}\beta_i\varphi(\check{x}^{l+1}_{(A_{l+1}),i}) - \beta_i\mathbb{E}\varphi(\check{X}^{l+1}_i)\right)^2 \to 0. \tag{23}$$

As for the case $l = 1$, this will be true if we prove that $\mathbb{E}\varphi(\check{x}^{l+1}_{(A_{l+1}),i})\varphi(\check{x}^{l+1}_{(A_{l+1}),j}) - \mathbb{E}\varphi(\check{X}^{l+1}_i)\mathbb{E}\varphi(\check{X}^{l+1}_j) \to 0$ for $i \neq j$. We write

$$\begin{aligned}
\check{x}^{l+1}_{(A_{l+1}),i} - \check{x}^{l+1}_{(A_{l+1}\cup\{j\}),i} &= \sum_{r\notin A_{l+1}\cup\{j\}} W_{ir}\left(f(\check{x}^l_{(A_l),r}) - f(\check{x}^l_{(A_l\cup\{j\}),r})\right) + W_{ij}f(\check{x}^l_{(A_l),j}) \\
&= \chi_1 + \chi_2.
\end{aligned}$$

We obviously have $\mathbb{E}\chi_2^2 \to 0$. Moreover, since the random vectors $[W_{ir}]_r$ and $[\check{x}^l_{(A_l),r}, \check{x}^l_{(A_l\cup\{j\}),r}]_r$ in the expression above are independent, we have

$$\mathbb{E}\chi_1^2 = \mathbb{E}\mathbb{E}\left[\chi_1^2 \mid \mathscr{F}_{-i}\right] = \sum_{r\notin A_{l+1}\cup\{j\}} ts_{ir}\mathbb{E}\left(f(\check{x}^l_{(A_l),r}) - f(\check{x}^l_{(A_l\cup\{j\}),r})\right)^2$$

which converges to zero by using the recurrence assumption (20), the bound (18), and the pseudo-Lipschitz property of f as a polynomial. We thus obtain that

$$\mathbb{E}\left(\breve{x}^{l+1}_{(A_{l+1}),i} - \breve{x}^{l+1}_{(A_{l+1}\cup\{j\}),i}\right)^2 \to 0,$$

and the convergence (20) is true for $l+1$. Using the pseudo-Lipschitz property of φ, this last result, and the bound (18), we also have

$$\begin{aligned}\mathbb{E}\varphi(\breve{x}^{l+1}_{(A_{l+1}),i})\varphi(\breve{x}^{l+1}_{(A_{l+1}),j}) &= \mathbb{E}\varphi(\breve{x}^{l+1}_{(A_{l+1}\cup\{j\}),i})\varphi(\breve{x}^{l+1}_{(A_{l+1}\cup\{i\}),j}) + o_n(1)\\ &= \mathbb{E}\varphi(\varsigma^{l+1}_{(A_{l+1}\cup\{j\}),i}\xi_1)\varphi(\varsigma^{l+1}_{(A_{l+1}\cup\{i\}),j}\xi_2) + o_n(1),\end{aligned}$$

where $[\xi_1, \xi_2]^\top \sim \mathcal{N}(0, I_2)$ is a vector independent of everything else. The remainder of the proof is similar to the case $l = 1$ with the difference that now we need the recurrence assumption to obtain that $\mathbb{E}(\varsigma^{l+1}_{(A_{l+1}\cup\{j\}),i} - \sqrt{\breve{q}^{l+1}_i})^2 \to 0$. This leads to (23), and this convergence can be upgraded to any moment thanks to (18). The proof of Proposition 13 is complete. □

Lemma 14. For each $\varepsilon > 0$, there is a polynomial f_ε such that the iterates (15) with $f = f_\varepsilon$ satisfy

$$\limsup_n \max_{i\in[n]} \mathbb{E}(\breve{x}^k_i - \tilde{x}^k_i)^2 \leqslant \varepsilon.$$

This polynomial can be chosen in such a way that $f_\varepsilon(0) = \operatorname{Tanh}(0)$, and

$$\max_{\alpha\in[0,\sqrt{2}]} \mathbb{E}\left(f_\varepsilon(\alpha\xi) - \operatorname{Tanh}(\alpha\xi)\right)^2 \leqslant e$$

for some $e > 0$ that depends only on ε.

Proof. Given a small $e > 0$, Lemma 9 shows that there exists a polynomial p_e such that $p_e(0) = \operatorname{Tanh}(0)$, and

$$\max_{\alpha\in[0,\sqrt{2}]} \mathbb{E}\left(p_e(\alpha\xi) - \operatorname{Tanh}(\alpha\xi)\right)^2 \leqslant e.$$

Unfolding the iterations (15) with $f = p_e$, we shall show by recurrence on l that for each $l \in [k]$,

$$\begin{aligned}&\limsup_n \max_{A_l,i} \mathbb{E}(\breve{x}^l_{(A_l),i} - \tilde{x}^l_{(A_l),i})^2 \leqslant Ce, \text{ and}\\ &\limsup_n \max_{A_l,i} \mathbb{E}\left(p_e(\breve{x}^l_{(A_l),i}) - \operatorname{Tanh}(\tilde{x}^l_{(A_l),i})\right)^2 \leqslant Ce,\end{aligned}$$

where $C > 0$ is a constant that can change from an iteration to another. Setting $e = \varepsilon/C$ at the k^{th} iteration, we obtain our result with $f_\varepsilon = p_e$.

The following results are needed before starting our recurrence. For any polynomial f, we have from the expressions (14) and (15) that for each A_{l+1} and each $i \notin A_{l+1}$ that

$$\mathscr{L}\left(\begin{bmatrix}\breve{x}^{l+1}_{(A_{l+1}),i}\\ \tilde{x}^{l+1}_{(A_{l+1}),i}\end{bmatrix} \mid \mathscr{F}_i\right) = \mathcal{N}\left(0, R^{l+1}_{(A_{l+1}),i}\right) \quad \text{with} \quad R^{l+1}_{(A_{l+1}),i} = \begin{bmatrix}[R^{l+1}_{(A_{l+1}),i}]_{11} & [R^{l+1}_{(A_{l+1}),i}]_{12}\\ [R^{l+1}_{(A_{l+1}),i}]_{21} & [R^{l+1}_{(A_{l+1}),i}]_{22}\end{bmatrix}$$

satisfying

$$\begin{aligned}&[R^{l+1}_{(A_{l+1}),i}]_{11} = t\sum_{r\notin A_l} s_{ir} f(\breve{x}^l_{(A_l),r})^2, \quad [R^{l+1}_{(A_{l+1}),i}]_{22} = t\sum_{r\notin A_l} s_{ir}\operatorname{Tanh}(\tilde{x}^l_{(A_l),r})^2, \text{ and}\\ &[R^{l+1}_{(A_{l+1}),i}]_{12} = t\sum_{r\notin A_l} s_{ir} f(\breve{x}^l_{(A_l),r})\operatorname{Tanh}(\tilde{x}^l_{(A_l),r})\end{aligned}$$

with $A_l = A_{l+1} \cup \{i\}$. By consequence,

$$\mathscr{L}\left(\breve{x}^{l+1}_{(A_{l+1}),i} - \tilde{x}^{l+1}_{(A_{l+1}),i} \mid \mathscr{F}_i\right) = \mathcal{N}\left(0, t\sum_{r\notin A_l} s_{ir}\left(f(\breve{x}^l_{(A_l),r}) - \operatorname{Tanh}(\tilde{x}^l_{(A_l),r})\right)^2\right).$$

We now tackle our recurrence. For $l = 1$, considering a set A_1 and an index $i \notin A_1$, we have $\check{x}^1_{(A_1),i} = \tilde{x}^1_{(A_1),i} \sim \mathcal{N}(0, [R^1_{(A_1),i}]_{ab})$ with

$$\forall a, b \in \{1, 2\}, \quad [R^1_{(A_1),i}]_{ab} = t \operatorname{Tanh}(0)^2 \sum_{r \notin A_1} s_{ir} \leqslant 1$$

by Lemma 10. By the construction of p_e, we therefore have that

$$\mathbb{E}\left(p_e(\check{x}^1_{(A_1),i}) - \operatorname{Tanh}(\tilde{x}^1_{(A_1),i})\right)^2 \leqslant e,$$

and the recurrence assumption is true for $l = 1$. Assume it is for $l > 1$. Let $i \in A_l$ and $A_{l+1} = A_l \backslash \{i\}$. Then we have

$$\begin{aligned}\mathbb{E}(\check{x}^{l+1}_{(A_{l+1}),i} - \tilde{x}^{l+1}_{(A_{l+1}),i})^2 &= \mathbb{E}\mathbb{E}\left[(\check{x}^{l+1}_{(A_{l+1}),i} - \tilde{x}^{l+1}_{(A_{l+1}),i})^2 \mid \mathscr{F}_{-i}\right] \\ &= t \sum_{r \notin A_{l+1}} s_{ir} \mathbb{E}\left(p_e(\check{x}^l_{(A_l),r}) - \operatorname{Tanh}(\tilde{x}^l_{(A_l),r})\right)^2\end{aligned}$$

which is bounded by Ce by the recurrence assumption. Since Tanh is Lipschitz, we can furthermore write

$$\begin{aligned}\mathbb{E}\left(p_e(\check{x}^{l+1}_{(A_{l+1}),i}) - \operatorname{Tanh}(\tilde{x}^{l+1}_{(A_{l+1}),i})\right)^2 \leqslant 2\mathbb{E}&\left(p_e(\check{x}^{l+1}_{(A_{l+1}),i}) - \operatorname{Tanh}(\check{x}^{l+1}_{(A_{l+1}),i})\right)^2 \\ &+ 2\mathbb{E}(\check{x}^{l+1}_{(A_{l+1}),i} - \tilde{x}^{l+1}_{(A_{l+1}),i})^2,\end{aligned}$$

and we need to control the first term at the right hand side. Using the bound (18), we can write

$$\begin{aligned}&\mathbb{E}\left(p_e(\check{x}^{l+1}_{(A_{l+1}),i}) - \operatorname{Tanh}(\check{x}^{l+1}_{(A_{l+1}),i})\right)^2 \\ &\leqslant \mathbb{E}\mathbb{E}\left[\left(p_e(\check{x}^{l+1}_{(A_{l+1}),i}) - \operatorname{Tanh}(\check{x}^{l+1}_{(A_{l+1}),i})\right)^2 \mid \mathscr{F}_{-i}\right] \mathbb{1}_{[R^{l+1}_{(A_{l+1}),i}]_{11} \leqslant 2} + C\mathbb{P}\left[[R^{l+1}_{(A_{l+1}),i}]_{11} > 2\right]^{1/2}.\end{aligned}$$

The first term at the right hand side is bounded by e by the construction of p_e. Regarding the second term, invoking the previous proposition with $\mathcal{S} = \{r \in [n] : s_{ir} > 0, r \notin A_l\}$, and using that $|A_l|$ is fixed, we obtain that

$$\max_{A_{l+1},i} \mathbb{E}\left([R^{l+1}_{(A_{l+1}),i}]_{11} - \check{q}^{l+1}_i\right)^2 \xrightarrow[n\to\infty]{} 0.$$

Since $\check{q}^{l+1}_i \leqslant 1$, we obtain that $\max_{A_{l+1},i} \mathbb{P}\left[[R^{l+1}_{(A_{l+1}),i}]_{11} > 2\right] \to 0$, and the recurrence assumption is verified for $l + 1$. □

We now manage the iterates $\check{x}^l$, $\check{z}^l$ and z^l which are all built around a polynomial activation function. Due to this polynomial nature, we can express these terms with the help of a tree formalism. The tree structure below is a simplification of the structure of [8].

Let $T = (V(T), E(T))$ be a rooted tree with the vertex set $V(T)$ and edge set $E(T)$. The root of this tree is denoted $\circ$, and the distance of a vertex u to $\circ$ is $|u|$. The root has one child (thus, this is a planted tree), and every vertex other than $\circ$ can have up to d children. We denote as $\pi(u)$ the parent of the vertex u, where the vertices are oriented towards the root. Thus, $u \to v$ is equivalent to $v = \pi(u)$. Every vertex v has a label $\ell(v) \in [n]$. The number of children of v is denoted $c(v)$. We denote as $L(T)$ the set of leaves of T. If a leaf $v \in L(T)$ has a maximal depth in the tree, we attribute to this leaf a number $c(v) \in \{0, \dots d\}$ as if this vertex was the parent of $c(v)$ children which were pruned from the tree. If the depth of this leaf $v \in L(T)$ is not maximal, then we keep the natural value $c(v) = 0$. We also use the following notations: $\overline{\mathcal{T}}^k$ is the set of such labelled trees, with depth k at most. $\check{\mathcal{T}}^k \subset \overline{\mathcal{T}}^k$ is the subset that satisfies the following condition: there is no path $v_1 = \circ \leftarrow v_2 \leftarrow \cdots \leftarrow v_i$ in which there exists two identical labels $\ell(\cdot)$. Furthermore, $\mathcal{T}^k \subset \overline{\mathcal{T}}^k$ is the subset that satisfies the following non-backtracking condition: if $v_1 = \circ \leftarrow v_2 \leftarrow \cdots \leftarrow v_i$, then the corresponding sequence of labels $\ell(\cdot)$ non-backtracking. This means that for each $j \in [i-2]$, the three labels $\ell(v_j)$, $\ell(v_{j+1})$ and $\ell(v_{j+2})$ are distinct. $\mathcal{T}_i^k \subset \mathcal{T}^k$ is the subset of trees in $\mathcal{T}^k$ for which the label $\ell(\circ)$ of the root is i, and of course, the label of the child v of the root satisfies $\ell(v) \neq i$. We apply a similar definition for $\check{\mathcal{T}}_i^k$. Notice that $\check{\mathcal{T}}_i^k \subset \mathcal{T}_i^k$. Such a tree will be called a non-backtracking tree (NBT).

Given a tree T, write

$$W(T) = \prod_{(u \to v) \in E(T)} W_{\ell(u)\ell(v)}, \quad \Gamma(T) = \prod_{(u \to v) \in E(T)} \alpha_{c(u)}, \quad \text{and} \quad \check{x}(T) = \prod_{v \in L(T)} (\check{x}^0_{\ell(v)})^{c(v)},$$

where we recall that the α_ℓ's are the coefficients of the polynomial f.

With this formalism, similarly to [8, Lemma 1], we have

$$\check{z}_i^k = \sum_{T \in \mathcal{T}_i^k} W(T)\Gamma(T)\check{x}(T) \quad \text{and} \quad \check{x}_i^k = \sum_{T \in \check{\mathcal{T}}_i^k} W(T)\Gamma(T)\check{x}(T).$$

Given an integer $b > 0$ and b trees $T_1, \ldots, T_b \in \overline{\mathcal{T}}_i^k$, define the graph $G = \boldsymbol{G}(T_1, \ldots, T_b)$ as being the rooted, undirected, and labelled graph obtained by merging the nodes of these trees that have the same label $\ell(\cdot)$. This common label will be the label of the resulting node in G. Of course, the root node $\circ$ of G will have the label $\ell(\circ) = i$. The other nodes are numbered, say, in the increasing order of their labels. The edges of G are furthermore unweighted.

For a tree T labelled as above and for $j, l \in [n]$, define

$$\phi(T)_{jl} = \left|\{(u \to v) \in E(T), \ \{\ell(u), \ell(v)\} = \{j, l\}\}\right|.$$

Finally, recalling Assumption 3, we define the set $\mathcal{K}$ as

$$\mathcal{K} = \{\{i, j\} \subset [n], \ s_{ij} > 0\},$$

and the section $\mathcal{K}_i$ for $i \in [n]$ as

$$\mathcal{K}_i = \{j \in [n], \ s_{ij} > 0\}.$$

The following lemma is proven in Appendix A.4.

Lemma 15. Let $b, r \geqslant 2$ be two integers. Let $\mathcal{A} \subset (\mathcal{T}_i^k)^{\otimes b}$ be such that each b–tuple $(T_1, \ldots, T_b) \in \mathcal{A}$ satisfies the two following conditions:

- $|V(\boldsymbol{G}(T_1, \ldots, T_b))| \leqslant r$.
- When $\sum_{k=1}^b \phi(T_k)_{jl} > 0$, it holds that $\{j, l\} \in \mathcal{K}$.

Then,

$$|\mathcal{A}| \leqslant C K_n^{r-1}.$$

Proposition 16. For each even integer $b \geqslant 2$, it holds that $\mathbb{E}\left(\check{x}_i^k - \check{z}_i^k\right)^b \leqslant C/K_n$ and $\mathbb{E}\left(\check{z}_i^k - z_i^k\right)^b \leqslant C/K_n$.

Combining the results of this proposition with the bound (18), we obtain that

$$\forall b > 0, \ \sup_n \max_{i \in [n]} \mathbb{E}\left|z_i^l\right|^b < \infty \tag{24}$$

for each $b > 0$, a bound that will be useful later.

Proof. We know that for each $T_1, \dots, T_b \in \mathcal{T}_i^k \backslash \breve{\mathcal{T}}_i^k$, it holds that $\sum_{j<l} \phi(T_1)_{jl} + \cdots + \phi(T_b)_{jl} \leqslant C_b$ with $C_b = b(1 + d + \cdots + d^{k-1})$. For $r \in [C_b]$, define the set $\mathcal{C}_i(r)$ as

$$\begin{aligned}\mathcal{C}_i(r) = \Big\{ (T_1, \dots, T_b) \,:\, & T_1, \dots, T_b \in \mathcal{T}_i^k \backslash \breve{\mathcal{T}}_i^k, \\ & \forall j < l,\ \phi(T_1)_{jl} + \cdots + \phi(T_b)_{jl} \neq 1, \\ & \forall j < l,\ \phi(T_1)_{jl} + \cdots + \phi(T_b)_{jl} > 0 \ \Rightarrow \{j, l\} \in \mathcal{K}, \\ & \sum_{j<l} \phi(T_1)_{jl} + \cdots + \phi(T_b)_{jl} = r \Big\}.\end{aligned}$$

With this definition, we have

$$\begin{aligned}\mathbb{E}(\breve{z}_i^k - \breve{x}_i^k)^b &= \mathbb{E} \sum_{T_1, \dots, T_b \in \mathcal{T}_i^k \backslash \breve{\mathcal{T}}_i^k} \Gamma(T_1) \dots \Gamma(T_b) x(T_1) \dots x(T_b) W(T_1) \dots W(T_b) \\ &= \mathbb{E} \sum_{r=2}^{C_b} \sum_{(T_1, \dots, T_b) \in \mathcal{C}_i(r)} \Gamma(T_1) \dots \Gamma(T_b) x(T_1) \dots x(T_b) W(T_1) \dots W(T_b) \\ &\leqslant C \sum_{r=2}^{C_b} \sum_{(T_1, \dots, T_b) \in \mathcal{C}_i(r)} \left| \mathbb{E} W(T_1) \dots W(T_b) \right|.\end{aligned}$$

Fix r and assume that $\mathcal{C}_i(r) \neq \varnothing$. For each $(T_1, \dots, T_b) \in \mathcal{C}_i(r)$, we have

$$\left| \mathbb{E} W(T_1) \dots W(T_b) \right| = \prod_{j<l} \left| \mathbb{E} W_{jl}^{\phi(T_1)_{jl} + \cdots + \phi(T_b)_{jl}} \right| \leqslant C K_n^{-r/2}.$$

We need to show that $|\mathcal{C}_i(r)| \leqslant C K_n^{r/2-1}$ to obtain the first bound in the statement. Given $T_1, \dots, T_b \in \mathcal{C}_i(r)$, the graph $G = \boldsymbol{G}(T_1, \dots, T_b)$ satisfies

$$|E(G)| = \sum_{j<l} \mathbb{1}_{\phi(T_1)_{jl} + \cdots + \phi(T_b)_{jl} \geqslant 2},$$

which shows that $|E(G)| \leqslant r/2$. For any connected graph G', it is well-known that $|V(G')| \leqslant |E(G')| + 1$ with equality if and only if G' is a tree. The crucial observation here is that since $T_1, \dots, T_b \in \mathcal{C}_i(r)$, the graph G is not a tree because there is at least one label that is repeated in some path belonging to T_1 (and similarly to $T_2, ..., T_b$). Therefore $|V(G)| \leqslant r/2$. It remains to apply Lemma 15 to obtain the first bound.

We now turn to the second bound. In a directed and labelled graph,

- A backtracking path of length 3 is a path $a \to b \to c \to d$ such that $\ell(a) = \ell(c)$ and $\ell(b) = \ell(d)$.
- A backtracking star is a structure $a, b \to c \to d$ where $\ell(a) = \ell(b) = \ell(d)$.

In [8, Lemma 3], it is shown that

$$z_i^k = \breve{z}_i^k + \sum_{T \in \mathcal{B}_i^k} W(T) \widetilde{\Gamma}(T) \breve{z}(T)$$

where $\widetilde{\Gamma}(T)$ is bounded and where $\mathcal{B}_i^k$ is a certain subset of $\overline{\mathcal{T}}_i^k$ such that each $T \in \mathcal{B}_i^k$ contains at least one backtracking path of length 3 or a backtracking star.

With this at hand, we have

$$\mathbb{E}(z_i^k - \breve{z}_i^k)^b \leqslant C \sum_{T_1, \dots, T_b \in \mathcal{B}_i^k} \left| \mathbb{E} W(T_1) \dots W(T_b) \right|.$$

Given an integer $r \in [C_b]$, we define the set $\mathcal{D}_i(r)$ similarly to $\mathcal{C}_i(r)$ above except for the fact that $T_1, \ldots, T_b \in \mathcal{B}_i^k$ instead of $\mathcal{T}_i^k \backslash \breve{\mathcal{T}}_i^k$. Fixing r such that $\mathcal{D}_i(r) \neq \varnothing$, we have that $|\mathbb{E} W(T_1) \ldots W(T_b)| \leqslant C K_n^{-r/2}$ when $(T_1, \ldots, T_b) \in \mathcal{D}_i(r)$. Furthermore, for $G = \boldsymbol{G}(T_1, \ldots, T_b)$, we recall that $|E(G)| = \sum_{j<l} \mathbb{1}_{\phi(T_1)_{jl} + \cdots + \phi(T_b)_{jl} \geqslant 2}$. Due to the presence of a backtracking path of length 3 or a backtracking star in each of the the trees $T_1, \ldots, T_b$, we observe that

$$2|E(G)| + 2 \leqslant \sum_{j<l} \phi(T_1)_{jl} + \cdots + \phi(T_b)_{jl} = r.$$

Since G is connected, $|V(G)| \leqslant |E(G)| + 1 \leqslant r/2$. It remains to apply Lemma 15 again. $\square$

Our last approximation result relates the iterates z^k with the x^k:

Lemma 17. Let $C_W > 0$ be a constant, and define the probability event $\mathcal{E} = [\|W\| \leqslant C_W]$. For each $\varepsilon > 0$, there is a polynomial f_ε such that the iterates z^l obtained with $f = f_\varepsilon$ satisfy

$$\limsup_n \mathbb{E}\left[\left\|z^k - x^k\right\|_n^2 \mathbb{1}_{\mathcal{E}}\right] \leqslant \varepsilon.$$

This polynomial can be chosen in such a way that $f_\varepsilon(0) = \operatorname{Tanh}(0)$ and

$$\max_{\alpha \in [0, \sqrt{2}]} \mathbb{E}\left(f_\varepsilon(\alpha\xi) - \operatorname{Tanh}(\alpha\xi)\right)^2 \leqslant c \quad \text{and} \quad \max_{\alpha \in [0, \sqrt{2}]} \mathbb{E}\left(f_\varepsilon'(\alpha\xi) - \operatorname{Tanh}'(\alpha\xi)\right)^2 \leqslant c.$$

for some $c > 0$ that depends on ε only.

The proof is close to [18, end of the proof of Theorem 2].

Proof. Given a small $c > 0$, Lemma 9 shows that there exists a polynomial p_c such that $p_c(0) = \operatorname{Tanh}(0)$,

$$\max_{\alpha \in [0, \sqrt{2}]} \mathbb{E}\left(p_c(\alpha\xi) - \operatorname{Tanh}(\alpha\xi)\right)^2 \leqslant c, \quad \text{and} \quad \max_{\alpha \in [0, \sqrt{2}]} \mathbb{E}\left(p_c'(\alpha\xi) - \operatorname{Tanh}'(\alpha\xi)\right)^2 \leqslant c.$$

In the proof, $\delta(c)$ will denote a generic function defined near zero in $\mathbb{R}_+$ such that $\delta(c) \to 0$ when $c \to 0$. Constructing the z^l's with $f = p_c$, we shall show by recurrence on l that for each $l \in [k]$,

$$\limsup_n \mathbb{E}\left\|z^l - x^l\right\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant \delta(c) \quad \text{and} \quad \limsup_n \mathbb{E}\left\|p_c(z^l) - \operatorname{Tanh}(x^l)\right\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant \delta(c),$$

where the function δ can change from an iteration to another. At Iteration k, it will be enough to choose c such that $\delta(c) \leqslant \varepsilon$ and to set $f_\varepsilon = p_c$ to obtain the result of the lemma.

Starting with $l = 1$, we have $z^1 = x^1 = W f(0) = W \operatorname{Tanh}(0)$. Similarly to the beginning of the proof of Lemma 14, we also have $\mathbb{E}\left\|p_c(z^1) - \operatorname{Tanh}(x^1)\right\|_n^2 \leqslant c$.

Assume now that the recurrence assumption is true for l. Recall the expressions (16) and (4) of the z^l's and the x^l's respectively. Our first task is to show that

$$\limsup_n \mathbb{E}\left\|\operatorname{diag}((W \odot W) p_c'(z^l)) p_c(z^{l-1}) - \operatorname{diag}(tS\mathbb{E}\operatorname{Tanh}'(X^l)) \operatorname{Tanh}(x^{l-1})\right\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant \delta(c). \quad (25)$$

By making use of the bound (24) along with Cauchy-Schwarz, we get

$$\begin{aligned}\mathbb{E}\left\|\operatorname{diag}((W \odot W - tS) p_c'(z^l)) p_c(z^{l-1})\right\|_n^2 &= \frac{1}{n} \sum_i \mathbb{E}\left[(W \odot W - tS) p_c'(z^l)\right]_i^2 p_c(z_i^{l-1})^2 \\ &\leqslant \frac{C}{n} \sum_i \left(\mathbb{E}\left[(W \odot W - tS) p_c'(z^l)\right]_i^4\right)^{1/2}\end{aligned}$$

which converges to zero by [18, Lemma 21]. We now write

$$
\begin{aligned}
&\left\|\operatorname{diag}(tSp_c'(z^l))p_c(z^{l-1}) - \operatorname{diag}(tS\mathbb{E}\operatorname{Tanh}'(X^l))\operatorname{Tanh}(x^{l-1})\right\|_n \\
&\leqslant \left\|\operatorname{diag}(tS(p_c'(z^l) - \mathbb{E}p_c'(\check{X}^l))p_c(z^{l-1})\right\|_n + \left\|\operatorname{diag}(tS\mathbb{E}p_c'(\check{X}^l))(p_c(z^{l-1}) - \operatorname{Tanh}(x^{l-1})\right\|_n \\
&\quad + \left\|\operatorname{diag}(tS(\mathbb{E}p_c'(\check{X}^l) - \mathbb{E}\operatorname{Tanh}'(X^l))\operatorname{Tanh}(x^{l-1})\right\|_n .
\end{aligned}
$$

By Cauchy-Schwarz and the bound (24), we have

$$
\mathbb{E}\left\|\operatorname{diag}(tS(p_c'(z^l) - \mathbb{E}p_c'(\check{X}^l))p_c(z^{l-1})\right\|_n^2 \leqslant \frac{C}{n}\sum_i \left(\mathbb{E}\left(\sum_r s_{ir}\left(p_c'(z_r^l) - \mathbb{E}p_c'(\check{X}_r^l)\right)\right)^4\right)^{1/2}
$$

which converges to zero by Propositions 13 and 16.

We also have that

$$
|\mathbb{E}p_c'(\check{X}^l)| \leqslant |\mathbb{E}p_c'(\check{X}^l) - \mathbb{E}\operatorname{Tanh}'(\check{X}_l)| + |\mathbb{E}\operatorname{Tanh}'(\check{X}_l)| \leqslant 1 + \sqrt{c},
$$

therefore, $\|\operatorname{diag}(tS\mathbb{E}p_c'(\check{X}^l))\|$ is bounded, and thus,

$$
\mathbb{E}\left\|\operatorname{diag}(tS\mathbb{E}p_c'(\check{X}^l))(p_c(z^{l-1}) - \operatorname{Tanh}(x^{l-1}))\right\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant C\mathbb{E}\left\|p_c(z^{l-1}) - \operatorname{Tanh}(x^{l-1})\right\|_n^2 \mathbb{1}_{\mathcal{E}}
$$

which $\limsup$ is bounded by $\delta(c)$ by the recurrence assumption.

We finally have that

$$
|\mathbb{E}p_c'(\check{X}_i^l) - \mathbb{E}\operatorname{Tanh}'(X_i^l)| \leqslant |\mathbb{E}p_c'(\check{X}_i^l) - \mathbb{E}\operatorname{Tanh}'(\check{X}_i^l)| + |\mathbb{E}\operatorname{Tanh}'(\check{X}_i^l) - \mathbb{E}\operatorname{Tanh}'(X_i^l)| \leqslant \delta(c)
$$

by the construction of p_c' and by Lemma 10 that shows that $\|\check{q}^l - q^l\|_\infty$ is small. With this, we obtain that

$$
\mathbb{E}\left\|\operatorname{diag}(tS(\mathbb{E}p_c'(\check{X}^l) - \mathbb{E}\operatorname{Tanh}'(X^l))\operatorname{Tanh}(x^{l-1})\right\|_n^2 \leqslant \delta(c),
$$

and (25) follows.

With this, we have

$$
\begin{aligned}
\|z^{l+1} - x^{l+1}\|_n \mathbb{1}_{\mathcal{E}} &\leqslant C_W\|z^l - x^l\|_n \mathbb{1}_{\mathcal{E}} \\
&\quad + \left\|\operatorname{diag}((W\odot W)p_c'(z^l))p_c(z^{l-1}) - \operatorname{diag}(tS\mathbb{E}\operatorname{Tanh}'(X^l))\operatorname{Tanh}(x^{l-1})\right\|_n \mathbb{1}_{\mathcal{E}}
\end{aligned}
$$

and we get from the recurrence assumption and from (25) that $\mathbb{E}\|z^{l+1} - x^{l+1}\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant \delta(c)$. Also, writing

$$
\begin{aligned}
\|p_c(z^{l+1}) - \operatorname{Tanh}(x^{l+1})\|_n \mathbb{1}_{\mathcal{E}} &\leqslant \|p_c(z^{l+1}) - \operatorname{Tanh}(z^{l+1})\|_n + \|z^{l+1} - x^{l+1}\|_n \mathbb{1}_{\mathcal{E}} \\
&\leqslant \|p_c(z^{l+1}) - \mathbb{E}p_c(\check{X}^{l+1})\|_n + \|\operatorname{Tanh}(z^{l+1}) - \mathbb{E}\operatorname{Tanh}(\check{X}^{l+1})\|_n \\
&\quad + \|\mathbb{E}p_c(\check{X}^{l+1}) - \mathbb{E}\operatorname{Tanh}(\check{X}^{l+1})\|_n + \|z^{l+1} - x^{l+1}\|_n \mathbb{1}_{\mathcal{E}},
\end{aligned}
$$

we see that $\limsup_n \mathbb{E}\|p_c(z^{l+1}) - \mathbb{E}p_c(\check{X}^{l+1})\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant \delta(c)$, and the recurrence assumption is verified for $l+1$. □

2.3.1 Theorem 5: end of proof

Given $\varepsilon > 0$, we know that we can choose $c > 0$ small enough in the statement of Lemma 17 so that the conclusions of this lemma are satisfied. If we make c smaller if necessary, then the conclusion of Lemma 14 will be true for this same polynomial f_ε. In this situation, combining Lemma 17, Proposition 16 with $b = 2$, and Lemma 14, we obtain that

$$
\limsup_n \mathbb{E}\left\|\tilde{x}^k - x^k\right\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant 2\varepsilon,
$$

which implies that

$$\limsup_n \mathbb{E}\left\|\mathrm{Tanh}(\tilde{x}^k) - \mathrm{Tanh}(x^k)\right\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant 2\varepsilon.$$

Using Lemmas 12 and 11, we obtain that

$$\limsup_n \mathbb{E}\left\|m - \mathrm{Tanh}(x^k)\right\|_n^2 \mathbb{1}_{\mathcal{E}} \leqslant 2\varepsilon + 4(\log 2)^k.$$

Denoting as $\|M\|_\infty$ the max norm of the matrix M, we know from [4, Th. 1.1] that

$$\mathbb{E}\|W^{(n)}\| \leqslant T^{(n)},$$

where

$$T^{(n)} = (1+\delta)\left(2\left|\left|\left|S^{(n)}\right|\right|\right|^{1/2} + \frac{6}{\sqrt{\log(1+\delta)}}(\|S^{(n)}\|_\infty \log n)^{1/2}\right)$$

for an arbitrary $\delta > 0$. Furthermore, by Gaussian concentration,

$$\mathbb{P}\left[\|W^{(n)}\| \geqslant T^{(n)} + t\right] \leqslant \exp(-t^2/(2\|S^{(n)}\|_\infty)^2),$$

for $\delta \in (0, 1/2]$, as given by [4, Cor. 3.9]. By consequence, since $K_n \geqslant \log n$, it holds that there exists $C_W > 0$ for which $\mathbb{1}_{\mathcal{E}} \to_n 1$ almost surely. This implies that

$$\limsup_n \mathbb{E}\left\|m - \mathrm{Tanh}(x^k)\right\|_n^2 \leqslant 2\varepsilon + 4(\log 2)^k,$$

thus, $\limsup_k \limsup_n \mathbb{E}\left\|m - \mathrm{Tanh}(x^k)\right\|_n^2 \leqslant 2\varepsilon$. Since ε is arbitrary, Theorem 5 holds true.

We close the paper with some remarks.

Remark 1. The proof for the bound (12) requires the Gaussian assumption on the entries of the matrix W. This assumption is not essential for the rest of the proof which is based on the approach of [8, 18]. If we manage to generalize the bound (12) to the non-necessarily Gaussian case, the proof above will continue to work after some easy adaptations.

Remark 2. The condition $K_n \geqslant \log n$ is an artifact of our proof due to the fact that we needed to bound the spectral norm $\|W\|$ in order to approximate our AMP algorithm with a polynomial activation function with the AMP algorithm with the Tanh activation function. Another possible proof technique would be to pass from the iterates $\tilde{x}^l$ to the iterates x^l without the need of introducing the polynomial intermediates. This is left for future research.

Acknowledgement

I would like to thank Christian Brennecke for a useful discussion.

A Appendices

A.1 Proof of Lemma 1

By a derivative calculation and the use of the Gaussian integration by parts formula (see, *e.g.* the proof of [25, Proposition 1.3.8]), we can show that g is 1–Lipschitz. By Assumptions 1 and 2, it holds that $\sup_n \left|\left|\left|tS^{(n)}\right|\right|\right| < 1$. Thus, for each $n > 0$, the function $q \in \mathbb{R}_+^n \mapsto tS^{(n)}g(q)$ is a contraction for the $\|\cdot\|_\infty$ norm on $\mathbb{R}_+^n$, and the result follows by Banach's fixed point theorem. Since $0 \leqslant g(q) < 1$, we also have $\|q^{(n)}\|_\infty \leqslant \left|\left|\left|tS^{(n)}\right|\right|\right| \|g(q^{(n)})\|_\infty < \log 2$ by Assumption 2.

A.2 Proof of Corollary 4

We take out the superscripts $^{(n)}$. Recall that $S = [\psi(|i-j|)]$. For n large enough, let S_{circ} be the circulant deformation of S given as the $n \times n$ circulant symmetric matrix which first row is

$$\begin{bmatrix}\psi(0) & \cdots & \psi(K_n) & 0 & \cdots & 0 & \psi(K_n) & \cdots & \psi(1)\end{bmatrix}.$$

Given a $n \times n$ matrix X taken from the Gaussian Orthogonal Ensemble, let $W = (tS)^{\odot 1/2} \odot X$, and $\widetilde{W} = [\widetilde{W}_{ij}] = (tS_{\text{circ}})^{\odot 1/2} \odot X$ where $[A_{ij}]^{\odot 1/2} = [A_{ij}^{1/2}]$. Recall that $H(\sigma) = \sigma^\top W \sigma/2 + h\,(\sigma \cdot 1)$ with the free energy F_n. Define the Hamiltonian $\widetilde{H}(\sigma) = \sigma^\top \widetilde{W}\sigma/2 + h\,(\sigma \cdot 1)$, and let $\widetilde{F}_n$ be the associated free energy. Knowing from Corollary 3 that the convergence (3) holds true for $\widetilde{F}_n$, all we need to prove is that $F_n - \widetilde{F}_n \to 0$. We can write $\widetilde{H}(\sigma) = H(\sigma) + E(\sigma)$ where $E(\sigma)$ is given as

$$E(\sigma) = \sum_{i=1}^{K_n} \sum_{j=n-K_n+i}^{n} \sigma_i \sigma_j \widetilde{W}_{ij}.$$

Note that $E(\sigma)$ and $H(\sigma)$ are independent and that the $\widetilde{W}_{ij}$'s in the expression of $E(\sigma)$ above satisfy $\mathbb{E}\widetilde{W}_{ij}^2 = t\psi(|i-j+n|)$. By Jensen's inequality, $\mathbb{E}\log(\sum e^{H(\sigma)+E(\sigma)}/\sum e^{H(\sigma)}) = \mathbb{E}\log\left\langle e^{E(\sigma)}\right\rangle \geqslant \mathbb{E}\left\langle E(\sigma)\right\rangle = 0$, thus, $\widetilde{F}_n \geqslant F_n$. We also have that

$$\mathbb{E}E(\sigma)^2 = t\sum_{i=1}^{K_n} \sum_{j=n-K_n+i}^{n} \psi(|i-j+n|) \leqslant tK_n.$$

Jensen's inequality applied to the expectation with respect to the law of $E(\sigma)$ leads to

$$\widetilde{F}_n \leqslant \frac{1}{n}\mathbb{E}\log\sum_{\sigma} e^{H(\sigma)+\mathbb{E}E(\sigma)^2/2} \leqslant \frac{tK_n}{2n} + F_n,$$

and the corollary is established.

A.3 Proof of Lemma 6

Using Identity (6), it is enough to bound $\mathbb{E}(\delta_i m_j^{[i]})^2$. In the derivations below, we use Itô's lemma to obtain the equality. To obtain the first inequality, we extract the terms $k = j$ from the two sums at the right hand side of the equality, we observe that $m_{jj}^{[i]} \in [0,1]$, and we use the inequality $ab \leqslant (a^2+b^2)/2$. We obtain

$$\begin{aligned}
&\mathbb{E}(\delta_i m_j^{[i]})^2 \\
&= -2\sum_{k\neq i} s_{ik} \int_0^{tu} \mathbb{E}\delta_i m_j^{[i]}(v)\delta_i\left(m_k^{[i]} m_{kj}^{[i]}\right)(v)dv + \sum_{k\neq i} s_{ik}\int_0^{tu} \mathbb{E}(\varepsilon_i m_{kj}^{[i]}(v))^2 dv \\
&\leqslant 3s_{ij}t + \sum_{k\neq i,j} s_{ik}\int_0^t \left(\mathbb{E}(\varepsilon_i m_{kj}^{[i]}(v))^2 + \mathbb{E}(\delta_i\left(m_k^{[i]} m_{kj}^{[i]}\right)(v))^2\right)dv + \left(\sum_{k\neq i,j} s_{ik}\right)\int_0^t \mathbb{E}(\delta_i m_j^{[i]}(v))^2 dv \\
&\leqslant 3s_{ij}t + \boldsymbol{C}_{\text{row}} \max_{k\neq i,j} \int_0^t \left(\mathbb{E}(\varepsilon_i m_{kj}^{[i]}(v))^2 + \mathbb{E}(\delta_i\left(m_k^{[i]} m_{kj}^{[i]}\right)(v))^2\right)dv + \boldsymbol{C}_{\text{row}}\int_0^t \mathbb{E}(\delta_i m_j^{[i]}(v))^2 dv \\
&\leqslant 3s_{ij}t + \boldsymbol{C}_{\text{row}} \max_{\substack{k\neq i,j\\ \sigma_i=\pm 1}} \int_0^t \mathbb{E}(1+m_k^{[i]}(v)^2)m_{kj}^{[i]}(v)^2 dv + \boldsymbol{C}_{\text{row}}\int_0^t \mathbb{E}(\delta_i m_j^{[i]}(v))^2 dv.
\end{aligned}$$

In the last inequality, we used that

$$\begin{aligned}
\left(\frac{x(1)+x(-1)}{2}\right)^2 + \left(\frac{y(1)x(1)-y(-1)x(-1)}{2}\right)^2 &\leqslant \frac{x(1)^2+x(-1)^2+y(1)^2x(1)^2+y(-1)^2x(-1)^2}{2} \\
&\leqslant \max_{i=\pm 1} x(i)^2(1+y(i)^2).
\end{aligned}$$

We now use that $m^{[i]}_{kj} = (1-(m^{[i]}_k)^2)\delta_k m^{[i,k]}_j$ to obtain that

$$\mathbb{E}(\delta_i m^{[i]}_j)^2 \leqslant 3 s_{ij} t + \boldsymbol{C}_{\text{row}} \max_{\substack{k\neq i,j\\ \sigma_i=\pm 1}} \int_0^t \mathbb{E}(\delta_k m^{[i,k]}_j(v))^2 dv + \boldsymbol{C}_{\text{row}} \int_0^t \mathbb{E}(\delta_i m^{[i]}_j(v))^2 dv$$

Using Gronwäll's inequality which states that if $\varphi(t) \leqslant \alpha(t) + C\int_0^t \varphi(v)dv$, then $\varphi(t) \leqslant \alpha(t) + C\int_0^t \alpha(v)\exp(C(t-v))dv$, we obtain by making an Integration by Parts that

$$\mathbb{E}(\delta_i m^{[i]}_j)^2 \leqslant 6 s_{ij} t + \boldsymbol{C}_{\text{row}} \max_{\substack{k_1\neq i,j\\ \sigma_i=\pm 1}} \int_0^t e^{\boldsymbol{C}_{\text{row}}(t-v_1)} \mathbb{E}(\delta_{k_1} m^{[i,k_1]}_j(v_1))^2 dv_1.$$

We now similarly consider $\delta_{k_1} m^{[i,k_1]}_j(v_1)$, and focus on the dependence of this random variable on $(W_{k_1k_2}(t))_{k_2\neq i,k_1}$. Applying the same argument as above to $\mathbb{E}(\delta_{k_1} m^{[i,k_1]}_j(v_1))^2$, we obtain

$$\mathbb{E}(\delta_{k_1} m^{[i,k_1]}_j(v_1))^2 \leqslant 6 s_{k_1 j} t + \boldsymbol{C}_{\text{row}} \max_{\substack{k_2\neq i,j,k_1\\ \sigma_{k_1}=\pm 1}} \int_0^t e^{\boldsymbol{C}_{\text{row}}(t-v_2)} \mathbb{E}(\delta_{k_2} m^{[i,k_1,k_2]}_j(v_1,v_2))^2 dv_2,$$

and thus,

$$\begin{aligned}\mathbb{E}(\delta_i m^{[i]}_j)^2 \leqslant{}& \frac{6\boldsymbol{C}_s}{K_n} t\left(1 + (e^{\boldsymbol{C}_{\text{row}} t} - 1)\right)\\ &+ \boldsymbol{C}_{\text{row}}^2 \max_{\substack{k_1\neq i,j\\ \sigma_i=\pm 1}} \max_{\substack{k_2\neq i,j,k_1\\ \sigma_{k_1}=\pm 1}} \int_0^t\int_0^t e^{\boldsymbol{C}_{\text{row}}(2t-v_1-v_2)} \mathbb{E}(\delta_{k_2} m^{[i,k_1,k_2]}_j(v_1,v_2))^2 dv_1 dv_2.\end{aligned}$$

Iterating, we end up with

$$\begin{aligned}\mathbb{E}(\delta_i m^{[i]}_j)^2 \leqslant{}& \frac{6\boldsymbol{C}_s}{K_n} t\left(1 + (e^{\boldsymbol{C}_{\text{row}} t} - 1) + \cdots + (e^{\boldsymbol{C}_{\text{row}} t} - 1)^{n-3}\right)\\ &+ \boldsymbol{C}_{\text{row}}^{n-2} \max_{\substack{k_1\neq i,j\\ \sigma_i=\pm 1}} \cdots \max_{\substack{k_{n-2}\neq i,j,k_1,\ldots,k_{n-3}\\ \sigma_{k_{n-2}}=\pm 1}} \int_0^t \cdots \int_0^t e^{\boldsymbol{C}_{\text{row}}\sum_{\ell=1}^{n-2}(t-v_\ell)} \times\\ &\qquad \mathbb{E}(\delta_{k_{n-2}} m^{[i,k_1,\ldots,k_{n-2}]}_j(v_1,v_2,\ldots,v_{n-2}))^2 dv_1\ldots dv_{n-2},\end{aligned}$$

which leads to the result.

A.4 Proof of Lemma 15

Let us denote as $\mathcal{G}_i$ the set of rooted, undirected, labelled and connected graphs such that $\ell(\circ) = i$, $|V(G)| \leqslant r$, and the property

$$\{u,v\} \in E(G) \ \Rightarrow\ \ell(u) \in \mathcal{K}_{\ell(v)}.$$

We denote as $\mathcal{R}$ the set of all the elements of $\mathcal{G}_i$ but without the labels. Given a graph $G \in \mathcal{G}_i$, let us denote as $\bar{G} = \boldsymbol{U}(G) \in \mathcal{R}$ the unlabelled version of G. With these notations, we have

$$|\mathcal{A}| = \sum_{\bar{G}\in\mathcal{R}} \sum_{\substack{G\in\mathcal{G}_i:\\ \boldsymbol{U}(G)=\bar{G}}} \left|\{(T_1,\ldots,T_b) \in \mathcal{A} \,:\, \boldsymbol{G}(T_1,\ldots,T_b) = G\}\right|. \tag{26}$$

The summand in this expression is bounded by a constant independent of G. We need to show that

$$\left|\left\{G \in \mathcal{G}_i \,:\, \boldsymbol{U}(G) = \bar{G}\right\}\right| \leqslant C K_n^{r-1}. \tag{27}$$

Given $\bar{G} \in \mathcal{R}$, denote as $\circ$ the root node of $\bar{G}$, write $M = |V(\bar{G})| - 1$, and write $V(\bar{G}) \setminus \{\circ\} = [M]$. Recalling that $\bar{G}$ is connected, let us consider a spanning tree of this graph rooted in $\circ$. Denote as $\boldsymbol{\pi}(v)$ the parent of the node v in this tree. Writing $j_\circ = i$, we obtain that

$$\left|\left\{G \in \mathcal{G}_i \,:\, \boldsymbol{U}(G) = \bar{G}\right\}\right| \leqslant \left|\left\{(j_1, \ldots, j_M) \in [n]^M \,:\, \forall k \in [M],\ j_k \in \mathcal{K}_{j_{\boldsymbol{\pi}(k)}}\right\}\right|.$$

Denoting as $L \subset [M]$ the set of the leaves of the spanning tree, we can write

$$\begin{aligned}
\left|\left\{(j_1, \ldots, j_M) \in [n]^M \,:\, \forall k \in [M],\ j_k \in \mathcal{K}_{j_{\boldsymbol{\pi}(k)}}\right\}\right| &= \sum_{\substack{j_1, \ldots, j_M \in [n]\,: \\ \forall k \in [M],\ j_k \in \mathcal{K}_{j_{\boldsymbol{\pi}(k)}}}} 1 \\
&= \sum_{k \in [M] \setminus L} \sum_{j_k \in \mathcal{K}_{j_{\boldsymbol{\pi}(k)}}} \Big(\sum_{p \in L} \sum_{j_p \in \mathcal{K}_{j_{\boldsymbol{\pi}(p)}}} 1 \Big) \\
&\leqslant C K_n^{|L|} \sum_{k \in [M] \setminus L} \sum_{j_k \in \mathcal{K}_{j_{\boldsymbol{\pi}(k)}}} 1,
\end{aligned}$$

recalling that $|\mathcal{K}_j| \leqslant C K_n$ for all j and using the inequality $|L| K_n \leqslant K_n^{|L|}$ as soon as $K_n \geqslant 2$. If we prune the leaves of the original spanning tree, what remains is a tree made of the nodes that constitute the first sum above plus the root node. We can apply the pruning operation to the new tree as above, and iterate until exhausting all the set $[M] = V(\bar{G}) \setminus \{\circ\}$. This leads to

$$\left|\left\{(j_1, \ldots, j_M) \in [n]^M \,:\, \forall k \in [M],\ j_k \in \mathcal{K}_{j_{\boldsymbol{\pi}(k)}}\right\}\right| \leqslant C K_n^M \leqslant C K_n^{r-1},$$

hence Inequality (27).

It is furthermore easy to check that

$$|\mathcal{R}| \leqslant C,$$

and the lemma is proven.

References

[1] A. Adhikari, C. Brennecke, P. von Soosten, and H.-T. Yau. Dynamical approach to the TAP equations for the Sherrington-Kirkpatrick model. *J. Stat. Phys.*, 183(3):Paper No. 35, 27, 2021.

[2] D. Alberici, F. Camilli, P. Contucci, and E. Mingione. The multi-species mean-field spin-glass on the Nishimori line. *arXiv preprint arXiv:2007.08891*, 2020.

[3] D. Alberici, P. Contucci, E. Mingione, and F. Zimmaro. Ferromagnetic Ising model on multiregular random graphs. *arXiv preprint arXiv:2403.14307*, 2024.

[4] A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479 – 2506, 2016.

[5] A. Barra, P. Contucci, E. Mingione, and D. Tantari. Multi-species mean field spin glasses. Rigorous results. *Ann. Henri Poincaré*, 16(3):691–708, 2015.

[6] E. Bates, L. Sloman, and Y. Sohn. Replica symmetry breaking in multi-species Sherrington-Kirkpatrick model. *J. Stat. Phys.*, 174(2):333–350, 2019.

[7] E. Bates and Y. Sohn. Balanced multi-species spin glasses. *arXiv preprint arXiv:2507.06522*, 2025.

[8] M. Bayati, M. Lelarge, and A. Montanari. Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.*, 25(2):753–822, 2015.

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

[10] H.-B. Chen. On free energy of non-convex multi-species spin glasses. *arXiv preprint arXiv:2411.13342*, 2024.

[11] H.-B. Chen, V. Issa, and J.-C. Mourrat. The convex structure of the Parisi formula for multi-species spin glasses. *arXiv preprint arXiv:2508.06397*, 2025.

[12] W.-K. Chen and S. Tang. On convergence of the cavity and Bolthausen's TAP iterations to the local magnetization. *Comm. Math. Phys.*, 386(2):1209–1242, 2021.

[13] F. Comets and J. Neveu. The Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case. *Comm. Math. Phys.*, 166(3):549–564, 1995.

[14] P.S. Dey and Q. Wu. Fluctuation results for multi-species Sherrington-Kirkpatrick model in the replica symmetric regime. *J. Stat. Phys.*, 185(3):Paper No. 22, 40, 2021.

[15] M. M. Džrbašyan and A. B. Tavadyan. On weighted uniform approximation by polynomials of functions of several variables. *Mat. Sb. (N.S.)*, 43(85):227–256, 1957.

[16] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth. A unifying tutorial on Approximate Message Passing. *Foundations and Trends in Machine Learning*, 15(4):335–536, 05 2022.

[17] A. Guionnet, J. Ko, F. Krzakala, and L. Zdeborová. Low-rank matrix estimation with inhomogeneous noise. *Information and Inference: A Journal of the IMA*, 14(2), 04 2025.

[18] W. Hachem. Approximate message passing for sparse matrices with application to the equilibria of large ecological Lotka-Volterra systems. *Stochastic Process. Appl.*, 170:Paper No. 104276, 34, 2024.

[19] A. Jagannath, J. Ko, and S. Sen. Max κ-cut and the inhomogeneous Potts spin glass. *Ann. Appl. Probab.*, 28(3):1536–1572, 2018.

[20] H. Kim. On the de Almeida–Thouless transition surface in the multi-species SK model with centered Gaussian external field. *arXiv preprint arXiv:2509.18066*, 2025.

[21] Q. Liu and Z. Dong. Some rigorous results for the diluted multi-species SK model. *J. Stat. Phys.*, 191(12):Paper No. 162, 25, 2024.

[22] H. N. Mhaskar. Weighted polynomial approximation. *J. Approx. Theory*, 46(1):100–110, 1986. Papers dedicated to the memory of Géza Freud.

[23] D. Panchenko. *The Sherrington-Kirkpatrick model.* Springer Monographs in Mathematics. Springer, New York, 2013.

[24] D. Panchenko. The free energy in a multi-species Sherrington-Kirkpatrick model. *Ann. Probab.*, 43(6):3494–3513, 2015.

[25] M. Talagrand. *Mean field models for spin glasses. Volume I*, volume 54 of *A Series of Modern Surveys in Mathematics*. Springer, 2011. Basic examples.

[26] M. Talagrand. *Mean field models for spin glasses. Volume II*, volume 55 of *A Series of Modern Surveys in Mathematics*. Springer, 2011. Advanced replica-symmetry and low temperature.

[27] Q. Wu. Thouless-Anderson-Palmer equations for the multi-species Sherrington-Kirkpatrick model. *J. Stat. Phys.*, 191(7):Paper No. 87, 14, 2024.