Journal of Theoretical Probability (2026) 39:19
https://doi.org/10.1007/s10959-025-01476-z

Approximate Message Passing for General Non-Symmetric Random Matrices

Mohammed-Younes Gueddari[1] · **Walid Hachem**[1] · **Jamal Najim**[1]

Received: 10 March 2025 / Revised: 24 November 2025 / Accepted: 16 December 2025

Abstract
Approximate message passing (AMP) algorithms are a family of iterative algorithms based on large random matrices with the special property of tracking the statistical properties of their iterates. They are used in various fields such as statistical physics, machine learning, communication systems, theoretical ecology, etc. In this article we consider AMP algorithms based on non-symmetric random matrices with a general variance profile, possibly sparse, a general covariance profile, and non-Gaussian entries. We hence substantially extend the results on elliptic random matrices that we developed in Gueddari et al. (Random Matrices: Theory Appl. 14, 2025). From a technical point of view, we enhance the combinatorial techniques developed in Bayati et al. (Ann. Appl. Prob. 25:753–822, 2015) and in Hachem (Stoch. Process. Appl. 170:104276, 2024). Our main motivation is the understanding of equilibria of large food-webs described by Lotka–Volterra systems of ordinary differential equations, continuing the work of Hachem (Stoch. Process. Appl. 170:104276, 2024), Akjouj et al. (J. Math. Biol. 89:61, 2024) and Gueddari et al. (Random Matrices: Theory Appl. 14, 2025), but the versatility of the model studied might be of interest beyond these particular applications.

Keywords Approximate message passing · non-hermitian random matrices · variance and covariance prole · sparse random matrix.

MSC code: primary 60B20 · secondary 60F05 · 60K35

✉ Jamal Najim
jamal.najim@univ-eiffel.fr

Mohammed-Younes Gueddari
gueddari.mohammedyounes@gmail.com

Walid Hachem
walid.hachem@univ-eiffel.fr

[1] CNRS, Laboratoire d'informatique Gaspard Monge (LIGM / UMR 8049), Université Gustave Eiffel, Champs-sur-Marne, France

Published online: 14 January 2026

1 Introduction

Approximate message passing (AMP) refers to a class of iterative algorithms that are built around a large random matrix, producing at each step a high-dimensional $\mathbb{R}^n$-valued random vector ($n \gg 1$) whose elements' empirical distribution can be identified as n goes to infinity. These algorithms take the following form

$$\boldsymbol{x}^{t+1} = W h_t(\boldsymbol{x}^t) - \{\text{corrective term}\},$$

where $\boldsymbol{x}^t = (x_i^t)$ is the $n \times 1$ vector at iteration t, W is a $n \times n$ random matrix, and $h_t(\boldsymbol{x}^t) = (h_t(x_i^t))_i$ is a vector based on the so-called *activation function* $h_t : \mathbb{R} \to \mathbb{R}$. The corrective term, known as the Onsager term, is carefully defined to facilitate the description of the statistical properties of $\boldsymbol{x}^t$ as $n \to \infty$.

In the fields of machine learning and statistical estimation, AMP algorithms were originally developed for studying compressed sensing and sparse signal recovery problems [8, 17]. They have since found applications across various fields, including high-dimensional estimation [15, 25], communication theory [6, 29], statistical physics [26], theoretical ecology [2, 19, 21], etc. AMP algorithms have undergone extensive recent developments and the goal of this article is to extend the AMP framework to general non-symmetric random matrices W.

In general, the random matrix model W may differ depending on the considered application, and most of AMP algorithms focus on symmetric matrices. For instance, in the problem of low-rank information extraction from noisy data matrix, the goal is to estimate the $n \times 1$ signal $\boldsymbol{x}^\star$ from noisy observations

$$Y = \sqrt{\lambda}\boldsymbol{x}^\star(\boldsymbol{x}^\star)^\top + W\,, \tag{1}$$

where W is a random matrix. In [16] and [27], the authors develop an AMP algorithm involving a symmetric matrix $W = \frac{1}{\sqrt{n}}G$ where G is drawn from the Gaussian Orthogonal Ensemble (GOE(n)) to study the problem (1). More precisely, each entry $G_{ij} \sim \mathcal{N}(0, 1 + 1_{(i=j)})$, where $1_{(i=j)}$ equals one if $i = j$ and zero else, and all the entries on and above the diagonal are independent. The $1/\sqrt{n}$ normalization factor is standard in Random Matrix Theory and has the effect to ensure that the spectral norm of W is $\mathcal{O}(1)$.

In [9, 20, 23, 28], the authors develop an AMP algorithm involving a symmetric random matrix W with a block-wise variance profile S to study the problem (1) in the case of an inhomogeneous noise. More precisely, W is now written as

$$W = \frac{1}{\sqrt{n}}\, S^{\odot 1/2} \odot G\,, \tag{2}$$

where $G \sim \text{GOE}(n)$ and S is a symmetric, deterministic, block-constant matrix of non-negative elements. Matrix S has a finite number of rectangular blocks which dimensions scale with n, the elements of $S^{\odot 1/2}$ are the square roots of those of S, and $\odot$ is the Hadamard or entry-wise product. In the recent paper [5], Bao et al. consider

an AMP algorithm based on Gaussian matrices with a variance profile and provide non-asymptotic results.

Our main motivation to develop AMP algorithms associated to new matrix models comes from theoretical ecology and the study of large Lotka–Volterra systems of ODEs. In such models, the random matrix W is used to model the interactions between n living species that coexist within an ecosystem, and the time evolution of the abundances is described by the multi-dimensional Lotka–Volterra differential equation. In [2], Akjouj *et al.* consider the GOE model for the matrix of interactions, and use an AMP approach to describe the statistical properties of the equilibrium point of the resulting Lotka–Volterra dynamical system when this equilibrium is globally stable. Dealing with a more realistic interaction matrix model, [21] considers a symmetric random matrix with a variance profile as in (2), with the main difference that the variance profile matrix S can be sparse. Including correlations between the elements of the interaction matrix is an important feature in theoretical ecology. In this direction, a non-symmetric elliptic matrix W is considered in [19], where each entry pair $(\sqrt{n}W_{ij}, \sqrt{n}W_{ji})$ is a standard two-dimensional centered Gaussian vector with a covariance $\rho \in [-1, 1]$, and where all the different pairs are independent.

All these cases are particular cases of the model we study in this article.

1.1 The Random Matrix Model

The model under investigation here combines an arbitrary variance profile, possibly sparse, with a correlation profile. To this end, we first introduce the notion of a T-correlated matrix. Let $[n] = \{1, \cdots, n\}$.

Definition 1.1 Let $T = (\tau_{ij})_{1\le i,j\le n}$ be a symmetric $n \times n$ matrix with entries in $[-1, 1]$. The $n \times n$ random matrix X is T-correlated if

- Every entry X_{ij} is centered random variable with variance 1.
- For $(i, j) \in [n]^2$, $i < j$, the covariance matrix of the pair (X_{ij}, X_{ji}) is

$$\begin{pmatrix} 1 & \tau_{ij} \\ \tau_{ji} & 1 \end{pmatrix}.$$

- The random elements in the set $\{X_{ii}, (X_{ij}, X_{ji}),\ (i, j) \in [n]^2,\ i < j\}$ are independent.

Remark 1.2 Notice that the diagonal elements of T are not specified in this definition. A natural convention could be to set $\tau_{ii} = 1$, as it represents the correlation of X_{ii} with itself, but their exact values (as long as it is bounded) have no impact on the presented results.

Let X be a $\mathbb{R}^{n\times n}$–valued T-correlated matrix and $S = (s_{ij})_{i,j\in[n]}$ be a deterministic $n \times n$ matrix with non-negative elements. The random matrix model considered in this paper is

$$W = S^{\odot 1/2} \odot X = \left(\sqrt{s_{ij}}X_{ij}\right)_{1\le i,j\le n}. \tag{3}$$

Notice that the entries need not to be Gaussian and contrary to (2), the normalization is embedded into matrix S. We refer to S as the *variance profile* of matrix W and to T as its *correlation profile*. Such a model is fairly general as it encompasses most of the classical random matrix models (Wigner, Elliptic, Circular models) and many important features required in the applications (sparsity, variance profile, etc.).

1.2 A Primer to Approximate Message Passing

For a random matrix W such that $\sqrt{n}W \sim \mathrm{GOE}(n)$, an AMP algorithm starting at $\boldsymbol{x}^0 = (x_0, \cdots, x_0)^\top$ using a set of Lipschitz activation functions $(h_t)_{t\geq 0}$ is given by the following recursion equation; for all $t \geq 0$,

$$\boldsymbol{x}^{t+1} = W h_t(\boldsymbol{x}^t) - b_t h_{t-1}(\boldsymbol{x}^{t-1}) \quad \text{where} \quad b_t = \frac{1}{n}\sum_{i=1}^{n} h_t'(x_i^t)\,, \tag{4}$$

with the convention that $h_{-1} \equiv 0$.

The crucial term in this recursion is the Onsager term, i.e. "$\mathrm{ONS}_t := b_t h_{t-1}(\boldsymbol{x}^{t-1})$" that we subtract from the power method iteration term at each step t. The effect of the Onsager term is that for a fixed t and as $n \to \infty$, it "cancels" the dependence due to the repeated use of matrix W at each iteration:

$$\boldsymbol{x}^{t+1} = W h_t(W h_{t-1}(W \cdots) - \mathrm{ONS}_{t-1}) - \mathrm{ONS}_t\ .$$

With the correction of the Onsager term, the asymptotic behavior of $\boldsymbol{x}^t$ is similar to the behavior of $\tilde{\boldsymbol{x}}^t$ generated with the "power method iteration" but with a new sampled independent random matrix W^t at each step t, i.e.

$$\tilde{\boldsymbol{x}}^{t+1} = W^t h_t(\tilde{\boldsymbol{x}}^t) \qquad \text{with} \qquad \sqrt{n}W^t \overset{i.i.d.}{\sim} \mathrm{GOE}(n)\,.$$

Notice that in the latter case, it is easy to characterize the asymptotic behavior of the empirical distribution $\mu^{\tilde{\boldsymbol{x}}^t}$ of the entries of the vector $\tilde{\boldsymbol{x}}^t = (\tilde{x}_i^t)$,

$$\mu^{\tilde{\boldsymbol{x}}^t} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\tilde{x}_i^t}\,.$$

Roughly speaking $\mu^{\boldsymbol{x}^t} \approx \mu^{\tilde{\boldsymbol{x}}^t}$ as $n \to \infty$. Beware however that the correlation between consecutive iterations $\boldsymbol{x}^t$ and $\boldsymbol{x}^{t+1}$ differs from the correlation between iterates $\tilde{\boldsymbol{x}}^t$ and $\tilde{\boldsymbol{x}}^{t+1}$ which turn out to be asymptotically decorrelated.

Given the iterates $\boldsymbol{x}^1 = (x_i^1), \cdots, \boldsymbol{x}^t = (x_i^t)$ produced by (4), the main result associated to AMP is the description of the limiting distribution of

$$\mu^{(\boldsymbol{x}^1, \cdots, \boldsymbol{x}^t)} := \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i^1, \cdots, x_i^t)}$$

as $n \to \infty$ in terms of a multivariate Gaussian vector whose covariance matrix is described by the Density Evolution Equations.

1.3 Density Evolution Equations

Density Evolution (DE) equations are a set of recursive equations that define a sequence of deterministic, symmetric, positive semi-definite matrices, which are central objects in the analysis of AMP algorithms. These matrices are covariance matrices associated to multivariate normal distributions which describe the asymptotic behavior of the AMP iterates (and their correlations) as n goes to infinity.

Given a set of activation functions $h_t : \mathbb{R} \to \mathbb{R}$ and a initial constant vector $\boldsymbol{x}^0 = x_0 \mathbf{1}_n \in \mathbb{R}^n$, the Density Evolution equations associated to the AMP (4) with $\sqrt{n}W \sim \mathrm{GOE}(n)$ is a sequence of $t \times t$ matrices $(R^t)_{t \in \mathbb{N}^\star}$ defined recursively as follows,

$$R^1 = (h(x_0))^2 \qquad \text{and} \qquad R^{t+1} = \mathbb{E} \begin{bmatrix} h_t(x_0) \\ h_t(Z_1) \\ \cdots \\ h_t(Z_t) \end{bmatrix} \begin{bmatrix} h_t(x_0)\ h_t(Z_1) \cdots h_t(Z_t) \end{bmatrix},$$

where $(Z_1, \cdots, Z_t) \sim \mathcal{N}_t(0, R^t)$. Notice that in particular, the variances $\sigma_t^2 = \mathbb{E}\, Z_t^2$ satisfy a simple recursion equation given by:

$$\sigma_0^2 = h_0^2(x_0) \qquad \text{and} \qquad \sigma_{t+1}^2 = \mathbb{E}\, h_t^2(\sigma_t \xi) \qquad \text{where} \quad \xi \sim \mathcal{N}(0,1)\,. \tag{5}$$

With the family of covariance matrices (R^t) at hand, we can express the limiting statistical properties of measure $\mu^{(\boldsymbol{x}^1, \cdots, \boldsymbol{x}^t)}$ which captures both the asymptotic properties of the iterates $\boldsymbol{x}^t$ and the dependence between the iterates $\boldsymbol{x}^1, \cdots, \boldsymbol{x}^t$:

$$\mu^{(\boldsymbol{x}^1, \cdots, \boldsymbol{x}^t)} \xrightarrow[n \to \infty]{\text{weak}, L^2} \mathcal{N}_t(0, R^t)$$

in probability (see [18] for sharper convergence results). Stated differently, for any test functions $\varphi : \mathbb{R}^t \to \mathbb{R}$ and $\psi : \mathbb{R} \to \mathbb{R}$,

$$\frac{1}{n} \sum_{i=1}^n \varphi(x_i^1, \cdots, x_i^t) \xrightarrow[n \to \infty]{\mathbb{P}} \mathbb{E}\, \varphi(Z_1, \cdots, Z_t) \qquad \text{and} \qquad \frac{1}{n} \sum_{i=1}^n \psi(x_i^t) \xrightarrow[n \to \infty]{\mathbb{P}} \mathbb{E}\, \psi(\sigma_t \xi)\,, \tag{6}$$

where $\xi \sim \mathcal{N}(0,1)$, $\xrightarrow{\mathbb{P}}$ stands for the convergence in probability and $(\sigma_t)_{t \geq 0}$ is a sequence of positive numbers defined recursively by (5).

In [19], we show that the DE equations used to study an AMP with an elliptic matrix do not depend on the correlation coefficient, the latter being included in the formulation of the AMP recursion, and more specifically in the Onsager term. In [21], the case of a symmetric random matrix with a general variance profile S is handled.

In the case of a general variance profile, the description of the asymptotic behavior of the iterates becomes more involved and instead of having a multivariate Gaussian vector $(Z_1, \cdots, Z_t)$ we have a family of n-dimensional vectors $(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t)$.

In the following definition, we give a general description of the DE equations associated to a variance profile matrix S. We now consider that the activation function depend on an additional parameter $\boldsymbol{\eta}$ and we no longer express the dependence in t using a subscript, it is now included in the arguments of function h.

Definition 1.3 Let $\boldsymbol{x}^0 = (x_i^0) \in \mathbb{R}^n$ and $\boldsymbol{\eta} = (\eta_i) \in \mathbb{R}^n$ be two deterministic vectors, $S = \left(s_{ij}\right)_{1 \le i,j \le n}$ a matrix with non-negative elements and $h : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ an activation function.

a) *Initialization.* For any $i \in [n]$, define the non-negative numbers H_i^0 and R_i^1 as

$$H_i^0 := h^2(x_i^0, \eta_i, 0) \qquad \text{and} \qquad R_i^1 := \sum_{j=1}^n s_{ij} H_j^0 \,.$$

Let $Z_i^1 \sim \mathcal{N}(0, R_i^1)$, assume that for all $i \in [n]$, the Z_i^1's are independent and set

$$\boldsymbol{Z}^1 = (Z_i^1)_{i \in [n]} \,.$$

b) *Step 1.* Let $\boldsymbol{Z}^1 = (Z_i^1)_{i \in [n]}$ be given and $i \in [n]$ be fixed. Let

$$H_i^1 = \mathbb{E}\begin{bmatrix} h(x_i^0, \eta_i, 0) \\ h(Z_i^1, \eta_i, 1) \end{bmatrix}\begin{bmatrix} h(x_i^0, \eta_i, 0) \,, & h(Z_i^1, \eta_i, 1) \end{bmatrix} \qquad \text{and} \qquad R_i^2 = \sum_{j=1}^n s_{ij} H_j^1 \,.$$

Notice that the 1×1 upper left corner of R_i^2 coincides with R_i^1. Let Z_i^2 be such that $\vec{Z}_i^2 := (Z_i^1, Z_i^2) \sim \mathcal{N}_2(0, R_i^2)$, and such that for all $i \in [n]$, the $\vec{Z}_i^2$'s are independent. Set $\boldsymbol{Z}^2 = (Z_i^2)$.

c) *Step t.* Let the covariance matrix $R_i^t \in \mathbb{R}^{t \times t}$ and the $\mathbb{R}^n$ vectors $\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t$ be given, where

$$\vec{Z}_i^t := (Z_i^1, \cdots, Z_i^t) \sim \mathcal{N}_t(0, R_i^t) \,,$$

and where all the $\vec{Z}_i^t$'s are independent for $i \in [n]$. Let

$$H_i^t = \mathbb{E}\begin{bmatrix} h(x_i^0, \eta_i, 0) \\ h(Z_i^1, \eta_i, 1) \\ \vdots \\ h(Z_i^t, \eta_i, t) \end{bmatrix}\begin{bmatrix} h(x_i^0, \eta_i, 0) & h(Z_i^1, \eta_i, 1) & \cdots & h(Z_i^t, \eta_i, t) \end{bmatrix}$$

and $R_i^{t+1} = \sum_{j=1}^n s_{ij} H_j^t$. Notice that the $t \times t$ upper left corner of matrix R_i^{t+1} coincides with R_i^t. Let Z_i^{t+1} be such that

$$\vec{Z}_i^{t+1} := (Z_i^1, Z_i^2, \cdots, Z_i^{t+1}) \sim \mathcal{N}_{t+1}(0, R_i^{t+1})$$

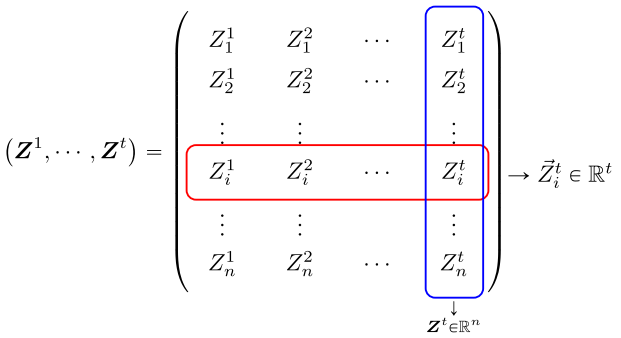


Fig. 1 The Gaussian matrix $(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t)$, the notations $\boldsymbol{Z}^t$ and $\vec{Z}_i^t$. Rows $Z_i = (Z_i^t,\ t \geq 1)$ are independent. The correlations within each row are described by the DE equations: $\vec{Z}_i^t \sim \mathcal{N}_t(0, R_i^t)$, see Definition 1.3.

and such that for all $i \in [n]$, the $\vec{Z}_i^{t+1}$'s are independent. Set $\boldsymbol{Z}^{t+1} = (Z_i^{t+1})$. Consider the sequence of n-dimensional Gaussian random vectors $\left(\boldsymbol{Z}^t\right)_{t\in\mathbb{N}}$. We denote

$$\left(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t\right) \sim \mathrm{DE}\left(S, h, \boldsymbol{x}^0, \boldsymbol{\eta}, t\right) .$$

We also define $Z_i = (Z_i^t)_{t\geq 1}$. The sequences $\{Z_i\}_{i\in[n]}$ are centered, Gaussian, and independent. The notations $\boldsymbol{Z}^t$ and $\vec{Z}_i^t$ are described in Fig. 1.

1.4 Main Result (Informal)

As already mentioned, numerous studies [7, 19, 21, 28] have extended the AMP algorithm to cover more complex random matrix models W. For each new matrix model, two key questions must be addressed:

a) How to define a proper Onsager term?
b) What are the associated DE equations ?

In this paper, we answer both questions for the matrix model described in Sect. 1.1. We show that the DE equations are given by Definition 1.3; in particular they only depend on the variance profile and not on the correlation profile. Let W be given by (3), $h : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ an activation function, $\boldsymbol{x}^0, \boldsymbol{\eta} \in \mathbb{R}^n$ deterministic vectors and

$$V = \left(S \odot S^\top\right)^{\odot 1/2} \odot T ,$$

where S and T are, respectively, the variance and correlation profiles of the random matrix W, and $(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_t)$ be given by the DE equations. We identify a possible Onsager term as

$$\mathrm{ONS}_t = \mathrm{diag}\left(V \mathbb{E}\frac{\partial h}{\partial x}(\boldsymbol{Z}^t, \boldsymbol{\eta}, t)\right) h(\boldsymbol{x}^{t-1}, \eta, t-1) ,$$

and consider the AMP

$$\boldsymbol{x}^{t+1} = Wh(\boldsymbol{x}^t, \boldsymbol{\eta}, t) - \mathrm{ONS}_t \,.$$

We shall prove that for any appropriate test function $\varphi : \mathbb{R}^{t+1} \to \mathbb{R}$ and uniformly bounded sequence $(\beta_i^{(n)})_{i\in[n]}$ of real numbers, the following convergence holds true

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\beta_i^{(n)}\varphi(\eta_i, x_i^1, \cdots, x_i^t) - \beta_i^{(n)}\varphi(\eta_i, \vec{Z}_i^t)\right\} \xrightarrow[n\to\infty]{\mathbb{P}} 0\,,$$

where the $\vec{Z}_i^t$'s are defined in Definition 1.3. The formal assumptions and statement are provided in Sect. 2.

Remark 1.4 As a consequence of the variance profile structure, each t-uple $(x_i^1, \cdots, x_i^t)$ needs to be compared to $\vec{Z}_i^t$ in the convergence above, a situation substantially more complex than in (6).

1.5 Motivation from Theoretical Ecology

The analysis of large ecological networks (foodwebs) and complex systems has garnered significant attention in recent years, with numerous studies leveraging tools from random matrix theory

[3, 11, 13]. In this perspective, large Lotka–Volterra (LV) models [1] describe the dynamics of the vector of the species abundances $\boldsymbol{x}(s) = (x_i(s))_{i\in[n]}$ for $s \in [0, \infty)$ in a series of coupled differential equations where the interactions are encoded by a random matrix A whose entries A_{ij}'s represent the effect of species j on species i. The more complex the matrix model A, the better the modeling of the network.

In a series of articles [2, 19, 21], AMP algorithms were designed in this context to analyze the statistical properties of the globally stable equilibrium $\boldsymbol{x}^\star$ (when it exists) of the vector $\boldsymbol{x}(s)$, depending on the random matrix model (symmetric models in [2, 21], elliptic model in [19]). More specifically, let $\boldsymbol{z} \in \mathbb{R}^n$ be the solution of the fixed-point equation:

$$\boldsymbol{z} = (A - I_n)\,\boldsymbol{z}^+ + \mathbf{1}_n\,, \qquad \boldsymbol{z}^+ = \boldsymbol{z} \vee 0\,,$$

which can be shown to be unique under a condition on A (see [2] for details), then the equilibrium $\boldsymbol{x}^\star$ is given by $\boldsymbol{x}^\star = \boldsymbol{z}^+$. Extracting statistical information from $\boldsymbol{x}^\star$ is a non-trivial task as the dependence of $\boldsymbol{x}^\star$ to A is highly nonlinear. However this task can be performed by designing a specific AMP algorithm.

In a foodweb, the effect $j \to i$ of species j on species i is a priori different from the effect $i \to j$. Moreover, recent empirical evidence [10] has shown that in a foodweb of size n a given species only interacts with a small number $K_n \ll n$ of other species. One may want to go one step further in modeling foodwebs, and for instance consider block structures with subpopulations with homogeneous statistical features [12].

All these desirable features naturally motivate the study of non-Symmetric and possibly sparse random matrices, with variance and correlation profiles. Such a model is at the heart of the AMP developed in this article.

In a forthcoming work, we intend to design improved matrix models for foodwebs and to analyze via AMP techniques the equilibria of associated large LV models.

1.6 Outline of the Article

In Sect. 2 we formally state the assumptions and the main result of the article, namely Theorem 2.1, together with examples, an extension to non-centered random matrices, and open questions. The remaining sections are devoted to the proof of the main result (see also Sect. 2.8 for a precise roadmap of the proof). In Sect. 3, we state a matrix AMP for polynomial activation functions, see Theorem 3.3. Section 4 is the heart of the proof of Theorem 3.3. It is based on combinatorial techniques which build upon [7] and [21]. In Sect. 5 we generalize the previous AMP for more general functions, and relax the assumption that matrix W should have null diagonal (an assumption made to handle the combinatorics in the proof of Theorem 3.3).

1.7 Notations

Denote by $|\mathcal{S}|$ the cardinality of a set $\mathcal{S}$. We often (but not systematically) use bold letters for vectors $\boldsymbol{a} = (a_i)_{i\in[n]}$, $\boldsymbol{b} = (b_j)_{j\in[k]}$, etc. If $\boldsymbol{a} = (a_\ell) \in \mathbb{R}^q$ and $\boldsymbol{m} = (m_\ell) \in \mathbb{N}^q$ is a multi-index, we denote by $\boldsymbol{a}^{\boldsymbol{m}} = \prod_{\ell\in[q]} a_\ell^{m_\ell}$.

Denote by $\mathbf{1}_n$ (or $\mathbf{1}$ if the context is obvious) the $n \times 1$ vector of ones and by $\mathbf{1}_{n\times p}$ the matrix $\mathbf{1}_{n\times p} = \mathbf{1}_n\mathbf{1}_p^\top$ where matrix $A^\top$ stands for the transpose of A. For $\boldsymbol{a} \in \mathbb{R}^n$, $\mathrm{diag}(\boldsymbol{a})$ stands for the $n \times n$ diagonal matrix with diagonal elements the a_i's. If $\boldsymbol{a} \in \mathbb{R}^n$ is a vector, $\|\boldsymbol{a}\|$ stands for its Euclidian norm and $\|\boldsymbol{a}\|_n := \|\boldsymbol{a}\|/\sqrt{n}$ for its normalized Euclidian norm. If A is a matrix, $\|A\|$ stands for its spectral norm.

If $f : \mathbb{R} \to \mathbb{R}$ and $\boldsymbol{a} = (a_i)_{i\in[n]}$ a vector, denote by $f(\boldsymbol{a}) = (f(a_i))_{i\in[n]}$ with obvious generalizations $f(\boldsymbol{a}, \boldsymbol{b}) = (f(a_i, b_i))$ for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$. Let $f(x, y, t)$ a real function with $(x, y, t) \in \mathbb{R}^2 \times \mathbb{N}$, denote by $\partial f = \frac{\partial f}{\partial x}$. Let $\boldsymbol{a} \in \mathbb{R}^n$ and $I \subset [n]$, then $\langle \boldsymbol{a}\rangle_n = \frac{1}{n}\sum_{i\in[n]} a_i$ and $\langle \boldsymbol{a}\rangle_I = \frac{1}{|I|}\sum_{i\in I} a_i$. The empirical measures $\mu^{\boldsymbol{a}}$ and $\mu^{\boldsymbol{a}^1,\cdots,\boldsymbol{a}^t}$ of vector $\boldsymbol{a} = (a_i)_{i\in[n]}$ and vectors $\boldsymbol{a}^1, \cdots, \boldsymbol{a}^t$ in $\mathbb{R}^n$ stand for

$$\mu^{\boldsymbol{a}} = \frac{1}{n}\sum_{i\in[n]} \delta_{a_i} \qquad \text{and} \qquad \mu^{\boldsymbol{a}^1,\cdots,\boldsymbol{a}^t} = \frac{1}{n}\sum_{i\in[n]} \delta_{(a_i^1,\cdots,a_i^t)},$$

where δ_x is the Dirac measure on $\mathbb{R}$ and $\delta_{(x^1,\cdots,x^t)}$, the Dirac measure on $\mathbb{R}^t$. Convergence in probability is denoted by $\xrightarrow{\mathbb{P}}$.

2 AMP for General Non-Symmetric Random Matrices

Assumptions are introduced in Sect. 2.2. The main result, Theorem 2.1, is stated in Sect. 2.3. In Sect. 2.5, we provide two examples, one focusing on the correlation profile, the second on a sparse variance profile. In Sect. 2.6, we extend the AMP result

to a non-centered random matrix model. Finally, we provide in Sect. 2.8 a detailed outline of the proof of the main theorem.

2.1 The General Framework of the AMP Recursions

Let X be a $n \times n$ T-correlated matrix and S a $n \times n$ matrix with non-negative coefficients. Recall the definition of $W = S^{\odot 1/2} \odot X$ in Eq. (3) and define matrix V as follows

$$V = \left(V_{ij}\right)_{i,j=1}^{n} = \left(S \odot S^{\top}\right)^{\odot 1/2} \odot T\,. \tag{7}$$

Notice that $\mathbb{E}\left[W \odot W^{\top}\right] = V$.

Let $h : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ be a measurable function such that for all $(\eta, t) \in \mathbb{R} \times \mathbb{N}$, the derivative $\partial h(\cdot, \eta, t)$ exists almost everywhere[1]. We denote as ∂h any measurable function that coincides with this derivative almost everywhere. For $\boldsymbol{x}, \boldsymbol{\eta} \in \mathbb{R}^n$ and $t \in \mathbb{N}$, denote $h(\boldsymbol{x}, \boldsymbol{\eta}, t) = (h(x_i, \eta_i, t))_{i \in [n]}$.

Definition 2.1 Let X be a $n \times n$ T-correlated matrix following Definition 1.1, W, V given by (3), (7), and $\boldsymbol{x}^0, \boldsymbol{\eta} \in \mathbb{R}^n$. Let $h : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ a measurable function such that ∂h exists. Let $\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t$ be $\mathbb{R}^n$-valued Gaussian vectors defined in Def. 1.3. Define the $\mathbb{R}^n$-valued random sequence $(\boldsymbol{x}^t)_{t \geq 1}$ recursively as follows,

$$\begin{cases} \boldsymbol{x}^1 & = W h(\boldsymbol{x}^0, \boldsymbol{\eta}, 0)\,, \\ \boldsymbol{x}^{t+1} & = W h(\boldsymbol{x}^t, \boldsymbol{\eta}, t) - \operatorname{diag}\left(V \mathbb{E}\partial h(\boldsymbol{Z}^t, \boldsymbol{\eta}, t)\right) h(\boldsymbol{x}^{t-1}, \boldsymbol{\eta}, t-1) \quad \text{for} \quad t \geq 1\,. \end{cases} \tag{8}$$

The following notation will be used in the sequel:

$$\left(\boldsymbol{x}^t\right)_{t \geq 1} = \text{AMP-Z}\left(X, S, h, \boldsymbol{x}^0, \boldsymbol{\eta}\right)\,, \quad \boldsymbol{x}^0, \boldsymbol{\eta} \in \mathbb{R}^n\,. \tag{9}$$

Remark 2.2 The parameter $\boldsymbol{\eta} \in \mathbb{R}^n$ which is fixed once for all in the recursions can be seen as an extra degree of freedom in the design of the algorithm.

Remark 2.3 (versatility) Definition 2.1 generalizes many frameworks found in the literature.

a) For a symmetric matrix X where $T = \mathbf{1}_{n \times n}$ and $S = \frac{\mathbf{1}_{n \times n}}{n}$, one gets the AMP in [7].
b) By taking a sparse symmetric matrix S, one recovers the AMP in [21].
c) The elliptic AMP studied in [19] is obtained by taking $S = \frac{\mathbf{1}_{n \times n}}{n}$ and $T = \rho \mathbf{1}_{n \times n}$ for $\rho \in [-1, 1]$. In the latter, the AMP recursion writes

$$\boldsymbol{x}^{t+1} = W\, h\left(\boldsymbol{x}^t, \boldsymbol{\eta}, t\right) - \rho \left\langle \partial h\left(\boldsymbol{x}^t, \boldsymbol{\eta}, t\right) \right\rangle_n h\left(\boldsymbol{x}^{t-1}, \boldsymbol{\eta}, t-1\right)\,.$$

One can notice that the Onsager term is slightly different. We will come back to this later in Sect. 2.4.

[1] Notice that if h is Lipschitz with respect to the first variable, then it is differentiable almost everywhere by Rademacher's theorem.

2.2 Assumptions

We present hereafter the assumptions that will be used in the sequel, some of which already appeared in [21].

Assumption A- 1 *(moments) Let $T = (\tau_{ij})_{1\le i,j\le n}$ be a symmetric matrix with $\tau_{ij} \in [-1,1]$ and X a random T-correlated matrix following Definition 1.1. For every $k \ge 1$ there exists a positive real number $C_{mom}(k) > 0$ such that for every $n \ge 1$ and all $i, j \in [n]$*

$$\left(\mathbb{E}\left|X_{ij}\right|^k\right)^{1/k} \le C_{mom}(k)\,.$$

Assumption A- 2 *(variance profile) Let (K_n) a sequence of positive integers diverging to $+\infty$ and satisfying $K_n \le n$. The deterministic $n \times n$ matrix $S = (s_{ij})_{1\le i,j\le n}$ has non-negative elements and satisfies the following: there exist positive constants $C_{card}, C_S, c_S > 0$ such that for every $n \ge 1$ and all $i, j \in [n]$,*

$$\left|\left\{j \in [n] : s_{ij} > 0\right\}\right| \le C_{card}\,K_n\,, \qquad s_{ij} \le \frac{C_S}{K_n} \quad \text{and} \quad \sum_{\ell=1}^{n} s_{i\ell} \ge c_S\,.$$

Remark 2.4 (on Assumption A-2)

(a) This assumption describes the sparsity of the variance profile $S \in \mathbb{R}^{n\times n}$. It consists of three sub-assumptions: there exists a sequence (K_n), with $K_n \le n$ and $K_n \to \infty$, and positive constants $C_{\text{card}}, C_S, c_S$ such that for every $n \ge 1$ and all $i, j \in [n]$: (i) the number of nonzero elements in the i-th row of S is at most $C_{\text{card}}K_n$; (ii) each variance satisfies $s_{ij} \le C_S K_n^{-1}$; and (iii) the row sums of S are bounded below by c_S.

(b) Although these three conditions are grouped under a single assumption A-2, they are not all used with the same frequency or importance. For instance, in all combinatorial arguments, A-2-(iii) is not required, and we expect that this sub-assumption is mainly a limitation of our proof technique. Indeed, A-2-(iii) becomes unnecessary when the AMP is built using only polynomial activation functions; it is therefore a technicality to ensure that we can approximate general activation functions by polynomial ones.

Remark 2.5 In the works [5] and [22], the authors study AMP-type algorithms with symmetric random matrices equipped with a variance profile. In [5], the analysis focuses on the classical AMP with a Gaussian matrix, while [22] investigates a broader class of generalized first-order methods, of which AMP is a particular case (see the canonical examples in Sect. 2.1). Both works provide non-asymptotic analyses of the corresponding algorithms.

For comparison, the random matrix models considered in [5], [22], and in our work are, respectively

$$\frac{1}{\sqrt{n}}\,S^{(1)} \odot G, \qquad \frac{1}{\sqrt{n}}\,S^{(2)} \odot A, \qquad \text{and} \qquad \frac{1}{\sqrt{K_n}}\,S^{(3)} \odot X,$$

where $K_n \to +\infty$ controls the sparsity level. While one could mimic sparsity in [5] and [22] by setting many entries of the variance profiles to zero, the normalization factor $1/\sqrt{n}$ in those works does not depend on sparsity and thus does not yield the correct scaling in the genuinely sparse regime. In contrast, our framework explicitly models sparsity through both a structural condition on the variance profile $S^{(3)}$ $(= K_n^{1/2} S^{\odot 1/2})$ (at most CK_n nonzero entries per row) and the normalization factor $1/\sqrt{K_n}$, which scales appropriately with the sparsity level. In particular, if one considers a sparse matrix normalized by $1/\sqrt{n}$ with $K_n/n \to 0$, the density-evolution equations degenerate to a trivial limit.

The following technical assumption ensures that the spectral norm of the matrix W is almost surely bounded by a constant as n goes to infinity.

Assumption A- 3 *(lower bound on the sparsity level) Let A-1 and A-2 hold for the random matrix X and the variance profile S, and consider associated C_{mom} and (K_n). There exist positive real numbers $\nu, C > 0$ such that for every $k, n \geq 1$*

$$C_{mom}(k) \leq C\, k^{\nu/2} \qquad \textit{and} \qquad K_n \geq C\, \log^{(\nu \vee 1)}(n)\,.$$

Remark 2.6 (on Assumption A-3)

(a) The moment condition $C_{\mathrm{mom}}(k) \leq C\, k^{\nu/2}$ is standard. For example, it is fulfilled with $\nu = 1$ for subGaussian entries.

(b) Assumptions A-2 and A-3 describe the sparsity level one can expect for matrix W. The sequence K_n is an upper bound of the number of non-vanishing elements of W per row. It must be at least logarithmic in n (up to the power $\nu \vee 1$) but can be much smaller than n.

(c) As will appear later in Proposition 5.5, the logarithmic lower bound on K_n and the upper bound for the moments of X's entries are technical conditions needed for bounding the spectral norm of the random matrix W.

We also consider initial conditions for the initial vector $\boldsymbol{x}^0$ and for the parameter vector $\boldsymbol{\eta} \in \mathbb{R}^n$.

Assumption A- 4 *(initial and parameter vectors) Let $\boldsymbol{x}^0 = (x_i^0) \in \mathbb{R}^n$, $\boldsymbol{\eta} = (\eta_i) \in \mathbb{R}^n$ be deterministic vectors and consider the sequences $(\boldsymbol{x}^0)_n$ and $(\boldsymbol{\eta})_n$. There exist two compact sets $\mathcal{Q}_x \subset \mathbb{R}$ and $\mathcal{Q}_\eta \subset \mathbb{R}$ such that*

$$\{x_i^0\,,\ i \in [n]\,,\ n \geq 1\} \subset \mathcal{Q}_x \qquad \textit{and} \qquad \{\eta_i\,,\ i \in [n]\,,\ n \geq 1\} \subset \mathcal{Q}_\eta\,.$$

Remark 2.7 (on Assumption A-4) This assumption requires the AMP initialization to lie in a compact set $\mathcal{Q}_x$, which facilitates uniform bounds throughout the proofs, for instance, in controlling the covariance matrices in Lemma 5.2 and some quantities in the combinatorial arguments used for example in Proposition 4.2 (see Equation (27) in Theorem 3.3). This assumption is mainly technical and simplifies several bounds; we believe the results should still hold for non-compact initializations (e.g., Gaussian vectors), at the cost of more involved arguments. An interesting alternative to explore

could be the following condition

$$\sum_{i=1}^{n} \exp(\|x_i^0\|^2/C) \leq Cn \quad \text{with probability tending to 1,}$$

found in (C, d)-regular sequence condition of [7] (see assumption 3 in Definition 4).

Assumption A- 5 *(Regularity of the activation functions) Let $h : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ be a measurable function. For every $t \in \mathbb{N}$, there exists a positive number L such that for every $x, y, \eta \in \mathbb{R}$,*

$$|h(x, \eta, t) - h(y, \eta, t)| \quad \leq \quad L\,|x - y|\,.$$

For every $t \in \mathbb{N}$, there exists a continuous non-decreasing function $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ with $\kappa(0) = 0$ and a compact set $\mathcal{Q}_\eta \subset \mathbb{R}$ such that for every $x \in \mathbb{R}$ and $\eta, \eta' \in \mathcal{Q}_\eta$,

$$\left|h(x, \eta, t) - h(x, \eta', t)\right| \leq \kappa\left(\left|\eta - \eta'\right|\right)(1 + |x|)\,.$$

Assumption A- 6 *(non-degeneracy condition over h) Let $h : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ be a measurable function. There exist two compact sets $\mathcal{Q}_x \subset \mathbb{R}$ and $\mathcal{Q}_\eta \subset \mathbb{R}$ with the following properties:*

(1) There exists a constant $c > 0$ such that

$$\inf_{x \in \mathcal{Q}_x, \eta \in \mathcal{Q}_\eta} h^2(x, \eta, 0) \quad \geq \quad c\,.$$

(2) For every $t \geq 1$, there exist two positive real numbers $c_h(t), D_h(t) > 0$ such that

$$\inf_{\eta \in \mathcal{Q}_\eta} \int_{-D_h(t)}^{D_h(t)} h^2(x, \eta, t)dx \quad \geq \quad c_h(t)\,.$$

There are tight links between the assumptions. In particular, the parameter ν of A-3 controls the moments bounds $(C_{mom}(k))$ given by A-1 and the sparsity level K_n given by A-2, the compact sets $\mathcal{Q}_\eta$ and $\mathcal{Q}_x$ of A-5 and A-6 will be given by A-4.

2.3 Main Result

Recall the definition of a *pseudo-Lipschitz* function. A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be pseudo-Lipschitz (PL) if there exists a constant L such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ the following inequality is satisfied:

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \;\leq\; L\,\|\boldsymbol{x} - \boldsymbol{y}\|\,(1 + \|\boldsymbol{x}\| + \|\boldsymbol{y}\|)\,.$$

We are now in position to state our main result.

Theorem 2.1 *Let Assumptions A-1 to A-6 hold true, with associated v, $\mathcal{Q}_\eta$ and $\mathcal{Q}_x$. Consider the AMP*

$$(\boldsymbol{x}^t)_{t\geq 1} = \text{AMP-Z}\left(X, S, h, \boldsymbol{x}^0, \boldsymbol{\eta}\right)$$

as defined in Definition 2.1, and the sequence of n-dimensional Gaussian random vectors $\left(\boldsymbol{Z}^t\right)_{t\in\mathbb{N}}$ defined by the DE equations in Definition 1.3:

$$(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t) \sim \text{DE}(S, h, \boldsymbol{x}^0, \boldsymbol{\eta}, t)\,.$$

Let $t \geq 1$ and $\boldsymbol{\beta} = (\beta_i^{(n)}) \in \mathbb{R}^n$ uniformly bounded, i.e. $\sup_n \max_{i\in[n]} |\beta_i^{(n)}| < \infty$. For any pseudo-Lipschitz test function $\varphi : \mathbb{R}^{t+1} \to \mathbb{R}$, it holds that

$$\frac{1}{n}\sum_{i\in[n]} \beta_i^{(n)} \left\{\varphi\left(\eta_i, x_i^1, \cdots, x_i^t\right) - \mathbb{E}\left[\varphi\left(\eta_i, Z_i^1, \cdots, Z_i^t\right)\right]\right\} \xrightarrow[n\to\infty]{\mathbb{P}} 0\,.$$

2.4 Alternative Onsager Terms

Recall the introduction of AMP-Z in Definition 2.1. It might be convenient to consider alternative Onsager terms in the AMP recursion and replace the diagonal matrix $\text{diag}(V\mathbb{E}\partial h(\boldsymbol{Z}^t, \boldsymbol{\eta}, t))$ by one of the two following terms

$$\text{diag}\left(V\,\partial h(\boldsymbol{x}^t, \boldsymbol{\eta}, t)\right) \quad \text{or} \quad \text{diag}\left(W \odot W^\top \partial h(\boldsymbol{x}^t, \boldsymbol{\eta}, t)\right)\,, \tag{10}$$

depending on the context.

For example, the Onsager term built upon $\text{diag}\left(W \odot W^\top \partial h(\boldsymbol{x}^t, \boldsymbol{\eta}, t)\right)$ is better suited for the combinatorial arguments developed in Sect. 4 as it directly involves the entries of matrix W, and the loss with respect to the original recursion should be asymptotically negligible since $\mathbb{E}(W \odot W^\top) = V$. The Onsager term built upon $\text{diag}\left(V\,\partial h(\boldsymbol{x}^t, \boldsymbol{\eta}, t)\right)$ naturally appears in [2, 19].

In this perspective we introduce new notations. Denote by

$$\left(\boldsymbol{x}^t\right)_{t\geq 1} := \text{AMP-W}\left(X, S, h, \boldsymbol{x}^0, \boldsymbol{\eta}\right), \tag{11}$$

the recursive procedure defined by

$$\begin{cases} \boldsymbol{x}^1 &= Wh(\boldsymbol{x}^0, \boldsymbol{\eta}, 0)\,, \\ \boldsymbol{x}^{t+1} &= Wh(\boldsymbol{x}^t, \boldsymbol{\eta}, t) - \text{diag}\left(W \odot W^\top \partial h(\boldsymbol{x}^t, \boldsymbol{\eta}, t)\right) h(\boldsymbol{x}^{t-1}, \boldsymbol{\eta}, t-1) \quad \text{for} \quad t \geq 1\,. \end{cases}$$

Similarly, denote by

$$\left(\boldsymbol{x}^t\right)_{t\geq 1} := \text{AMP}\left(X, S, h, \boldsymbol{x}^0, \boldsymbol{\eta}\right)\,, \tag{12}$$

the recursive procedure defined by

$$\begin{cases} \boldsymbol{x}^1 & = Wh(\boldsymbol{x}^0, \boldsymbol{\eta}, 0)\,, \\ \boldsymbol{x}^{t+1} & = Wh(\boldsymbol{x}^t, \boldsymbol{\eta}, t) - \operatorname{diag}\left(V\partial h(\boldsymbol{x}^t, \boldsymbol{\eta}, t)\right) h(\boldsymbol{x}^{t-1}, \boldsymbol{\eta}, t-1) \quad \text{for} \quad t \geq 1\,. \end{cases}$$

We believe that none of these three Onsager terms should change the general asymptotics of the AMP. However, a complete proof of this fact is not established. We only prove the AMP result for the AMP-Z formulation, see Theorem 2.1.

In conclusion, there are three possible formulations of the Onsager term, leading to the AMP-Z, AMP, and AMP-W algorithms depending on whether the correction involves an expectation over Gaussian variables, the current iterate, or the random matrix itself. Among these, the "AMP" formulation appears to be the most natural, as it yields a deterministic Onsager term depending only on the variance profile and the iterates, making it straightforward to implement. However, our proof relies on combinatorial arguments and polynomial approximations of the activation function, which are more conveniently handled within the "AMP-Z" formulation. Consequently, we establish our main AMP-type result for the "AMP-Z" version. Lemma 5.3 provides a comparison between the "AMP" and "AMP-W" formulations, but only for polynomial activations.

2.5 Examples of AMP

We provide hereafter three examples of matrix models where we work out the specific AMP recursion and DE equations. All three matrix models are of practical interest, with applications in fields such as theoretical ecology, where random matrices represent species interaction matrices in large ecological systems (see [1]).

Example 1: Elliptic Matrix Model

Let $\rho > 0$, and consider an $n \times n$ random matrix $X = (X_{ij})_{i,j=1}^n$ whose entries satisfy the following properties:

- The random variables X_{ij} are centered with variance 1.
- For each $i < j$, the pair (X_{ij}, X_{ji}) has covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$
- The collection $\{X_{ii},\ (X_{ij}, X_{ji}) : (i,j) \in [n]^2,\ i < j\}$ consists of independent random elements.

A particular instance of this model is the *Elliptic Ensemble*, in which the entries are Gaussian (the diagonal entries may have a different variance, but this can be handled without difficulty).

Consider the following matrix model:

$$W = \frac{1}{\sqrt{n}} X.$$

For a uniform variance profile the Density Evolution equations (1.3) become asymptotic and do not depend on the dimension n, i.e.

$$R^1 = f^2(x_0) \in \mathbb{R}^{1\times 1}, \qquad R^{t+1} = \mathbb{E}\begin{bmatrix} f(x_0) \\ f(Z_1) \\ \cdots \\ f(Z_t) \end{bmatrix} \begin{bmatrix} f(x_0) & f(Z_1) & \cdots & f(Z_t) \end{bmatrix} \in \mathbb{R}^{(t+1)\times(t+1)}, \quad (13)$$

where $(Z_1, \cdots, Z_t) \sim \mathcal{N}_t(0, R^t)$. The following is a corollary of our main theorem.

Corollary 2.2 *Let Assumption A-1 and A-3 hold for the above matrix model W. Let $\boldsymbol{x}^0$ be a constant an initialization vector $\boldsymbol{x}^0 = x_0\mathbb{1}$, and let h be an activation function satisfying Assumption A-5. Assume further that ∂h is continuous λ-almost everywhere, where λ denotes the Lebesgue measure. Consider the AMP sequence $(\boldsymbol{x}^t)_t$ defined recursively as*

$$\boldsymbol{x}^{t+1} = Wh(\boldsymbol{x}^t) - \rho\, \langle \partial h(\boldsymbol{x}^t)\rangle_n\, h(\boldsymbol{x}^{t-1})$$

For each $t \geq 1$, consider the t-dimensional centered Gaussian vector $(Z_1, \ldots, Z_t)$ whose covariance matrix R^t is defined recursively by the (asymptotic) Density Evolution equations.

Then, for any pseudo-Lipschitz test function $\varphi : \mathbb{R}^t \to \mathbb{R}$, we have

$$\frac{1}{n}\sum_{i=1}^{n} \varphi(x_i^1, \ldots, x_i^t) \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}\big[\varphi(Z_1, \ldots, Z_t)\big].$$

The proof of this corollary is in Appendix 1.

Remark 2.8 (1) The Onsager term considered in this corollary differs from the one used in the main theorem. A detailed discussion on the possible variants of the Onsager term is provided in Sect. 2.4.

(2) This result extends the work of [19], where the matrix model was assumed to be Gaussian. Here, the Gaussian assumption is relaxed, but we obtain convergence in probability rather than almost sure convergence.

(3) To establish the result under this modified Onsager term, we require an additional assumption on the activation function, namely that its derivative ∂h is continuous almost everywhere. This assumption is also present in [18, 19].

(4) In this corollary, we assume a constant initialization vector for simplicity. However, the result can be extended to more general initializations under additional mild regularity conditions.

Example 2: Blockwise Correlated Random Matrix

This example generalizes the elliptic matrix model characterized by a single correlation coefficient ρ. Here, the matrix is allowed to have different correlation coefficients for each block. Let $n = n_1 + n_2$, X a $n \times n$ matrix partitioned into four submatrices: $X^{(11)}$,

$X^{(12)}$, $X^{(21)}$, and $X^{(22)}$, of respective sizes $n_1 \times n_1$, $n_1 \times n_2$, $n_2 \times n_1$, and $n_2 \times n_2$:

$$X = \begin{pmatrix} X^{(11)} & X^{(12)} \\ X^{(21)} & X^{(22)} \end{pmatrix}.$$

Let $X^{(11)}$ and $X^{(22)}$ be (independent) elliptic random matrices with correlation coefficient ρ_1, while each entry in $X^{(12)}$ is correlated with its symmetrically corresponding entry in $X^{(21)}$ with a coefficient ρ_2. All the entries of the random matrix X have variance 1 and satisfy A-1. Consider the normalized version of X,

$$W = \frac{X}{\sqrt{n}}.$$

With our previous formalism, this model corresponds to choosing X as a T-correlated matrix and $W = S \odot X$ where S (variance profile) and T (correlation profile) are defined by

$$S = \frac{\mathbf{1}_{n\times n}}{n} \qquad \text{and} \qquad T = \begin{pmatrix} \rho_1 \mathbf{1}_{n_1\times n_1} & \rho_2 \mathbf{1}_{n_1\times n_2} \\ \rho_2 \mathbf{1}_{n_2\times n_1} & \rho_1 \mathbf{1}_{n_2\times n_2} \end{pmatrix}.$$

Let $r_n := \frac{n_1}{n}$, $I_1 = \{1, \cdots, n_1\}$ and $I_2 = [n] \setminus I_1$, assume that $r_n \to r \in (0, 1)$ and consider the following framework: $\boldsymbol{x}^0 = x_0 \mathbf{1}_n$, the activation function $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz. Notice that f satisfies A-5, neither depends on t nor on some extra parameter $\boldsymbol{\eta}$.

Consider the recursion $(\boldsymbol{x}^t)_{t\in\mathbb{N}} = \text{AMP}\left(X, S, f, \boldsymbol{x}^0\right)$. In particular,

$$\boldsymbol{x}^{t+1} = Wf(\boldsymbol{x}^t) - \text{diag}\left(Vf'(\boldsymbol{x}^t)\right) f(\boldsymbol{x}^{t-1}),$$

where $V = T/n$. The Onsager term can be simplified here by writing $Vf'(\boldsymbol{x}^t)$ as

$$\begin{aligned} Vf'(\boldsymbol{x}^t) &= \begin{pmatrix} r_n\rho_1\langle f'(\boldsymbol{x}^t)\rangle_{I_1}\mathbf{1}_{n_1} + (1-r_n)\rho_2\langle f'(\boldsymbol{x}^t)\rangle_{I_2}\mathbf{1}_{n_1} \\ r_n\rho_2\langle f'(\boldsymbol{x}^t)\rangle_{I_1}\mathbf{1}_{n_2} + (1-r_n)\rho_1\langle f'(\boldsymbol{x}^t)\rangle_{I_2}\mathbf{1}_{n_2} \end{pmatrix} \\ &= \begin{pmatrix} r_n\rho_1\mathbf{1}_{n_1} & (1-r_n)\rho_2\mathbf{1}_{n_1} \\ r_n\rho_2\mathbf{1}_{n_2} & (1-r_n)\rho_1\mathbf{1}_{n_2} \end{pmatrix} \begin{pmatrix} \langle f'(\boldsymbol{x}^t)\rangle_{I_1} \\ \langle f'(\boldsymbol{x}^t)\rangle_{I_2} \end{pmatrix}. \end{aligned}$$

Thus

$$\boldsymbol{x}^{t+1} = Wf(\boldsymbol{x}^t) - \left[\begin{pmatrix} r_n\rho_1\mathbf{1}_{n_1} & (1-r_n)\rho_2\mathbf{1}_{n_1} \\ r_n\rho_2\mathbf{1}_{n_2} & (1-r_n)\rho_1\mathbf{1}_{n_2} \end{pmatrix} \begin{pmatrix} \langle f'(\boldsymbol{x}^t)\rangle_{I_1} \\ \langle f'(\boldsymbol{x}^t)\rangle_{I_2} \end{pmatrix}\right] \odot f(\boldsymbol{x}^{t-1}),$$

Notice that the Onsager term generalizes here the one obtained in the elliptic case (see Remark 2.3).

Not surprisingly (and as mentioned in [19] in the elliptic case), the DE equations do not depend on the correlation structure of X and reduce to

$$R^1 = f^2(x_0) \in \mathbb{R}^{1\times 1}, \quad R^{t+1} = \mathbb{E}\begin{bmatrix} f(x_0) \\ f(Z_1) \\ \cdots \\ f(Z_t) \end{bmatrix}\begin{bmatrix} f(x_0) & f(Z_1) & \cdots & f(Z_t) \end{bmatrix} \in \mathbb{R}^{(t+1)\times(t+1)},$$

where $(Z_1, \cdots, Z_t) \sim \mathcal{N}_t(0, R^t)$. In this case, Theorem 2.1 implies that for any PL test function $\varphi : \mathbb{R}^t \in \mathbb{R}$ ur main theorem implies in this case that

$$\frac{1}{n}\sum_{i\in[n]} \varphi(x_i^1, \cdots, x_i^t) \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}\varphi(Z_1, \cdots, Z_t)\,.$$

Remark 2.9 This example can easily be generalized to $K \times K$ blocks and K correlation coefficients $\rho_1, \cdots, \rho_K$.

Example 3: *d*-Regular Random Matrix

In this example, we consider a symmetric matrix X where X_{ij} are independent centered random variables with variance 1 up to the symmetry, i.e. X is a T-correlated random matrix where $T = \mathbf{1}_{n\times n}$. Let Assumption 1 hold, let $d = d_n = \lfloor C \log^{(\nu\vee 1)}(n)\rfloor$ where $\nu > 0$ is given by Assumption 3. Let A be the $n \times n$ adjacency matrix of a d-regular non oriented graph, in particular

$$\big|\{j \in [n] \mid A_{ij} = 1\}\big| = d \qquad \text{and} \qquad \big|\{i \in [n] \mid A_{ij} = 1\}\big| = d\,,$$

and consider the variance profile matrix $S = \frac{1}{d}A$. Let $f : \mathbb{R} \to \mathbb{R}$ a Lipschitz function (hence satisfying Assumption 5) and set

$$W = S \odot X = \frac{1}{d}A \odot X\,, \qquad \boldsymbol{x}^0 = x_0\mathbf{1}_n \qquad \text{and} \qquad (\boldsymbol{x}^t)_{t\in\mathbb{N}} = \mathrm{AMP}\left(X, S, f, \boldsymbol{x}^0\right).$$

Introducing the sets $I_k := \{j \in [n] \mid A_{kj} = 1\}$ and the $n \times 1$ vector $\boldsymbol{v} = \big(\langle f'(\boldsymbol{x}^t)\rangle_{I_k},\ k \in [n]\big)$, the recursion writes

$$\boldsymbol{x}^{t+1} = Wf(\boldsymbol{x}^t) - \boldsymbol{v} \odot f(\boldsymbol{x}^{t-1})\,.$$

Let us now simplify the Density Evolution equations defined 1.3 for this particular case. We notice that $H_i^0 = (h(x_0))^2 =: H^0$ does not depend on i, so $R_i^1 = \sum_{j\in I_i} \frac{1}{d} H_i^0 = H^0 := R^1$ which is also independent of i and n. By induction, we can reduce DE equations to "asymptotic" DE equations, meaning that they do not depend on n. In fact, if $R_i^t \in \mathbb{R}^{t\times t}$ is independent of i, consider $(Z_i^1, \cdots, Z_i^t) \sim \mathcal{N}_t(0, R_i^t)$, these n t-dimensional random vectors have the same law. Now let $i \in [n]$ and consider the value of R_i^{t+1},

$$R_i^{t+1} = \sum_{j\in I_i} \frac{1}{d}\mathbb{E}\begin{bmatrix} f(x_0) \\ f(Z_i^1) \\ \cdots \\ f(Z_i^t) \end{bmatrix}\begin{bmatrix} f(x_0) & f(Z_i^1) & \cdots & f(Z_i^t) \end{bmatrix} = \mathbb{E}\begin{bmatrix} f(x_0) \\ f(Z_1) \\ \cdots \\ f(Z_t) \end{bmatrix}\begin{bmatrix} f(x_0) & f(Z_1) & \cdots & f(Z_t) \end{bmatrix}$$

where $(Z_1, \cdots, Z_t) \sim \mathcal{N}_t(0, R^t)$, thus R_i^{t+1} is also independent of i and n and we recover the "asymptotic" DE equations. Our main theorem implies in this case that

$$\mu^{\boldsymbol{x}^1,\cdots,\boldsymbol{x}^t} \xrightarrow[n\to\infty]{\mathbb{P}} \mathcal{L}(Z_1, \cdots, Z_t).$$

Remark 2.10 In this setting, the matrix sparsity is governed by the parameter d, which also serves as the normalization factor, corresponding to the sparsity level K_n in Assumption 2 (i.e., $K_n = d$). While extending the analysis to constant-degree (d finite) matrices would be interesting, we believe it is outside the scope of AMP theory, which typically requires $d \to \infty$. The current lower bound $d \gtrsim \log^{(\nu\vee 1)}(n)$ is mainly a technical limitation of our proof, needed to control the spectral norm of the sparse matrix (Proposition 5.5), and could likely be relaxed with sharper arguments.

2.6 Extension to Non-centered Random Matrices

We have considered so far an AMP algorithm with a centered random matrix. We extend our AMP result to consider a non-centered matrix model. More precisely, we add to our centered random matrix model a deterministic rank-one perturbation - notice that our result could easily be generalized to any finite-rank perturbation.

Let W be a random matrix model as in Theorem 2.1, with variance profile S and correlation profile T. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ two deterministic vectors satisfying $\|\boldsymbol{u}\|, \|\boldsymbol{v}\| = \mathcal{O}(n^{-1})$. Consider the following matrix model,

$$A = \lambda \boldsymbol{u}\boldsymbol{v}^\top + W. \tag{14}$$

Before stating the AMP recursion based on matrix A, we adapt the Density Evolution equations introduced in Definition 1.3. In this section, we shall use the notation $h_t(x, \eta)$ instead of $h(x, \eta, t)$ as simplification of the notations.

Definition 2.11 Let $\boldsymbol{x}^0 = (x_i^0) \in \mathbb{R}^n$, $\boldsymbol{\eta} = (\eta_i) \in \mathbb{R}^n$, $\boldsymbol{u} = (u_i) \in \mathbb{R}^n$ and $\boldsymbol{v} = (v_i) \in \mathbb{R}^n$ be deterministic vectors, $S = \left(s_{ij}\right)_{1\le i,j\le n}$ a matrix with non-negative elements and $h : \mathbb{R}^2 \times \mathbb{N} \to \mathbb{R}$ an activation function.

a) *Initialization.* For any $i \in [n]$, define the positive numbers H_i^0, R_i^1 and μ_1 as

$$H_i^0 := \left(h_0(x_i^0, \eta_i)\right)^2, \qquad R_i^1 := \sum_{j=1}^n s_{ij} H_j^0 \qquad \text{and} \qquad \mu_1 := \lambda \left\langle \boldsymbol{v}, h_0(\boldsymbol{x}^0, \boldsymbol{\eta}) \right\rangle.$$

Let $Z_i^1 \sim \mathcal{N}(0, R_i^1)$, assume that for all $i \in [n]$, the Z_i^1's are independent and set

$$\boldsymbol{Z}^1 = (Z_i^1)_{i\in[n]} .$$

b) *Step 1.* Let $i \in [n]$ be fixed. Given Z_i^1, let

$$H_i^1 = \mathbb{E}\begin{bmatrix} h_0\left(x_i^0, \eta_i\right) \\ h_1\left(Z_i^1 + \mu_0 u_i, \eta_i\right) \end{bmatrix}\left[h_0\left(x_i^0, \eta_i\right)\, h_1\left(Z_i^1 + \mu_0 u_i, \eta_i\right)\right]$$
$$R_i^2 = \sum_{j=1}^{n} s_{ij} H_j^1 \quad \text{and} \quad \mu_2 = \lambda \mathbb{E}\left[\left\langle \boldsymbol{v}, h_1\left(\boldsymbol{Z}^1 + \mu_1 \boldsymbol{u}, \boldsymbol{\eta}\right)\right\rangle\right].$$

Let $(Z_i^1, Z_i^2) \sim \mathcal{N}_2(0, R_i^2)$, denote by $\vec{Z}_i^2 = (Z_i^1, Z_i^2)$. Assume that for all $i \in [n]$, the $\vec{Z}_i^2$'s are independent. Set $\boldsymbol{Z}^2 = (Z_i^2)$.

c) *Step t.* Let $i \in [n]$ be fixed. Given $(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t)$ and $\vec{Z}_i^t = (Z_i^1, \cdots, Z_i^t)$, let

$$H_i^t = \mathbb{E}\begin{bmatrix} h_0(x_i^0, \eta_i) \\ h_1(Z_i^1 + \mu_1 u_i, \eta_i) \\ \vdots \\ h_t(Z_i^t + \mu_t u_i, \eta_i) \end{bmatrix}\left[h_0(x_i^0, \eta_i)\; h_1(Z_i^1 + \mu_1 u_i, \eta_i) \cdots h_t(Z_i^t + \mu_t u_i, \eta_i)\right] .$$

Denote

$$R_i^{t+1} = \sum_{j=1}^{n} s_{ij} H_j^t \qquad \text{and} \qquad \mu_{t+1} = \lambda \mathbb{E}\left[\left\langle \boldsymbol{v}, h_t\left(\boldsymbol{Z}^t + \mu_t \boldsymbol{u}, \boldsymbol{\eta}\right)\right\rangle\right] .$$

Let $(Z_i^1, Z_i^2, \cdots Z_i^{t+1}) \sim \mathcal{N}_{t+1}(0, R_i^{t+1})$, denote by $\vec{Z}_i^{t+1} = (Z_i^1, Z_i^2, \cdots, Z_i^{t+1})$. Assume that for all $i \in [n]$, the $\vec{Z}_i^{t+1}$'s are independent. Set $\boldsymbol{Z}^{t+1} = (Z_i^{t+1})$.

Consider the sequence of n-dimensional Gaussian random vectors $\left(\boldsymbol{Z}^t\right)_{t\in\mathbb{N}}$. We denote

$$\left(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t\right) \sim \widetilde{\mathrm{DE}}\left(h, \boldsymbol{x}^0, S, t, \boldsymbol{u}, \boldsymbol{v}\right) .$$

We are now in position to state the AMP recursion.

$$\boldsymbol{x}^{t+1} = A h_t(\boldsymbol{x}^t, \boldsymbol{\eta}) - \mathrm{diag}\left(V \mathbb{E} \partial h_t(\boldsymbol{Z}^t + \mu_t \boldsymbol{u}, \boldsymbol{\eta})\right) h_{t-1}(\boldsymbol{x}^{t-1}, \boldsymbol{\eta}), \tag{15}$$

where $\boldsymbol{Z}^t$ and μ^t are defined as in Definition 2.11.

The following theorem describes the asymptotic behavior of $\left(\boldsymbol{x}^t\right)_{t\in\mathbb{N}}$ when n goes to infinity.

Theorem 2.3 *Let Assumptions A-1 to A-6 hold true, with associated ν, $\mathcal{Q}_\eta$ and $\mathcal{Q}_x$. Consider the AMP sequence $(\boldsymbol{x}^t)_t$ defined in (15). Consider the sequence n-dimensional Gaussian random vectors $\left(\boldsymbol{Z}^t\right)_{t\in\mathbb{N}}$ and the scalars $(\mu_t)_t$ defined by the DE equations in Definition 2.11.*

Let $t \geq 1$ and $\boldsymbol{\beta} = (\beta_i^{(n)}) \in \mathbb{R}^n$ uniformly bounded, i.e. $\sup_n \max_{i\in[n]} |\beta_i^{(n)}| < \infty$. For any pseudo-Lipschitz test function $\varphi : \mathbb{R}^{t+1} \to \mathbb{R}$, it holds that

$$\frac{1}{n}\sum_{i\in[n]} \beta_i^{(n)} \left\{\varphi\left(\eta_i, x_i^1, \cdots, x_i^t\right) - \mathbb{E}\left[\varphi\left(\eta_i, Z_i^1 + \mu_1 u_i, \cdots, Z_i^t + \mu_t u_i\right)\right]\right\} \xrightarrow[n\to\infty]{\mathbb{P}} 0\,.$$

This theorem can be seen as a corollary to Theorem 2.1, the proof is provided in Appendix 1.

2.7 Open Questions

(1) Currently, the sparsity level is of order $\log^{\nu\vee 1}(n)$. Would it be possible to lower this level, and to dissociate the sparsity assumption from the parameter ν which is associated to the moments of the matrix entries?
(2) In Sect. 2.5 (Example 3), we discussed the example of an AMP algorithm for sparse d-regular matrices and noted that the AMP theory requires $d \to \infty$ for such matrices. An interesting open question is whether these results can be extended to the constant-degree regime, i.e., to d-regular matrices with finite d, although it would be unlikely to get Gaussian limits in this case (the limit might not be universal and may depend upon d). More generally, one may ask whether it is possible to design AMP algorithms for sparse matrices with a constant level of sparsity. A promising direction could be to rely on non-asymptotic AMP results (cf. [5, 22]), and establish bounds explicitly depending on d.
(3) Would it be possible to improve the convergence in probability in Theorem 2.1 to an almost sure convergence?
(4) Our current assumptions over the entries of the matrix necessitate all the moments. Would it be possible by truncation techniques to lower this assumption?
(5) Would it be possible to establish the counterpart of Theorem 2.1 for AMP schemes (11) or (12)?

2.8 Outline of the Proof

Building on the methods developed in [7] and [21], we start by analyzing a particular case of the approximate message passing (AMP) algorithm with polynomial activation functions (Sect. 3.1), which motivates the adoption of combinatorial techniques. In our setting, the variance profile is non-symmetric, and the matrix contains correlations between symmetric entries, necessitating modifications to the combinatorial approaches used in both [7] and [21] to fit our case. The combinatorial heart of the proof is presented in Sect. 4. We then use density arguments to extend the results to non-polynomial activation functions that exhibit at most polynomial growth (Sect. 5.1).

It should be noted that the combinatorial methods in [7] and [21] rely on the assumption of a zero-diagonal variance profile, i.e., $S_{ii} = 0$ for all $i \in [n]$, which simplifies the derivations. We adopt this assumption in Sects. 3.1, 3.2 and 5.1 and then lift it via a perturbation argument in Sect. 5.2. Unless otherwise specified, we assume that the

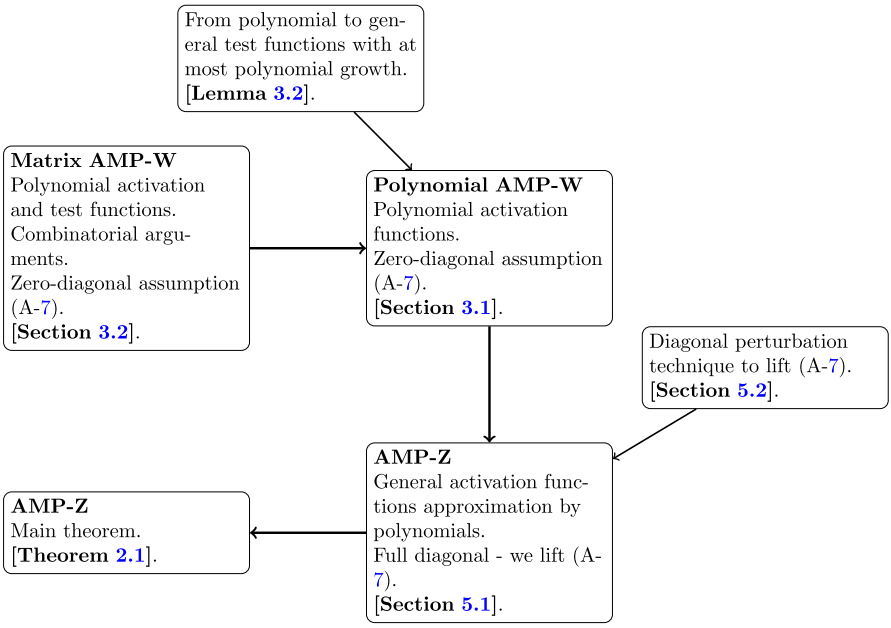


Fig. 2 Proof steps.

matrix S has a zero-diagonal, implying, without loss of generality, that the random matrix X also has a zero diagonal $X_{ii} = 0$ (Fig. 2).

3 AMP and Matrix AMP for Polynomial Activation Functions

We present hereafter the AMP algorithm for polynomial activation functions, a suitable framework to establish the proof by combinatorial techniques, see [7, 21]. In Sect. 3.1, we state Theorem 3.1 for iterates that are $\mathbb{R}^n$-valued.

In Sect. 3.2, we state a result for iterates that are $\mathbb{R}^{n\times q}$-valued, a more general result that will imply Theorem 3.1. The extension to general pseudo-Lipschitz functions will be performed in Sect. 5.1.

The following technical assumption (to be lifted in Sect. 5.2) will be used hereafter.

Assumption A- 7 *(variance profile with vanishing diagonal) The deterministic $n \times n$ matrix $S = (s_{ij})_{1\le i,j\le n}$ has non-negative elements with null elements on the diagonal:*

$$S_{ii} = 0 \quad \textit{for} \quad i \in [n]\,.$$

Remark 3.1 Assumption A-7 is very convenient to establish the statistical properties of the AMP iterates for polynomial activation functions, as the proof relies on combinatorial techniques. The fact that the diagonal of the variance profile S is zero substantially simplifies the combinatorics. This assumption is relaxed in Theorem 2.1 by means of perturbation arguments (see Sect. 5.2).

3.1 AMP for Polynomial Activation Functions

Let $d \geq 1$ be a fixed positive integer independent from n. For every integer $t \geq 1$, consider a uniformly bounded triangular array of real coefficients

$$\Big(\alpha_\ell(i,t,n)\,,\ \ell \leq d\,,\ i \in [n]\,,\ n \geq 1\Big) \qquad \text{with} \qquad \sup_n \max_{\ell \leq d} \max_{i \in [n]} |\alpha_\ell(i,t,n)| < \infty\,. \tag{16}$$

The following function will play a key role in the sequel:

$$\begin{aligned} p : \mathbb{R} \times [n] \times \mathbb{N} &\to \mathbb{R}, \\ (u,i,t) \quad &\mapsto p(u,i,t) = \sum_{\ell=1}^{d} \alpha_\ell(i,t,n) u^\ell. \end{aligned} \tag{17}$$

Function p is a polynomial in u with degree bounded by d. It depends on n via the coefficients $\alpha_\ell(i,t,n)$. To lighten the notations, we drop the dependence of $\alpha_\ell(i,t,n)$ in n and simply write $\alpha_\ell(i,t)$ and do not indicate the dependence of p in n.

Following Definition (11), let $\check{\boldsymbol{x}}^0 \in \mathbb{R}^n$ be deterministic and define

$$\big(\check{\boldsymbol{x}}^t\big)_{t \geq 1} = \text{AMP-W}\left(X, S, p, \check{\boldsymbol{x}}^0\right), \quad \check{\boldsymbol{x}}^0 \in \mathbb{R}^n\,,$$

that is

$$\check{\boldsymbol{x}}^{t+1} = W p(\check{\boldsymbol{x}}^t, \cdot, t) - \text{diag}\Big(W \odot W^\top \partial p(\check{\boldsymbol{x}}^t, \cdot, t)\Big) p(\check{\boldsymbol{x}}^{t-1}, \cdot, t-1), \tag{18}$$

where $p(\boldsymbol{x}, \cdot, t) = [p(x_i, i, t)]_{i=1}^n$ and $\partial p(\boldsymbol{x}, \cdot, t) = [\partial p(x_i, i, t)]_{i=1}^n$ for any $\boldsymbol{x} \in \mathbb{R}^n$.

We now present the AMP result for polynomial activation functions.

Theorem 3.1 *Let A-1, A-2 and A-7 hold true. Let $d \geq 1$ be fixed, (α_ℓ) and p given by (16) and (17). Let $\check{\boldsymbol{x}}^0 = (\check{x}_i^0) \in \mathbb{R}^n$. Assume that there exists a compact set $\mathcal{Q}_{\check{\boldsymbol{x}}} \subset \mathbb{R}$ such that $\check{x}_i^0 \in \mathcal{Q}_{\check{\boldsymbol{x}}}$. Consider*

$$\big(\check{\boldsymbol{x}}^t\big)_{t \geq 1} = \text{AMP-W}\left(X, S, p, \check{\boldsymbol{x}}^0\right).$$

Let $(\check{\boldsymbol{Z}}^1, \cdots, \check{\boldsymbol{Z}}^t) \sim \text{DE}(S, p, \check{\boldsymbol{x}}^0, t)$ and denote by $\check{R}_i^t$ the covariance matrix of vector $\left(\check{Z}_i^1, \cdots, \check{Z}_i^t\right)$. Then for all $t, m \geq 1$

$$\sup_n \max_{i \in [n]} \|\check{R}_i^t\| < \infty \qquad \textit{and} \qquad \sup_n \max_{i \in [n]} \mathbb{E}|\check{x}_i^t|^m < \infty\,. \tag{19}$$

Given $t \geq 1$, let $d' \geq 1$ be fixed and consider function $\psi_n : \mathbb{R}^t \times [n] \to \mathbb{R}$, a multivariate polynomial with bounded degree:

$$\psi_n(x_1, \cdots, x_t, \ell) = \sum_{d_1 + \cdots + d_t \leq d'} \beta_n(d_1, \cdots, d_t, \ell) \prod_{i \in [t]} x_i^{d_i}\,,$$

with

$$\sup_{n\geq 1}\sup_{\ell\in[n]}\sup_{d_1+\cdots+d_t\leq d'}|\beta_n(d_1,\cdots,d_t,\ell)|<\infty\,.$$

Let $\mathcal{S}^{(n)}\subset[n]$ be such that $|\mathcal{S}^{(n)}|\leq CK_n$ where K_n is given by A-2. Then,

$$\frac{1}{K_n}\sum_{i\in\mathcal{S}^{(n)}}\left\{\psi_n(\check{x}_i^1,\ldots,\check{x}_i^t,i)-\mathbb{E}\psi_n(\check{Z}_i^1,\ldots,\check{Z}_i^t,i)\right\}\xrightarrow[n\to\infty]{\mathbb{P}}0\,,\quad\textit{and}\tag{20a}$$

$$\frac{1}{n}\sum_{i\in[n]}\left\{\psi_n(\check{x}_i^1,\ldots,\check{x}_i^t,i)-\mathbb{E}\psi_n(\check{Z}_i^1,\ldots,\check{Z}_i^t,i)\right\}\xrightarrow[n\to\infty]{\mathbb{P}}0\,.\tag{20b}$$

Remark 3.2 In this theorem, both the activation function and the test function used in the convergence formulation are polynomials. The general case for the activation function will be addressed later in Sect. 5.1. Regarding the test functions, we extend this result in the following lemma to encompass general continuous functions that grow at most polynomially near infinity. Notice also that Assumption A-3 is not needed when dealing with AMP sequences having polynomial activation functions, this assumption is purely technical and is used when a comparison between two AMP sequences is provided.

Remark 3.3 The interesting regime in (20a) is $|\mathcal{S}^{(n)}|\sim K_n$. If $|\mathcal{S}^{(n)}|\ll K_n$ then (20a) is trivial in the sense that one can easily prove that both terms

$$\frac{1}{K_n}\sum_{i\in\mathcal{S}^{(n)}}\psi_n\left(\check{x}_i^1,\cdots,\check{x}_i^t,i\right)\qquad\text{and}\qquad\frac{1}{K_n}\sum_{i\in\mathcal{S}^{(n)}}\mathbb{E}\left[\psi_n\left(\check{Z}_i^1,\cdots,\check{Z}_i^t,i\right)\right]$$

converge to zero[2].

Lemma 3.2 *Let $\check{\boldsymbol{x}}^0$ and $\boldsymbol{\eta}$ satisfy A-4. Let $(\check{\boldsymbol{x}}^t)_{t\in\mathbb{N}}$ and $(\check{\boldsymbol{Z}}^t)_{t\in\mathbb{N}}$ as in Theorem 3.1. Let $t,m\geq 0$ be fixed integers and let $\varphi:\mathcal{Q}_\eta\times\mathbb{R}^t\to\mathbb{R}$ be a continuous function such that*

$$|\varphi(\alpha,u_1,\cdots,u_t)|\leq C\left(1+|u_1|^m+\cdots+|u_t|^m\right)\,.$$

For any sequence $(\beta_i^{(n)}\in\mathbb{R},\ i\in[n],\ n\geq 1)$ such that $\sup_n\max_{i\in[n]}|\beta_i^{(n)}|<\infty$, the following convergence holds:

$$\frac{1}{n}\sum_{i\in[n]}\beta_i^{(n)}\varphi(\eta_i,\check{x}_i^1,\ldots,\check{x}_i^t)-\frac{1}{n}\sum_{i\in[n]}\beta_i^{(n)}\mathbb{E}\varphi(\eta_i,\check{Z}_i^1,\ldots,\check{Z}_i^t)\xrightarrow[n\to\infty]{\mathbb{P}}0.$$

Proof Define the two $t+2$ dimensional random measures μ_n and ν_n as follows

$$\mu_n=\frac{1}{n}\sum_{i\in[n]}\delta_{(\beta_i,\eta_i,\check{x}_i^1,\cdots,\check{x}_i^t)}\quad\text{and}\quad\nu_n=\mathcal{L}\left(\beta_\theta,\eta_\theta,\check{Z}_\theta^1,\cdots,\check{Z}_\theta^t\right),$$

[2] By $|\mathcal{S}^{(n)}|\sim K_n$, we mean that there exist $c,C>0$ such that $cK_n\leq|\mathcal{S}^{(n)}|\leq CK_n$ and by $|\mathcal{S}^{(n)}|\ll K_n$, we mean that $|\mathcal{S}^{(n)}|/K_n\to 0$.

where $\theta \sim \mathcal{U}([n])$ is independent. Consider the function $\psi(\beta, \eta, x_1, \cdots, x_t) = \beta\varphi(\eta, x_1, \cdots, x_t)$, and recall that (β_i), (η_i) and the covariance matrices (R_i^t) are bounded, thus by some slight modification to Lemma B.1 we get the desired result. □

3.2 Matrix AMP for Polynomial Activation Functions

In order to prove Theorem 3.1, we need to study a matrix version of the AMP algorithm where the iterates $\check{\boldsymbol{x}}^t$ are $\mathbb{R}^{n\times q}$–valued matrices, $q \geq 1$ being a fixed integer. Using this framework, we only need to express the convergence result in Theorem 3.1 using test functions acting only on the t^{th} iterates instead of all previous iterates. Consider the function

$$f : \mathbb{R}^q \times [n] \times \mathbb{N} \longrightarrow \mathbb{R}^q, \qquad f(\boldsymbol{u}, l, t) = \begin{pmatrix} f_1(\boldsymbol{u}, l, t) \\ \vdots \\ f_q(\boldsymbol{u}, l, t) \end{pmatrix}, \tag{21}$$

where each component f_r is a polynomial in $\boldsymbol{u} \in \mathbb{R}^q$, with degree bounded by d, written as

$$f_r(\boldsymbol{u}, \ell, t) = \sum_{\substack{\boldsymbol{i}=(i_1,\cdots,i_q)\in\mathbb{N}^q \\ i_1+\cdots+i_q\leq d}} \alpha_{\boldsymbol{i}}(r, \ell, t)\boldsymbol{u}^{\boldsymbol{i}},$$

(recall the notation $\boldsymbol{u}^{\boldsymbol{i}} = \prod_{s\in[q]} u_s^{i_s}$). Given a deterministic n-uple $(\boldsymbol{x}_1^0, \ldots, \boldsymbol{x}_n^0)$ where $\boldsymbol{x}_i^0$ is a q-dimensional vector, the AMP iterates are recursively defined for $t \geq 1$ as follows:

$$x_i^{t+1}(r) = \sum_{\ell\in[n]} W_{i\ell} f_r(\boldsymbol{x}_\ell^t, \ell, t) - \sum_{s\in[q]} f_s(\boldsymbol{x}_i^{t-1}, i, t-1) \sum_{\ell\in[n]} W_{i\ell}W_{\ell i} \frac{\partial f_r}{\partial x(s)}(\boldsymbol{x}_\ell^t, \ell, t), \tag{22}$$

for $r \in [q]$ and $f(\cdot, \cdot, -1) \equiv 0$. We denote such a sequence by

$$\left(\boldsymbol{x}^t\right)_{t\geq 1} = \text{AMP-W}_q\left(X, S, f, \boldsymbol{x}^0\right), \quad \boldsymbol{x}^0 \in \mathbb{R}^{n\times q}.$$

DE Equations for Matrix AMP

Similarly to the DE equations for standard AMP introduced in Definition 1.3, we introduce here a $(\mathbb{R}^q)^n$-valued sequence of Gaussian random vectors $(U^t)_{t\in\mathbb{N}^\star}$ defined by

$$U^t = \begin{bmatrix} (U_1^t)^\top \\ \vdots \\ (U_n^t)^\top \end{bmatrix},$$

where $\{U_i^t\}_{i\in[n]}$ are $\mathbb{R}^q$-valued independent Gaussian random vectors, $U_i^t \sim \mathcal{N}\left(0, Q_i^t\right)$ and the $q \times q$ matrices Q_i^t are defined recursively in t by

$$Q_i^{t+1} = \sum_{\ell\in[n]} s_{i\ell}\mathbb{E} f(U_\ell^t, \ell, t) f(U_\ell^t, \ell, t)^\top \quad \text{for} \quad i \in [n], \tag{23}$$

with the convention that $U^0 := \boldsymbol{x}^0$. We denote

$$U^t \sim \mathrm{DE}_q\left(S, f, \boldsymbol{x}^0, t\right). \tag{24}$$

The following Theorem is the key component to the proof of Theorem 3.1.

Theorem 3.3 *Let Assumptions A-1 and A-2 hold true and $q \geq 1$ be fixed. Let f be defined by (21) and $\boldsymbol{x}^0 \in \mathbb{R}^{n\times q}$. Assume that for each $t \geq 1$, there exists a constant $C = C(t) > 0$ such that*

$$\left|\alpha_{i_1,\ldots,i_q}(r, l, t)\right| \leq C, \quad \text{and} \quad \sup_n \max_{i\in[n]} \left\|\boldsymbol{x}_i^0\right\| < \infty\,. \tag{25}$$

Consider the iterative algorithm $\left(\boldsymbol{x}^t\right)_{t\geq 1} = \text{AMP-W}_q\left(X, S, f, \boldsymbol{x}^0\right)$, and let Q_i^t and U^t be defined by (23)–(24). Then we have,

$$\forall t > 0\,, \quad \sup_n \max_{i\in[n]} \|Q_i^t\| < \infty. \tag{26}$$

Moreover,

$$\forall t > 0\,,\ \forall \boldsymbol{m} \in \mathbb{N}^q\,, \quad \sup_n \max_{i\in[n]} \mathbb{E}|(\boldsymbol{x}_i^t)^{\boldsymbol{m}}| < \infty. \tag{27}$$

Let $\psi : \mathbb{R}^q \times [n] \to \mathbb{R}$ be such that $\psi(\cdot, l)$ is a multivariate polynomial with a bounded degree and bounded coefficients as functions of (l, n). Let $\mathcal{S}^{(n)} \subset [n]$ be a non-empty set such that $|\mathcal{S}^{(n)}| \leq C K_n$. Then,

$$\frac{1}{K_n}\sum_{i\in\mathcal{S}^{(n)}} \psi(\boldsymbol{x}_i^t, i) - \mathbb{E}\psi(U_i^t, i) \xrightarrow[n\to\infty]{\mathbb{P}} 0 \quad \text{and} \tag{28a}$$

$$\frac{1}{n}\sum_{i\in[n]} \psi(\boldsymbol{x}_i^t, i) - \mathbb{E}\psi(U_i^t, i) \xrightarrow[n\to\infty]{\mathbb{P}} 0\,. \tag{28b}$$

Remark 3.4 In this theorem, and particularly in the convergence described in (28b), the result is not explicitly stated for all iterations from 1 to t, as was done in (20b). Consequently, Matrix AMP can be interpreted as a more compact formulation of the "standard" AMP. This distinction is further elucidated in the subsequent proof.

Proof of Theorem 3.1 Theorem 3.1 can be deduced from Theorem 3.3 by adequately choosing q as well as a precise construction of the activation function f using the

$\mathbb{R}$-valued polynomials p. Define the sequence $\left(\check{\boldsymbol{x}}^t\right)_{t\geq 1}$ as follows,

$$\left(\check{\boldsymbol{x}}^t\right)_{t\geq 1} = \text{AMP-W}\left(X, p, \check{\boldsymbol{x}}^0, S\right). \tag{29}$$

We shall establish the convergence (20b) for each t and prove that for all multivariate polynomials ψ we have

$$\frac{1}{n}\sum_{i\in[n]}\left\{\psi(\check{x}_i^1,\ldots,\check{x}_i^t,i) - \mathbb{E}\psi(\check{Z}_i^1,\ldots,\check{Z}_i^t,i)\right\} \xrightarrow[n\to\infty]{\mathbb{P}} 0\,.$$

where $(\check{\boldsymbol{Z}}^1,\cdots,\check{\boldsymbol{Z}}^t) \sim \text{DE}(S,p,\check{\boldsymbol{x}}^0,t)$. To this end, let $\tau > 0$ be fixed and chose $q=\tau$, construct the sequence $(\boldsymbol{x}^t)_{1\leq t\leq\tau}$ of $\mathbb{R}^{\tau\times\tau}$-valued matrices such that

$$\begin{aligned}
\boldsymbol{x}_i^1 &= \left(\check{x}_i^1\ 0\cdots 0\right)^\top,\\
\boldsymbol{x}_i^2 &= \left(\check{x}_i^1\ \check{x}_i^2\cdots 0\right)^\top,\\
&\vdots\\
\boldsymbol{x}_i^\tau &= \left(\check{x}_i^1\ \check{x}_i^2\cdots\check{x}_i^\tau\right)^\top.
\end{aligned}$$

Now using the polynomials p, we construct the function $f:\mathbb{R}^\tau\times[n]\times\mathbb{N}\to\mathbb{R}^\tau$ such that for all $i\in[n]$ and $0\leq\ell\leq\tau-1$ we have

$$f(\boldsymbol{x},i,\ell) = \left(p(x_i^0,i,0)\ p(\boldsymbol{x}_i(1),i,1)\cdots p(\boldsymbol{x}_i(\ell),i,\ell)\ 0\cdots 0\right)^\top.$$

For $\ell\geq\tau$, we set

$$f(\boldsymbol{x},i,\ell) = (0\ \cdots\ 0)\,.$$

In order to apply Theorem 3.3, we show that the sequence $(\boldsymbol{x}^t)$ is given by

$$(\boldsymbol{x}^t)_{t\geq 1} = \text{AMP-W}_\tau\left(X, S, f, \boldsymbol{x}^0\right). \tag{30}$$

Let $t\in[\tau-1]$. By definition, for $r\in[\tau]$ and $i\in[n]$ we have

$$x_i^{t+1}(r) = \begin{cases}\check{x}_i^r & \text{if } r\leq t+1\,,\\ 0 & \text{if } r>t+1\,.\end{cases}$$

In addition, by Eq. (29) we know that

$$\check{x}_i^r = \sum_{\ell\in[n]} W_{i\ell}\,p(\check{x}_\ell^{r-1},\ell,r-1) - \sum_{\ell\in[n]} W_{i\ell}W_{\ell i}\,\partial p(\check{x}_\ell^{r-1},\ell,r-1)p(\check{x}_i^{r-2},i,r-2),$$

which implies that for $r\leq\tau+1$,

$$
\begin{aligned}
x_i^{\tau+1}(r) &= \sum_{\ell\in[n]} W_{i\ell}\, p(x_\ell^\tau(r-1), \ell, r-1) \\
&\quad - \sum_{\ell\in[n]} W_{i\ell}W_{\ell i}\,\partial p(x_\ell^\tau(r-1), \ell, r-1)\, p(x_i^{\tau-1}(r-2), i, r-2)\,, \\
&= \sum_{\ell\in[n]} W_{i\ell}\, p(x_\ell^\tau(r-1), \ell, r-1) \\
&\quad - \sum_{s\in[t]} p(x_i^{\tau-1}(s-1), i, s-1) \sum_{\ell\in[n]} W_{i\ell}W_{\ell i}\,\partial p(x_\ell^\tau(r-1), \ell, r-1)\delta_{s,r-1}\,, \\
&= \sum_{\ell\in[n]} W_{i\ell}\, f_r(x_\ell^\tau, \ell, \tau) - \sum_{s\in[t]} f_s(x_i^{\tau-1}, i, \tau-1) \sum_{\ell\in[n]} W_{i\ell}W_{\ell i}\,\frac{\partial}{\partial x(s)} f_r(x_\ell^\tau, \ell, \tau)\,,
\end{aligned}
$$

which is precisely the recursion in (30).

We can now apply the result of Theorem 3.3 to the sequence $(\boldsymbol{x}^t)$, which implies that for all polynomial test functions $\psi(., \ell) : \mathbb{R}^\tau \to \mathbb{R}$ we have

$$
\frac{1}{n}\sum_{i\in[n]} \psi(\boldsymbol{x}_i^\tau, i) - \mathbb{E}\psi(U_i^\tau, i) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \forall\tau\in\mathbb{N}\,,
$$

which yields

$$
\frac{1}{n}\sum_{i\in[n]} \psi(\check{x}_i^1, \cdots, \check{x}_i^\tau, i) - \mathbb{E}\psi(U_i^\tau, i) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \forall\tau\in\mathbb{N}\,, \tag{31}
$$

where the U^τ is $(n\times\tau)$-dimensional random matrix with law $\mathrm{DE}_q(S, f, \boldsymbol{x}^0, \tau)$, the latter is defined in (24). Denote the columns of U^τ by $Z^1, \cdots, Z^\tau \in \mathbb{R}^n$, then it is clear that $(Z^1, \cdots, Z^\tau) \sim \mathrm{DE}(p, \check{x}^0, S, \tau)$. The convergence in (31) becomes

$$
\frac{1}{n}\sum_{i\in[n]} \psi(\check{x}_i^1, \cdots, \check{x}_i^\tau, i) - \mathbb{E}\psi(Z_i^1, \cdots, Z_i^\tau, i) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \forall\tau\in\mathbb{N}\,.
$$

with $(Z^1, \cdots, Z^\tau) \sim \mathrm{DE}(S, p, \check{x}^0, \tau)$. Convergence (20b) is established. One can prove similarly (20a), which concludes the proof of Theorem 3.1.

4 Proof of Theorem 3.3: A Combinatorial Approach

Taking polynomial activation functions in Theorem 3.3 is fundamental, as all iterations $\boldsymbol{x}^t$ can be written as multinomials on the entries of the matrix W and the initial point's coordinates $x_i^0(s)$. This makes the analysis purely combinatorial. At the first and second iterations $t = 1, 2$, and given simple polynomial activation functions

$f_r(u, \ell, 1) = f_r(u, \ell, 0) = u(1)^m$, one can write

$$x_i^1(r) = \sum_{\ell \in [n]} W_{i\ell} x_\ell^0(1)^m,$$

$$x_i^2(r) = \sum_{\ell, \ell_1, \cdots, \ell_m \in [n]} W_{i\ell} W_{i\ell_1} \cdots W_{i\ell_m} \left(x_{\ell_1}^0(1) \cdots x_{\ell_m}^0(1) \right)^m - \{\text{Onsager}\}.$$

We already notice that by the second iteration $t = 2$, the exact expression for x_i^2 as a multinomial expansion in terms of the entries of matrix W becomes increasingly complex. We hence need to find an alternative indexation scheme for the summation above, properly suited to extract the desired information and establish Theorem 3.3. We follow the combinatorial approach initiated in [7]. This approach is based on the introduction of "non-backtracking" trees associated to "non-backtracking" iterations.

4.1 Strategy of Proof

To prove that the AMP iterations have the simple deterministic equivalent described in Theorem 3.3 we first approximate the moments of $\boldsymbol{x}^t \in \mathbb{R}^{n \times q}$ with the moments of simpler objects $\boldsymbol{z}^t$ called the "non-backtracking" iterations, these are generated with the same matrix W used in the recursion (8), with a slightly different recursion scheme where the Onsager term is removed.

$$\mathbb{E}(\boldsymbol{x}_i^t)^{\boldsymbol{m}} \approx \mathbb{E}(\boldsymbol{z}_i^t)^{\boldsymbol{m}}, \quad \forall \boldsymbol{m} \in \mathbb{N}^q,$$

this is done in (Proposition 4.5) sect. 4.5. We then show a universality property of the iterations $\boldsymbol{z}^t$ in (Proposition 4.2) sect. 4.3. More specifically, we show that if $\tilde{\boldsymbol{z}}^t$ is another non-backtracking iteration sequence generated using another matrix $\tilde{W}$ satisfying the same assumptions as W but does not have the same distribution, then

$$\mathbb{E}(\boldsymbol{z}_i^t)^{\boldsymbol{m}} \approx \mathbb{E}(\tilde{\boldsymbol{z}}_i^t)^{\boldsymbol{m}}, \quad \forall \boldsymbol{m} \in \mathbb{N}^q.$$

This means that we can reduce our problem to an AMP constructed using a Gaussian matrix. Hence, without loss of generality we can suppose that W is Gaussian. Moreover, we approximate the non-backtracking iterations $\boldsymbol{z}^t$ with another non-backtracking iterations $\boldsymbol{y}^t$, but this time, in the recursion formula of $\boldsymbol{y}^t$, at each step t we independently pick a new random matrix $W^t \overset{\mathcal{L}}{=} W$ which is Gaussian,

$$\mathbb{E}(\tilde{\boldsymbol{z}}_i^t)^{\boldsymbol{m}} \approx \mathbb{E}(\boldsymbol{y}_i^t)^{\boldsymbol{m}}, \quad \forall \boldsymbol{m} \in \mathbb{N}^q.$$

this is done in (Proposition 4.4) sect. 4.4. $\boldsymbol{x}^t$ is now reduced to its simplest form $\boldsymbol{y}^t$. Finally, we show in (Proposition 4.7) sect. 4.6 that

$$\mathbb{E}(\boldsymbol{y}_i^t)^{\boldsymbol{m}} \approx \mathbb{E}(U_i^t)^{\boldsymbol{m}}, \quad \forall \boldsymbol{m} \in \mathbb{N}^q.$$

which is relatively easy given that $\boldsymbol{y}^t$ are Gaussian. This finishes the proof of Theorem 3.3.

The proof of all these steps follows the combinatorial approach described in both [7] and [21] and thus we begin by presenting the framework of "non-backtracking" trees in sect. 4.2. Notice that that the key difference between prior research and our approach is that the matrix W is no longer symmetric, and exhibits some correlations between its entries.

4.2 Description of the Tree Structure

The proof of Theorem 3.3 follows a combinatorial approach which aims at studying the moments of the AMP iterates. In order to simplify the expression of these moments, we use *planted and labeled trees* to index the sums in these expressions. We first define *planted trees* and then describe its labeling.

Definition 4.1 (Planted trees) We recall the following definition from graph theory.

- A rooted tree $T = (V(T), E(T))$ at $\circ \in V(T)$, where $V(T)$ and $E(T)$ denote respectively the set of vertices and edges, is said to be panted if the root $\circ$ has degree 1.
- We consider that all the edges are oriented toward the root, we say that $v \in V(T)$ is the parent of u if the edge $(u \to v)$ is in $E(T)$, in this case, we use the notation $\pi(u) = v$, we also say that u is a child of v.
- We denote by $L(T)$ the set of leaves of T, i.e. vertices $v \in V(T)$ with no children.
- Given a vertex $v \in V(T)$, we denote by $|v|$ its distance to the root $\circ$.
- Finally, we define a *path* starting at v_1 and ending at v_k as a sequence of vertices $(v_1, v_2, \cdots, v_k)$ such that $v_i = \pi(v_{i+1})$ for all $i \in [k-1]$.

We fix a integer $d, t \in \mathbb{N}$, throughout this proof we consider the class of planted trees $(T, \circ)$ of depth at most t such that for each vertex v, v can have at most d children.

We denote

$$\mathbb{N}^q_{\leq d} := \{(a_1, \cdots, a_q) \in \mathbb{N}^q \,,\ a_1 + \cdots + a_q \leq d\} \,,$$

where q is also a fixed integer.

Definition 4.2 (Labeled and planted trees) We now describe the labeling of the trees. A labeling of a tree T, is a triplet of functions (ℓ, r, c) such that

$$\ell : V(T) \to [n], \quad r : V(T) \setminus \{\circ\} \to [q], \quad c : L(T) \to \mathbb{N}^q_{\leq d}.$$

- For each vertex $u \in V(T)$, $\ell(u)$ is called the *type* of u.
- For each vertex $u \in V(T)$ except the root, $r(u)$ is called the *mark* of u.
- For each vertex $u \in V(T)$ which is not a leaf, we denote by $u[i]$ the number of children of u that have mark $i \in [q]$. We use the same notation to describe $c(u)$ for $u \in L(T)$; $c(u) = (u[1], \cdots, u[q]) \in \mathbb{N}^q_{\leq d}$. In what follows, this notation is used instead of $c(u)$.

- For a non-maximal leaf $u \in L(T)$, i.e. such that $|u|$ is less than the depth of T, we set $u[1] = \cdots u[q] = 0$.

We denote by $\overline{\mathcal{T}}^t$ the set of planted and labeled trees, with depth t at most.

Non-backtracking Trees

One class of planted and labeled trees that is particularly adapted to our specific study, is the class of trees satisfying the *non-backtracking* condition, we recall here the definition that can be found in [7]. A non-backtracking tree is a planted and labeled tree T such that for each path $(u_1 = \circ, u_2 \cdots, u_k)$ in T the types $(\ell(u_i), \ell(u_{i+1}), \ell(u_{i+2}))$ are distinct for each $i \in [k-2]$. We denote the class of these trees as $\mathcal{T}^t$. In addition, we introduce the following classes of trees, for given integers i, j and r, we denote by,

- $\mathcal{T}^t_{i\to j}(r) \subset \mathcal{T}^t$ the subset of trees in $\mathcal{T}^t$ for which the type of the root is i, the type of the child v of the root satisfies $\ell(v) \notin \{i, j\}$, and the mark of v is $r(v) = r$.
- $\mathcal{T}^t_i(r) \subset \mathcal{T}^t$ the subset of trees in $\mathcal{T}^t$ for which the type of the root is i, the type of the child v of the root satisfies $\ell(v) \neq i$, and the mark of v is $r(v) = r$.

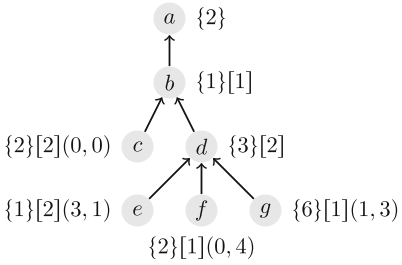
We can already use these trees to create the following objects. For a matrix $W \in \mathbb{R}^{n\times n}$, a vector $x \in \mathbb{R}^n$ and a family of real numbers $\boldsymbol{\alpha} = \left\{\alpha_\iota(r, \ell, s) \mid \iota \in \mathbb{N}^q_{\leq d}, (r, \ell, s) \in [q] \times [n] \times [t]\right\}$, we define,

$$
\begin{aligned}
W(T) &:= \prod_{(u\to v)\in E(T)} W_{\ell(v)\ell(u)} \,, \\
\Gamma(T, \boldsymbol{\alpha}, t) &:= \prod_{(u\to v)\in E(T)} \alpha_{u[1],\ldots,u[q]}\left(r(u), \ell(u), t - |u|\right) \,, \\
x(T) &:= \prod_{v\in L(T)} \prod_{s\in[q]} \left(x_{\ell(v)}(s)\right)^{v[s]} \,.
\end{aligned}
$$

To better illustrate the concepts previously defined, we present a simple example of a tree and demonstrate how it indexes the tree quantities W, Γ, and x (Fig. 3).

4.3 Non-backtracking Iterations

The non-backtracking iterations $(\boldsymbol{z}^t)_t$, are defined recursively similarly to $(\boldsymbol{x}^t)_t$ but minus the Onsager term and with a slight change in the contributing terms from the previous iteration. Recall that the purpose of having the Onsager term is to eliminate components that induce non-Gaussian behavior in the iterates in the high-dimensional regime. Basically, non-backtracking iterations evolve purposefully getting rid of parts that are source non-Gaussian behavior. In particular we do not need to have a corrective term.



$$
\begin{aligned}
W(T) &= W_{21}W_{12}W_{13}W_{31}W_{32}W_{36},\\
\Gamma(T,\boldsymbol{\alpha},3) &= \alpha_{0,2}(1,1,2)\alpha_{0,0}(2,2,1)\alpha_{2,1}(2,3,1)\\
&\quad\times \alpha_{3,1}(2,1,0)\alpha_{0,4}(1,2,0)\alpha_{1,3}(1,6,0),\\
x(T) &= (x_2(1))^0(x_2(2))^0(x_1(1))^3(x_1(2))^1\\
&\quad\times (x_2(1))^0(x_2(2))^4(x_6(1))^1(x_6(2))^3.
\end{aligned}
$$

Fig. 3 Example of a tree $T \in \overline{\mathcal{T}}^3$ for parameters $q = 2$, $d = 4$, $t = 3$ and $n = 6$. The types are written between braces, the marks are between brackets and leafs info is between parentheses. In this example, T is not a non-backtracking tree because of the two paths $(a \leftarrow b \leftarrow c)$ and $(b \leftarrow d \leftarrow e)$.

Given any $i, j \in [n]$ with $i \neq j$, we initialize the non-backtracking sequence with $\boldsymbol{z}^0_{i\to j} := \boldsymbol{x}^0_i$. We then define recursively $\boldsymbol{z}^{t+1}_{i\to j}$ using the previous iterations as follows

$$
z^{t+1}_{i\to j}(r) = \sum_{\ell\in[n]\setminus\{j\}} W_{i\ell} f_r(\boldsymbol{z}^t_{\ell\to i}, \ell, t), \quad \forall r \in [q], \tag{32}
$$

the case $l = i$ is excluded because $W_{ii} = 0$. In addition, we also define the vectors $(\boldsymbol{z}^t)_t$ by

$$
z^{t+1}_{i}(r) = \sum_{\ell\in[n]} W_{i\ell} f_r(\boldsymbol{z}^t_{\ell\to i}, \ell, t), \quad \forall r \in [q]. \tag{33}
$$

We provide here a non-recursive formulation of $z^t_{i\to j}$ and z^t_i described as sums indexed by trees in $\mathcal{T}^t_{i\to j}(r)$ and $\mathcal{T}^t_i(r)$.

Lemma 4.1 *(Lemma 1 of [7]) For all integers $t \in \mathbb{N}$, $i, j \in [n]$ and $r \in [q]$, we have,*

$$
\begin{aligned}
z^t_{i\to j}(r) &= \sum_{T\in\mathcal{T}^t_{i\to j}(r)} W(T)\Gamma(T,\boldsymbol{\alpha},t)x(T),\\
z^t_{i}(r) &= \sum_{T\in\mathcal{T}^t_{i}(r)} W(T)\Gamma(T,\boldsymbol{\alpha},t)x(T).
\end{aligned}
$$

Here $x(T) := \boldsymbol{x}^0(T)$, we drop the superscript from this notation.

Note that this lemma is purely structural, the proof is not impacted by our specific variance and correlation profiles.

To simplify the notations in the following proofs we introduce the following sets,

$$
\mathcal{K} = \left\{(i,j) \in [n]\times[n]\,,\ \ s_{ij} > 0\right\} \quad \text{and} \quad \mathcal{C} = \left\{(i,j) \in [n]\times[n]\,,\ \ \tau_{ij} \neq 0\right\}. \tag{34}
$$

We also define the row and column sections of $\mathcal{K}$,

$$
\mathcal{K}_i = \left\{j \in [n]\,,\ \ s_{ij} > 0\right\} \quad \text{and} \quad \mathcal{K}^j = \left\{i \in [n]\,,\ \ s_{ij} > 0\right\}. \tag{35}
$$

The next proposition shows that in the large dimensional regime, the moments of a vector z_i^t issued from the non-backtracking iterations depend for large n only on the first two moments of the elements of W.

Proposition 4.2 *(adaptation of Proposition 1 of [7]) Let $\widetilde{X}$ be a random matrix satisfying A-1, with distribution not necessarily identical to its analogue X. Assume that W fulfills A-2. Let $\widetilde{W}$ be the matrix constructed similarly to W, but with the X_{ij} replaced with the $\widetilde{X}_{ij}$. Starting with the set of $\mathbb{R}^q$–valued vectors $\{\tilde{z}^0_{i\to j},\ i,j\in[n],\ i\neq j\}$ given as $\tilde{z}^0_{i\to j}=\boldsymbol{x}_i^0$, define the vectors $\tilde{z}_i^t\in\mathbb{R}^q$ by the recursion (32) and the equation (33), where W is replaced with $\widetilde{W}$. Then, for each $t\geq 1$ and each $\boldsymbol{m}\in\mathbb{N}^q$,*

$$\left|\mathbb{E}(z_i^t)^{\boldsymbol{m}}-\mathbb{E}(\tilde{z}_i^t)^{\boldsymbol{m}}\right|=\mathcal{O}\left(\frac{1}{\sqrt{K_n}}\right).$$

Proof For simplicity and clarity of the arguments presented we restrict the proof to the case where the multi-index $\boldsymbol{m}$ satisfies

$$\boldsymbol{m}(s)=\begin{cases}0 & \text{if } s\neq r,\\ m & \text{if } s=r,\end{cases}$$

for some integer $m>0$. The proof for a general multi-index $\boldsymbol{m}\in\mathbb{N}^q$ is very similar to that of the single-valued case as long as q is finite, and is thus omitted.

By Lemma 4.1, we have

$$\mathbb{E}(z_i^t(r))^m=\sum_{T_1,\dots,T_m\in\mathcal{T}_i^t(r)}\left(\prod_{k=1}^m\Gamma(T_k,\boldsymbol{\alpha},t)\right)\mathbb{E}\left[\prod_{k=1}^m W(T_k)\right]\prod_{k=1}^m x(T_k).$$

For a tree T and $j,\ell\in[n]$, define

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

Based on the definition of $W(T)$, $\vec{\varphi}_{\ell j}(T)$ counts the number of edges in the tree T that represent the (ℓ,j) matrix entry $W_{\ell j}$. We also define $\varphi_{j\ell}$ for $j<\ell$ as

$$\varphi_{j\ell}(T)=\vec{\varphi}_{j\ell}(T)+\vec{\varphi}_{\ell j}(T),$$

this quantity represents the total number of edges in the tree T that represent either $W_{j\ell}$ or $W_{\ell j}$. We know that there is an integer constant $C_E=C_E(d,t,m)$ that bounds the total number of edges in the trees $T_1,\dots,T_m\in\mathcal{T}_i^t(r)$, thus

$$\sum_{k\in[m]}\sum_{j<\ell}\varphi_{j\ell}(T_k)\leq C_E=m\frac{d^t-1}{d-1}.$$

C_E is simply the maximum number of edges in the m-tuple of trees $T_1, \cdots, T_m$. Given an integer $\mu \in [C_E]$, recall that $\mathcal{K}$ is introduced in (34), define

$$\begin{aligned}\mathcal{A}_i(\mu) := \Big\{ & (T_1, \ldots, T_m)\,,\ T_k \in \mathcal{T}_i^t(r) \text{ for all } k \in [m], \\ & \forall j < \ell,\ \sum_{k\in[m]} \varphi_{j\ell}(T_k) \neq 1, \\ & \forall j, \ell,\ \sum_{k\in[m]} \vec{\varphi}_{j\ell}(T_k) > 0 \ \Rightarrow (j,l) \in \mathcal{K}, \\ & \sum_{k\in[m]} \sum_{j<\ell} \varphi_{j\ell}(T_k) = \mu \Big\}.\end{aligned}$$

Since the elements of W beneath the diagonal are centered and independent, then,

$$\mathbb{E}(z_i^t(r))^m = \sum_{\mu=1}^{C_E} \sum_{(T_1,\ldots,T_m)\in\mathcal{A}_i(\mu)} \left(\prod_{k=1}^m \Gamma(T_k, \boldsymbol{\alpha}, t)\right) \left(\prod_{k=1}^m x(T_k)\right) \mathbb{E}\left[\prod_{k=1}^m W(T_k)\right]. \tag{36}$$

Notice that the contributions of the m–uples of trees in the set

$$\left\{ (T_1, \cdots, T_m) \in \mathcal{A}_i(\mu)\,,\ \forall j < \ell,\ \sum_{k\in[m]} \varphi(T_k)_{j\ell} \in \{0, 2\} \right\},$$

are the same for $\mathbb{E}(z_i^t(r))^m$ and $\mathbb{E}(\widetilde{z}_i^t(r))^m$ by the assumptions on the matrices W and $\widetilde{W}$. Three cases can be considered for a couple of indices (j, ℓ) where $j < \ell$ and $\sum_{k\in[m]} \vec{\varphi}(T_k)_{j\ell} = 2$,

- $W_{j\ell}$ is represented two times in the trees $\Rightarrow$ contribution equal to $s_{j\ell}$,
- $W_{\ell j}$ is represented two times in the trees $\Rightarrow$ contribution equal to $s_{\ell j}$,
- $W_{j\ell}$ and $W_{\ell j}$ are both represented in the trees $\Rightarrow$ contribution equal to $\sqrt{s_{j\ell}s_{\ell j}}\tau_{j\ell}$.

Notice that in all three cases the contributions do not depend on the distributions of the entries of the matrix W but only on the first and second moments. Thus, defining the set

$$\check{\mathcal{A}}_i(\mu) = \left\{ (T_1, \ldots, T_m) \in \mathcal{A}_i(\mu)\,,\ \exists j < \ell,\ \sum_{k\in[m]} \vec{\varphi}(T_k)_{j\ell} \geq 3 \right\}, \tag{37}$$

the proposition can be proven if we prove that for all $\mu \in [C_E]$, the real number

$$\xi_\mu = \sum_{(T_1,\ldots,T_m)\in\check{\mathcal{A}}_i(\mu)} \left(\prod_{k=1}^m \Gamma(T_k, \boldsymbol{\alpha}, t)\right) \left(\prod_{k=1}^m x(T_k)\right) \mathbb{E}\left[\prod_{k=1}^m W(T_k)\right]$$

satisfies

$$|\xi_\mu| = \mathcal{O}\left(\frac{1}{\sqrt{K_n}}\right).$$

Using the bounds (25) provided in the statement of Theorem 3.3, it is clear that $\prod_{k=1}^{m} \Gamma(T_k, \boldsymbol{\alpha}, t)$ and $\prod_{k=1}^{m} x(T_k)$ are bounded as n goes to infinity.

Since there exists a constant C such that $|\mathbb{E}W_{j\ell}^{s}| \leq C K_n^{-s/2}$ for each integer $s > 0$ by A-1 and A-2, for each $(T_1, \ldots, T_m) \in \breve{\mathcal{A}}_i(\mu)$, we have

$$\begin{aligned}\left|\mathbb{E}\prod_{k=1}^{m} W(T_k)\right| &= \prod_{j<\ell}\left|\mathbb{E}W_{j\ell}^{\sum_{k=1}^{m}\vec{\varphi}_{j\ell}(T_k)} W_{\ell j}^{\sum_{k=1}^{m}\vec{\varphi}_{\ell j}(T_k)}\right| \\ &\leq \prod_{j<\ell}\left|\left(\mathbb{E}W_{j\ell}^{2\sum_{k=1}^{m}\vec{\varphi}_{j\ell}(T_k)}\right)^{1/2}\left(\mathbb{E}W_{\ell j}^{2\sum_{k=1}^{m}\vec{\varphi}_{\ell j}(T_k)}\right)^{1/2}\right. \\ &\leq C K_n^{-\frac{1}{2}\sum_{j<\ell}\sum_{k}\vec{\varphi}_{j\ell}(T_k)+\vec{\varphi}_{\ell j}(T_k)} \quad \leq \quad C K_n^{-\mu/2}.\end{aligned}$$

To complete the proof, we shall show that

$$\left|\breve{\mathcal{A}}_i(\mu)\right| = \mathcal{O}\left(K_n^{\frac{\mu-1}{2}}\right). \tag{38}$$

Given an m–uple $(T_1, \ldots, T_m) \in \breve{\mathcal{A}}_i(\mu)$ of trees, we construct a graph $G = \boldsymbol{G}(T_1, \ldots, T_m)$ by identifying the types of the vertices in all these trees (i.e., by merging the vertices of $T_1, \ldots, T_m$ that have the same type). The marks as well as the orientation of the edges are ignored. G is then a rooted and labeled graph whose root is the vertex obtained by merging the roots of the trees $T_1, \ldots, T_m$ (remember that they all have the same type i).
The number of edges of G is

$$|E(G)| = \sum_{j<\ell} \mathbb{1}_{\sum_k \varphi(T_k)_{j\ell}>0}.$$

Remember that when $\sum_k \varphi(T_k)_{j\ell} > 0$, this sum is greater than 2, so

$$\forall j < \ell, \quad \sum_{k} \varphi(T_k)_{j\ell} \geq 2\mathbb{1}_{\sum_k \varphi(T_k)_{j\ell}>0},$$

we also know that for some $j < \ell$ we have $\sum_k \varphi(T_k)_{j\ell} \geq 3$. Consequently,

$$2(|E(G)| - 1) + 3 \quad \leq \quad \sum_{j<\ell}\sum_{k=1}^{m} \varphi(T_k)_{j\ell},$$

thus,

$$|E(G)| \leq \frac{\mu - 1}{2}.$$

Note that since G is connected, as being obtained through the merger of planted trees with the same root's type,

$$|V(G)| \leq |E(G)| + 1,$$

which gives

$$|\{v \in V(G)\,,\ v \neq \circ\}| \leq (\mu - 1)/2.$$

Also, by construction, G satisfies the following property:

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

where $\mathcal{K}_i$ is defined in (35). And by A-2, this implies that G satisfies the following property: for any fixed labeled vertex $v \in V(G)$ if $(u \to v) \in E(G)$ then u can be labeled by at most CK_n different values.

We shall denote as $\mathcal{G}_i^\mu$ the set of rooted, undirected and labeled graphs G such that

- G is connected,
- $\ell(\circ) = i$, $|E(G)| \leq (\mu - 1)/2$,
- for any fixed labeled vertex $v \in V(G)$ if $(u, v) \in E(G)$ then u can be labeled by at most CK_n different values.

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

$$\left|\breve{\mathcal{A}}_i(\mu)\right| = \sum_{\bar{G} \in \mathcal{R}^\mu} \sum_{\substack{G \in \mathcal{G}_i^\mu : \\ \boldsymbol{U}(G) = \bar{G}}} \left|\left\{(T_1, \ldots, T_m) \in \breve{\mathcal{A}}_i(\mu)\,,\ \boldsymbol{G}(T_1, \ldots, T_m) = G\right\}\right|. \quad (39)$$

For each graph G, it is clear that

$$\left|\left\{(T_1, \ldots, T_m) \in \breve{\mathcal{A}}_i(\mu)\,,\ \boldsymbol{G}(T_1, \ldots, T_m) = G\right\}\right| \quad \leq \quad C\,, \quad (40)$$

where $C = C(d, t, m)$ is independent of G. Our goal now is to show that

$$\left|\left\{G \in \mathcal{G}_i^\mu\,,\ \boldsymbol{U}(G) = \bar{G}\right\}\right| \leq CK_n^{(\mu-1)/2}\,, \quad (41)$$

which is simply the number of all possible labelings of a graph $\bar{G}$ under the constraints described above. To see this, consider a breadth first search ordering of the vertices of the graph $v_0 = \circ < v_1 < \cdots < v_{|V(\bar{G})|-1}$ that begins at the root $\circ$, this ordering has the property of visiting each vertex once and that each new vertex is connected to an already visited vertex, i.e.

- $\{v_0 = \circ, v_1, \cdots, v_{|V(\bar{G})|-1}\} = V(\bar{G})$,
- $\forall j = 1, \cdots\ \exists k < j$ such that $(v_j \to v_k) \in E(\bar{G})$.

Now, starting with v_1 and by induction, after fixing the label of v_{j-1}, one can see that v_j can only be labeled in at most CK_n possible ways. So the number of all possible labelings of $\bar{G}$ is bounded by $CK_n^{|V(\bar{G})|-1} \leq CK_n^{(\mu-1)/2}$.

Furthermore, it is easy to check that

$$\left|\mathcal{R}^\mu\right| \leq C.$$

Getting back to equality (39), and using this last inequality along with inequalities (41) and (40), we obtain inequality (38), and the proposition is proved. □

Notice that for a tuple of trees $(T_1, \cdots, T_m)$ satisfying the following condition

$$\forall j < \ell, \ \sum_{k\in[m]} \varphi_{j\ell}(T_k) \in \{0, 2\},$$

if there exists a pair (j, ℓ) such that $\sum_{k\in[m]} \vec{\varphi}_{j\ell}(T_k) = 1$ and $(j, \ell) \in \mathcal{C}$, i.e. $\tau_{j\ell} \neq 0$, then $\mathbb{E}\left[\prod_{k=1}^m W(T_k)\right] = 0$. Consider the following subset $\tilde{\mathcal{A}}_i(\mu)$ of $\mathcal{A}_i(\mu)$ defined

$$\begin{aligned}\tilde{\mathcal{A}}_i(\mu) = \Big\{ & (T_1, \ldots, T_m) \in \mathcal{A}_i(\mu), \\ & \forall j < \ell, \ \sum_{k\in[m]} \varphi_{j\ell}(T_k) \in \{0, 2\}, \\ & \forall j, \ell, \ \sum_{k\in[m]} \vec{\varphi}_{j\ell}(T_k) = 1 \ \Rightarrow (j, \ell) \in \mathcal{C} \Big\}. \end{aligned} \tag{42}$$

If $(T_1, \cdots, T_m) \in \tilde{\mathcal{A}}_i(\mu)$ then the graph $G = \boldsymbol{G}(T_1, \cdots, T_m)$ constructed by merging the trees has exactly $\mu/2$ edges, and that can be seen by writing

$$\sum_{k=1}^m \varphi_{jl}(T_k) = 2\, 1_{\sum_{k=1}^m \varphi_{jl}(T_k)>0},$$

$$|E(G)| = \sum_{j<l} 1_{\sum_{k=1}^m \varphi_{jl}(T_k)>0} = \sum_{j<l} \frac{1}{2} \sum_{k=1}^m \varphi_{jl}(T_k) = \mu/2.$$

Define the set of graphs $\tilde{\mathcal{G}}_i^\mu$ analogously to $\mathcal{G}_i^\mu$ with the difference that we replace the requirement $|E(G)| \leq (\mu - 1)/2$ with $|E(G)| = \mu/2$. We can then write

$$\mathbb{E} z_i^t(r)^m = \sum_{\mu=1}^{C_E} \chi_\mu + \sum_{\mu=1}^{C_E} \xi_\mu, \tag{43}$$

where

$$\chi_\mu = \sum_{\bar{G}\in\mathcal{R}^\mu} \sum_{\substack{G\in\tilde{\mathcal{G}}_i^\mu : \\ \boldsymbol{U}(G)=\bar{G}}} \sum_{\substack{(T_1,\cdots,T_m)\in\tilde{\mathcal{A}}_i(\mu) : \\ \boldsymbol{G}(T_1,\cdots,T_m)=G}} \left(\prod_{k=1}^m \Gamma(T_k, \boldsymbol{\alpha}, t)\right) \left(\prod_{k=1}^m x(T_k)\right) \mathbb{E}\left[\prod_{k=1}^m W(T_k)\right]. \tag{44}$$

Recalling that $|\xi_\mu| = \mathcal{O}(K_n^{-1/2})$, we focus on the χ_μ. To that end, we further decompose the first sum on the unlabeled graphs $\bar{G} \in \mathcal{R}^\mu$ above into a sum on the graphs

which are trees and a sum on the graphs which are not trees, i.e., those that contain a cycle. Let us denote, respectively, the corresponding sums by χ_μ^{T} and χ_μ^{NT}, and write

$$\chi_\mu = \chi_\mu^{\mathrm{T}} + \chi_\mu^{\mathrm{NT}}.$$

We show in the following lemma that the contribution of the term χ_μ^{NT} is negligible.

Lemma 4.3 *Consider the same framework as in Proposition 4.2. We have*

$$\chi_\mu^{T} = \mathcal{O}(1) \quad \textit{and} \quad \chi_\mu^{NT} = \mathcal{O}\left(\frac{1}{K_n}\right).$$

Proof In the proof of Proposition 4.2, we have already got that $|\mathbb{E}\left[\prod_{k=1}^m W(T_k)\right]|$ is bounded by $CK_n^{-\mu/2}$, so we only need to study the quantity

$$\left|\left\{G \in \tilde{\mathcal{G}}_i^\mu \,,\ \boldsymbol{U}(G) = \bar{G}\right\}\right|,$$

in the case where $\bar{G}$ is a tree and where $\bar{G}$ in not a tree. Recall that for a given $G \in \tilde{\mathcal{G}}_i^\mu$ the graph G is connected and we have $|E(G)| = \mu/2$ so $|V(G) \setminus \{\circ\}| \leq \mu/2$ with the equality if and only if G is a tree. So repeating the same argument as in Proposition 4.2 we find that

$$\left|\left\{G \in \tilde{\mathcal{G}}_i^\mu \,,\ \boldsymbol{U}(G) = \bar{G}\right\}\right| \leq CK_n^{\mu/2} \quad \text{and} \quad \left|\left\{G \in \tilde{\mathcal{G}}_i^\mu \,,\ \boldsymbol{U}(G) = \bar{G}\right\}\right| \leq CK_n^{\mu/2-1},$$

in the case of $\bar{G}$ being a tree and not a tree respectively. Multiplying by $CK_n^{-\mu/2}$ yields to the desired result. □

4.4 Approximation of the Non-backtracking Iterations

For each n, let us now consider an i.i.d. sequence $(W^t)_{t=0,1,\ldots}$ of $n \times n$ matrices such that $W^t \overset{\mathcal{L}}{=} W$. We define the vectors $\boldsymbol{y}_{i\to j}^t$ and $\boldsymbol{y}_i^t$ recursively in t similarly to what we did for the vectors $\boldsymbol{z}_{i\to j}^t$ and $\boldsymbol{z}_i^t$, with the difference that we now replace the matrix W with the matrix W^t at step t. More precisely, we set $\boldsymbol{y}_{i\to j}^0 = \boldsymbol{x}_i^0$ for each $i, j \in [n]$ with $i \neq j$. Given $\{\boldsymbol{y}_{i\to j}^t \mid i, j \in [n],\ i \neq j\}$, we set

$$y_{i\to j}^{t+1}(r) = \sum_{\ell \in [n]\setminus\{j\}} W_{i\ell}^t f_r(\boldsymbol{y}_{\ell\to i}^t, \ell, t), \quad i \neq j. \tag{45}$$

Also,

$$y_i^{t+1}(r) = \sum_{\ell \in [n]} W_{i\ell}^t f_r(\boldsymbol{y}_{\ell\to i}^t, \ell, t). \tag{46}$$

We introduce here a similar quantity to $W(T)$ for a given labeled tree which is adapted to the computations related to the iterations y_i^t. We define $\overline{W}(T,t)$ by

$$\overline{W}(T,t) = \prod_{(u\to v)\in E(T)} W_{\ell(v)\ell(u)}^{t-|u|},$$

where we recall that $|u|$ denotes the distance of the vertex u to the root $\circ$ in the tree T. We can prove similar structural identities for y_i^t and $y_{i\to j}^t$ as what we did with the iterates z_i^t and $z_{i\to j}^t$. In fact, we have

$$y_{i\to j}^t(r) = \sum_{T\in\mathcal{T}_{i\to j}^t(r)} \overline{W}(T,t)\Gamma(T,\boldsymbol{\alpha},t)x(T),$$
$$y_i^t(r) = \sum_{T\in\mathcal{T}_i^t(r)} \overline{W}(T,t)\Gamma(T,\boldsymbol{\alpha},t)x(T).$$

Proposition 4.4 *Let $(\boldsymbol{z}^t)$ and $(\boldsymbol{y}^t)$ two sequences defined in (33) and (46) respectively, then for each $t\geq 1$ and each $\boldsymbol{m}\in\mathbb{N}^q$, we have that for each $i\in[n]$,*

$$\left|\mathbb{E}(\boldsymbol{z}_i^t)^{\boldsymbol{m}} - \mathbb{E}(\boldsymbol{y}_i^t)^{\boldsymbol{m}}\right| = \mathcal{O}\left(\frac{1}{\sqrt{K_n}}\right).$$

Proof We follow the same strategy of proof as in Proposition 4.2. For simplicity let us fix $\boldsymbol{m}(r)=m$ for a certain $r\in[q]$. We have

$$\mathbb{E}\left[y_i^t(r)^m\right] = \sum_{\mu=1}^{C_E}\sum_{(T_1,\cdots,T_m)\in\mathcal{A}_i(\mu)}\left(\prod_{k=1}^m\Gamma(T_k,\boldsymbol{\alpha},t)\right)\left(\prod_{k=1}^m x(T_k)\right)\mathbb{E}\left[\prod_{k=1}^m\overline{W}(T_k,t)\right].$$

As in the case of (z_i^t), we can also decompose this sum into a sum over trees $(T_1,\cdots,T_m)$ in the set $\breve{\mathcal{A}}_i(\mu)$ (defined in (37)) and trees that are in the set $\tilde{\mathcal{A}}_i(\mu)$ (defined in (42)). The contribution of m-tuples of trees in $\breve{\mathcal{A}}_i(\mu)$ is of order $K_n^{-1/2}$, so we may focus on m-tuples of trees in $\tilde{\mathcal{A}}_i(\mu)$. Recall the definition of a graph $G\in\mathcal{G}_i^\mu$ as the merger of trees $(T_1,\cdots,T_m)$ where we identify vertices u that have the same label $\ell(u)$. As in the previous proof, we further partition these graphs into trees and graphs that contain at least a cycle. The latter have a contribution of order K_n^{-1} so we may focus on the contribution of graphs G that are trees. Write

$$\bar{\chi}_\mu^T = \sum_{\substack{\bar{G}\in\mathcal{R}^\mu\\ \bar{G}\text{ is a tree}}}\sum_{\substack{G\in\tilde{\mathcal{G}}_i^\mu:\\ \boldsymbol{U}(G)=\bar{G}}}\sum_{\substack{(T_1,\cdots,T_m)\in\tilde{\mathcal{A}}_i(\mu):\\ \boldsymbol{G}(T_1,\cdots,T_m)=G}}\left(\prod_{k=1}^m\Gamma(T_k,\boldsymbol{\alpha},t)\right)\left(\prod_{k=1}^m x(T_k)\right)\mathbb{E}\left[\prod_{k=1}^m\overline{W}(T_k)\right].$$

The proof of this proposition will be completed if we can show that $\chi_\mu^T=\bar{\chi}_\mu^T$.

First, notice that the terms $\prod_{k=1}^{m}\Gamma(T_k,\boldsymbol{\alpha},t)$ and $\prod_{k=1}^{m}x(T_k)$ are the same in the expressions of χ_{μ}^{T} (defined in (44)) and $\bar{\chi}_{\mu}^{T}$. So it suffices study the term $\mathbb{E}\left[\prod_{k=1}^{m}\overline{W}(T_k)\right]$. Two cases can be studied, whether this term is zero or nonzero.

Consider any m-tuple of trees $(T_1,\cdots,T_m)\in\tilde{\mathcal{A}}_i(\mu)$, if

$$\mathbb{E}\left[\prod_{k=1}^{m}\overline{W}(T_k)\right]\neq 0\,,$$

then for every matrix entry (i,j) which is represented in the trees $T_1,\cdots,T_m$ there exist exactly two edges $(a\to b)$ and $(c\to d)$ such that $\{\ell(a),\ell(b)\}=\{\ell(c),\ell(d)\}=\{i,j\}$, in addition $|a|=|c|$ otherwise $\mathbb{E}\left[\prod_{k=1}^{m}\overline{W}(T_k)\right]=0$, we then obtain a second moment of W which means that

$$\mathbb{E}\left[\prod_{k=1}^{m}\overline{W}(T_k)\right]=\mathbb{E}\left[\prod_{k=1}^{m}W(T_k)\right]\,.$$

Now suppose for the sake of contradiction that

$$\mathbb{E}\left[\prod_{k=1}^{m}\overline{W}(T_k)\right]=0\quad\text{and}\quad\mathbb{E}\left[\prod_{k=1}^{m}W(T_k)\right]\neq 0\,,$$

we show that in this case the graph $G=\boldsymbol{G}(T_1,\cdots,T_m)$ is not a tree which is a contradiction. There exists a matrix entry (i,j) with $i<j$ which is represented in the trees $(T_1,\cdots,T_m)$ by two edges $(a\to b)$ and $(c\to d)$ such that vertices a and c do not have the same distance to the root $\circ$, i.e. $|a|>|c|$ for example. This is because $\mathbb{E}\left[\prod_{k=1}^{m}\overline{W}(T_k)\right]=0$ and because $\tau_{ij}\neq 0$, $s_{ij}\neq 0$ and $s_{ji}\neq 0$. Three possible cases can be considered:

- $(a\to b)$ and $(c\to d)$ exist on the same path of a certain tree: by the non-backtracking condition, these edges should be separated by at least one vertex say e of label $k\notin\{i,j\}$, i.e.:

$$\cdots\to a\to b\to e\to\cdots\to c\to d\cdots\to\circ.$$

  As for the graph G, this means that starting from a vertex of label i we should pass through a vertex of label $k\notin\{i,j\}$ and then return to the vertex of label i which creates a cycle.
- $(a\to b)$ and $(c\to d)$ exist in two different trees say T_1 and T_2 respectively:

$$\begin{aligned}\cdots\to a\to b\to\cdots\to\cdots\to\circ\quad(T_1)\\ \cdots\to *\to c\to\ d\to\cdots\to\circ\quad(T_2)\end{aligned}$$

  First notice that the labels of the vertices in each of these two paths are different: if two vertices on the same path have the same label say k then due to the non-backtracking condition they should be separated by at least two other vertices which

result in a cycle in the graph G. Recall that the roots $\circ_{T_1}$ and $\circ_{T_2}$ are identified in the graph G which means that in G there exist a path from the vertex $\ell(b)$ to $\circ$ and another path from $\ell(d)$ to $\circ$, these two paths are distinct as they have different lengths which is a consequence of the condition $|a| < |c|$. In addition $\ell(b)$ and $\ell(d)$ are either equal or linked in G, this creates a cycle in the graph.

- $(a \to b)$ and $(c \to d)$ exist in two different paths of the same tree: similar to the previous case.

□

4.5 Approximation of the AMP Iterations

Let us now establish the relationship between AMP iterates $(\boldsymbol{x}^t)_t$ and the non-backtracking iterations $(\boldsymbol{z}^t)_t$. We see in the following proposition that the moments of $\boldsymbol{x}^t$ can be approximated by the moments of $\boldsymbol{z}^t$.

Recall that we denote by $\overline{\mathcal{T}}_i^t$ the set of planted and labeled trees of depth at most t, such that the type of the root is equal to i. Analogously, $\mathcal{T}_i^t$ denotes the subset of trees satisfying, in addition, the non-backtracking condition (see Sect. 4.2).

Proposition 4.5 *For each $t \geq 1$ and each $\boldsymbol{m} \in \mathbb{N}^q$, we have that for each $i \in [n]$,*

$$\left|\mathbb{E}(\boldsymbol{x}_i^t)^{\boldsymbol{m}} - \mathbb{E}(\boldsymbol{z}_i^t)^{\boldsymbol{m}}\right| = \mathcal{O}\left(\frac{1}{\sqrt{K_n}}\right).$$

In order to prove this proposition we need the following structural lemma that connects $x_i^t(r)$ to $z_i^t(r)$ for $i \in [n], r \in [q]$ and $t \in \mathbb{N}$. Consider $\bar{\mathcal{U}}_i^t$ (resp. $\mathcal{U}_i^t$) the set of unmarked trees of the set $\bar{\mathcal{T}}_i^t$ (resp. $\mathcal{T}_i^t$). We can consider that these sets are constructed by identifying the trees with the same structure and labels. Denote also by $\mathcal{U}$ the map that assigns to a tree T its unmarked version $\hat{T} := \mathcal{U}(T)$. The two equations in Lemma 4.1 can be reformulated as:

$$z_{i\to j}^t(r) = \sum_{\hat{T} \in \mathcal{U}_{i\to j}^t} W(\hat{T})\Gamma(\hat{T}, r, t)x(\hat{T}),$$
$$z_i^t(r) = \sum_{\hat{T} \in \mathcal{U}_i^t} W(\hat{T})\Gamma(\hat{T}, r, t)x(\hat{T}),$$

where $W(T)$ and $x(T)$ are invariant with respect to the marking of the tree, and

$$\Gamma(\hat{T}, r, t) := \sum_{T \in \mathcal{T}^t(r) \,:\, \mathcal{U}(T) = \hat{T}} \Gamma(T, \boldsymbol{\alpha}, t), \quad \forall \hat{T} \in \mathcal{U}_i^t.$$

Consider $\mathcal{B}^t \subset \bar{\mathcal{U}}^t$ to be the set of trees T such that for each $(u \to v) \in E(T)$ we have $\ell(u) \neq \ell(v)$, in addition at least one of the following conditions holds,

- there exists a backtracking path of length 3: a path $a \to b \to c \to d$ such that $\ell(a) = \ell(c)$ and $\ell(b) = \ell(d)$,

- there exists a backtracking star: $a \to b \to c$ and $a' \to b \to c$ such that $\ell(a) = \ell(a') = \ell(c)$.

Lemma 4.6 *For each t, r, i there exists a $\tilde{\Gamma}(., t, r)$ such that $\tilde{\Gamma}(T, r, t) = \mathcal{O}(1)$ uniformly in T and*

$$x_i^t(r) = z_i^t(r) + \sum_{T \in \mathcal{B}_i^t} W(T)\tilde{\Gamma}(T, r, t)x(T).$$

Proof We prove this lemma by induction on t. The cases $t = 0, 1$ are simple, suppose that $t \geq 2$, and that the equation is valid for t. Recall the AMP recursion given by,

$$x_i^{t+1}(r) = \sum_{\ell=1}^{n} W_{i\ell} f_r(x_\ell^t) - \sum_{s=1}^{q}\sum_{\ell=1}^{n} W_{i\ell} W_{\ell i} f_s(x_i^{t-1})\partial_{x(s)} f_r(x_\ell^t).$$

Here we omit the dependence of f on ℓ and t, i.e. $f_r(x_\ell^t, \ell, t) = f_r(x_\ell^t)$. Recall that f_r is a multivariate polynomial, so by Taylor's expansion at $z_{\ell\to i}^t$, we can write

$$\begin{aligned} f_r(x_\ell^t) = f_r(z_{\ell\to i}^t) &+ \sum_{s\in[q]} \left(x_\ell^t(s) - z_{\ell\to i}^t(s)\right) \partial_{x(s)} f_r(z_{\ell\to i}^t) \\ &+ \sum_{\boldsymbol{k}\,:\, k_1+\cdots+k_q\geq 2} \left[\prod_{s=1}^{q} \frac{\left(x_\ell^t(s) - z_{\ell\to i}^t(s)\right)^{k_s}}{k_s!}\right] D^{\boldsymbol{k}} f_r(z_{\ell\to i}), \end{aligned} \tag{47}$$

where for $\boldsymbol{k} \in \mathbb{N}^q$ and $x \in \mathbb{R}^q$ we denote by $D_x^{\boldsymbol{k}}$ the following differential operator

$$D^{\boldsymbol{k}} g(x) = \frac{\partial^{k_1+\cdots+k_q} g(x)}{\partial x(1)^{k_1} \cdots x(q)^{k_q}}.$$

Let $e_\ell^t(r) := \sum_{T\in\mathcal{B}_\ell^t} W(T)\tilde{\Gamma}(T, r, t)x(T)$, by the induction hypothesis we have

$$\begin{aligned} x_\ell^t(r) &= z_\ell^t(r) + e_\ell^t(r) \\ &= z_{\ell\to i}^t(r) + z_{\ell,i}^t(r) + e_\ell^t(r), \end{aligned}$$

where we use the notation $z_{\ell,i}^t(r) := W_{\ell i} f_r(z_{i\to\ell}^{t-1})$. Plugging this equation into (47) gives

$$\begin{aligned} f_r(x_\ell^t) = f_r(z_{\ell\to i}^t) &+ \sum_{s\in[q]} \left(z_{\ell,i}^t(s) + e_\ell^t(s)\right) \partial_{x(s)} f_r(z_{\ell\to i}^t) \\ &+ \sum_{k\,:\, k_1+\cdots+k_q\geq 2} \left[\prod_{s=1}^{q} \frac{\left(z_{\ell,i}^t(s) + e_\ell^t(s)\right)^{k_s}}{k_s!}\right] D^{k} f_r(z_{\ell\to i}), \end{aligned} \tag{48}$$

Now, multiplying by $W_{i\ell}$ on both sides and summing over ℓ gives the following

$$\sum_{\ell\in[n]} W_{i\ell} f_r(x_\ell^t) = z_i^{t+1}(r) + \sum_{\ell\in[n],s\in[q]} W_{i\ell}\left(z_{\ell,i}^t(s) + e_\ell^t(s)\right)\partial_{x(s)} f_r(z_{\ell\to i}^t) + \sum_{\ell\in[n],\ k_1+\cdots+k_q\geq 2} W_{i\ell}\left[\prod_{s=1}^q \frac{\left(z_{\ell,i}^t(s) + e_\ell^t(s)\right)^{k_s}}{k_s!}\right] D^k f_r(z_{\ell\to i}). \tag{49}$$

The first term is obtained by the definition of $z_i^{t+1}(r)$, see Eq (33). The second term can be decomposed into the two following sums,

$$\sum_{\ell\in[n],s\in[q]} W_{i\ell}W_{\ell i} f_r(z_{i\to\ell}^{t-1})\partial_{x(s)} f_r(z_{\ell\to i}^t) + \sum_{\ell\in[n],s\in[q]} W_{i\ell} e_\ell^t(s)\partial_{x(s)} f_r(z_{\ell\to i}^t).$$

Now subtracting the Onsager term from both sides of Eq (49) gives the following

$$x_i^{t+1}(r) = z_i^{t+1}(r) - \sum_{\ell\in[n],s\in[q]} W_{i\ell}W_{\ell i}\left(f_r(x_i^{t-1})\partial_{x(s)} f_r(x_\ell^t) - f_r(z_{i\to\ell}^{t-1})\partial_{x(s)} f_r(z_{\ell\to i}^t)\right) + \sum_{\ell\in[n],s\in[q]} W_{i\ell} e_\ell^t(s)\partial_{x(s)} f_r(z_{\ell\to i}^t) + \sum_{\ell\in[n],\ k_1+\cdots+k_q\geq 2} W_{i\ell}\left[\prod_{s=1}^q \frac{\left(z_{\ell,i}^t(s) + e_i^t(s)\right)^{k_s}}{k_s!}\right] D^k f_r(z_{\ell\to i}). \tag{50}$$

Denote by S_1, S_2 and S_3, respectively, the three terms in the right-hand side of the previous equation except $z_i^{t+1}(r)$. One wants to prove that these three terms can be written as sums over trees in $T \in \mathcal{B}_i^{t+1}$ of terms having the form,

$$W(T)\tilde{\Gamma}(T, r, t)x(T),$$

where $\tilde{\Gamma}(T, r, t)$ is obtained by construction, the exact form of this term is not important, we only need it to be bounded as n goes to infinity.

The term S_2

The second term is given by the following formula,

$$S_2 = \sum_{\ell\in[n],\ s\in[q]} W_{i\ell} e_\ell^t(s)\partial_{x(s)} f_r(z_{\ell\to i}^t).$$

The terms in this sum are given by

$$e_\ell^t(s) = \sum_{T \in \mathcal{B}_\ell^t} W(T)\Gamma(T, s, t)x(T),$$

$$\partial_{x(s)} f_r(z_{\ell\to i}) = \sum_{k_1+\cdots+k_q \le d} \alpha_{k_1,\cdots,k_q}(r, \ell, t) k_s \left(z_{\ell\to i}^t(s)\right)^{k_s-1} \prod_{u\in[q]\setminus\{s\}} \left(z_{\ell\to i}^t(u)\right)^{k_u},$$

with

$$z_{\ell\to i}^t(u) = \sum_{T\in\mathcal{U}_{\ell\to i}^t} W(T)\Gamma(T, u, t)x(T).$$

S_2 can thus be interpreted as a sum over trees $T \in \mathcal{B}_i^{t+1}$ constructed as follows:

- The root $\circ$ has a type equal to i, and $\circ$ has a child, say $\square$, of type ℓ. This is due to $W_{i\ell}$.
- The vertex $\square$ is the root of a tree in $\mathcal{B}_\ell^t$. This is due to the term $e_\ell^t(s)$.
- The root's child $\square$ is also the root of $k_1 + \cdots + (k_s - 1) + \cdots + k_q$ additional trees in $\mathcal{U}_{\ell\to i}^t$. This is due to the term $\partial_{x(s)} f_r(z_{\ell\to i})$. Note that in total, $\square$ has at most $d \ge k_1 + \cdots + (k_s - 1) + \cdots + k_q + 1$ children.

By construction, we can easily see that T is in $\mathcal{B}_i^{t+1}$.

The term S_1

The first term is given by the following formula,

$$S_1 = \sum_{\ell\in[n], s\in[q]} W_{i\ell}W_{\ell i}\left(f_r(x_i^{t-1})\partial_{x(s)}f_r(x_\ell^t) - f_r(z_{i\to\ell}^{t-1})\partial_{x(s)}f_r(z_{\ell\to i}^t)\right).$$

Doing a Taylor expansion of the polynomial $g : (x, x') \mapsto f_r(x)\partial_{x(s)}f_r(\boldsymbol{x}')$ around $(z_{i\to\ell}^{t-1}, z_{\ell\to i}^t)$ gives

$$S_1 = \sum_{\ell\in[n]s\in[q]} W_{i\ell}W_{\ell i} \sum_{|j|+|k|\ge 1} \left[\prod_{u=1}^q \frac{\left(z_{i,\ell}^{t-1}(u) + e_i^{t-1}(u)\right)^{j_u}\left(z_{\ell,i}^t(u) + e_\ell^t(u)\right)^{k_u}}{(j_u+k_u)!}\right] D^{(j,k)}g(z_{i\to\ell}^{t-1}, z_{\ell\to i}^t).$$

S_1 can be seen as a sum, up to multiplication factors, of the following terms

$$W_{i\ell}W_{\ell i}\prod_{u=1}\left(z_{i,\ell}^{t-1}(u) + e_i^{t-1}(u)\right)^{j_u}\left(z_{\ell,i}^t(u) + e_\ell^t(u)\right)^{k_u}\left(z_{i\to\ell}^{t-1}(u)\right)^{a_u}\left(z_{\ell\to i}^t(u)\right)^{b_u},$$

with the constraint that $\sum_{u=1}^{q}(j_u + k_u) \geq 1$. To show that S_1 can be seen a sum of trees T belonging to $\mathcal{B}_i^{t+1}$, two cases should be considered, either it exists a u such that $j_u \geq 1$ or $k_u \geq 1$.

- If there exists a u such that $j_u \geq 1$, we construct a tree in $\mathcal{B}^{t+1}$ as follows:
  - The root $\circ$ has a type equal to i, and $\circ$ has a child, say $\square$, of type ℓ. This is due to $W_{i\ell}$.
  - The vertex $\square$ is the root of trees in $\mathcal{U}_{\ell\to i}^{t}$, which is due to the multiplication by $z_{\ell\to i}^{t}(u)$.
  - The vertex $\square$ has a child, say $\diamondsuit$, of type i, which is due to $W_{\ell i}$.
  - The vertex $\diamondsuit$ is the root of trees in $\mathcal{U}_{i\to\ell}^{t-1}$, which is due to $z_{i\to\ell}^{t-1}(u)$.

  Now because $j_u \geq 1$, at least one of the following holds:
  - The vertex $\diamondsuit$ is the root of trees in $\mathcal{B}_i^{t-1}$, which obviously results in a tree $T \in \mathcal{B}_i^{t+1}$.
  - The vertex $\diamondsuit$ is has a child of type ℓ, which creates a backtracking path of length 3 of types $\ell \to i \to \ell \to i$ which also results in a tree $T \in \mathcal{B}_i^{t+1}$. This child is the root of a tree in $\mathcal{U}_{i\to\ell}^{t-1}$. And this is due to the term $z_{i,\ell}^{t-1}$.
- If there exists a u such that $k_u \geq 1$, we repeat the same argument. This time, the multiplication by $z_{\ell,i}^{t}(u)$ gives a backtracking star $[i,\ i \to \ell \to i]$, which results in a tree $T \in \mathcal{B}_i^{t+1}$. Otherwise, the multiplication by $e_\ell^t(u)$ adds a tree in $\mathcal{B}_\ell^t$ which obviously results in a final tree T belonging to $\mathcal{B}_i^{t+1}$.

The term S_3

The third term is given by the following formula,

$$S_3 = \sum_{\ell\in[n],\ k_1+\cdots+k_q\geq 2} W_{i\ell}\left[\prod_{s=1}^{q}\frac{\left(z_{\ell,i}^{t}(s)+e_i^t(s)\right)^{k_s}}{k_s!}\right] D^k f_r(z_{\ell\to i}).$$

Similarly to the interpretation of S_2 as a sum of trees in $\mathcal{B}_i^{t+1}$, we can repeat the same arguments for S_3. The terms that have $e_i^t(s)$ as a multiplication factor naturally results in trees belonging to $\mathcal{B}_i^{t+1}$. In the other case, notice that the constraints $k_1+\cdots+k_q \geq 2$ implies that a term of the form $W_{i\ell}z_{\ell,i}^{t}(s)z_{\ell,i}^{t}(s')$ always exists, this term produces a backtracking star and thus the final tree T belongs to $\mathcal{B}_i^{t+1}$.

By studying the tree terms, we proved the existence of a $\tilde{\Gamma}(T,t,t+1)$ such that

$$x_i^{t+1}(r) = z_i^{t+1} + \sum_{T\in\mathcal{B}_i^{t+1}} W(T)\tilde{\Gamma}(T,r,t+1)x(T).$$

Where $\tilde{\Gamma}(T,t,t+1)$ is a function of $\tilde{\Gamma}(T,t,t)$ and the activation functions' coefficients. It remains to check that $\tilde{\Gamma}(T,t,t+1) = \mathcal{O}(1)$. This can be easily verified, and its proof will be omitted. □

Remark 4.3 The previous proof is a non-Symmetric adaptation of the techniques developed in [7] and [21] in the symmetric case. Instead of terms $W_{i\ell}^2$ in the symmetric case, we handle their counterparts $W_{i\ell}W_{\ell i}$ in the non-Symmetric case and properly interpret them as edges of a tree. Accordingly, we rely on an Onsager term based on matrix $W \odot W^\top$ instead of $W^{\odot 2}$.

Finally, we can prove Proposition 4.5 by repeating the same arguments used in the proof of Proposition 4.2.

Proof of Proposition 4.5 We can restrict ourselves to the case of $\boldsymbol{m}(r) = m$ and $\boldsymbol{m}(s) = 0$ for $s \neq r$. The m-th power of $x_i^t(r)$ is given by

$$\mathbb{E}\left(x_i^t(r)\right)^m - \mathbb{E}\left(z_i^t(r)\right)^m = \mathbb{E}\left(z_i^t(r) + \sum_{T\in\mathcal{B}_i^t} W(T)\tilde{\Gamma}(T,r,t)x(T)\right)^m - \mathbb{E}\left(z_i^t(r)\right)^m$$
$$\leq C \sum_{T_1\in\mathcal{B}_i^t} \sum_{T_2,\cdots,T_m\in\mathcal{B}_i^t\cup\mathcal{T}_i^t(r)} \left|\mathbb{E}\prod_{i=1}^m W(T_i)\right|.$$

The key observation here is to notice that the graph obtained by merging the trees $(T_1,\cdots,T_m)$ has an edge which is the result of the fusion of at least three edges, and this is because T_1 has a backtracking path or a backtracking star. This implies a bound on the number of edges of the resulting graph.

4.6 End of Proof of Theorem 3.3

We now show that the sequence of Gaussian vectors (U^t) defined in (24) by the Density Evolution equations approximate the iterations $(\boldsymbol{y}^t)$ defined in (45) and (46) where the matrices $(W^t)_{t\in\mathbb{N}}$ are independent and Gaussian.

Proposition 4.7 *Let W be a random matrix defined in (3) and satisfying assumptions A-1 and A-2, suppose in addition that W is gaussian. Let $(W^t)_{t\in\mathbb{N}}$ be a sequence of independent copies of W. Then for each multi-index $\boldsymbol{m}\in\mathbb{N}^q$ and each integer $t>0$ we have*

$$\max_{i\in[n]} \left|\mathbb{E}\left[(\boldsymbol{y}_i^t)^{\boldsymbol{m}}\right] - \mathbb{E}\left[(U_i^t)^{\boldsymbol{m}}\right]\right| \longrightarrow 0.$$

Remark 4.4 Recall that the random matrix $(U_1^t,\cdots,U_n^t)^\top \in \mathbb{R}^{n\times q}$ is defined such that $(U_i^t)_{i\in[n]}$ are independent and such that $U_i^t \sim \mathcal{N}(0, Q_i^t)$ where $(Q_i^t)_t$ is a sequence of $k\times k$ covariance matrices defined recursively by

$$Q_i^{t+1} = \sum_{\ell\in[n]} s_{i\ell}\mathbb{E}\left[f(U_\ell^t,\ell,t)f(U_\ell^t,\ell,t)^\top\right].$$

In particular, the law of U does not depend on our correlation profile.

We also recall that the iterations y^t are defined by $y^0_{i\to j} = x^0_i$ and

$$y^{t+1}_{i\to j} = \sum_{\ell\in[n]\setminus\{j\}} W^t_{i\ell} f(y^t_{\ell\to i}, \ell, t)$$

which implies that the conditional distribution of $y^{t+1}_{i\to j}$ given $\mathcal{F}_t := \sigma\{W^0, \cdots, W^{t-1}\}$ is $\mathcal{N}_k\left(0, H^{t+1}_{ij}\right)$ where $(H^t_{ij})_t$ is a sequence of $q\times q$ covariance matrices defined for each $t\in\mathbb{N}^\star$ by the following recursion

$$H^{t+1}_{ij} = \sum_{\ell\in[n]\setminus\{j\}} s_{i\ell}\mathbb{E}\left[f(y^t_{\ell\to i}, \ell, t) f(y^t_{\ell\to i}, \ell, t)^\top\right].$$

We therefore notice that the conditional distribution of $y^{t+1}_{i\to j}$ given $\mathcal{F}_t$ is unchanged if we replace the matrix W with a random symmetric matrix $\tilde{W}$ having the same variance profile as W. By doing so, we can directly apply the result in [21, Proposition 15].

Combining the previous results we get the following convergence for each multi-index $\boldsymbol{m}$

$$\max_{i\in[n]} \left|\mathbb{E}\left[(x^t_i)^{\boldsymbol{m}}\right] - \mathbb{E}\left[(U^t_i)^{\boldsymbol{m}}\right]\right| \longrightarrow 0.$$

We can then use the triangular inequality to get this same result for any multivariate polynomial with bounded coefficients instead considering only the monomial $X^{\boldsymbol{m}}$.

Proposition 4.8 *Let $\psi : \mathbb{R}^q \times [n] \to \mathbb{R}$ such that $\psi(., \ell)$ is a multivariate polynomial with bounded degree and bounded coefficients. Then for each subset $\mathcal{S}^{(n)}$ of $[n]$ with $|\mathcal{S}^{(n)}| \to \infty$, it holds that*

$$\frac{1}{|\mathcal{S}^{(n)}|}\sum_{i\in\mathcal{S}^{(n)}} \mathbb{E}\left[\psi(x^t_i, i)\right] - \mathbb{E}\left[\psi(U^t_i, i)\right] \longrightarrow 0.$$

Finally, in order to get the convergence in probability stated in Theorem 3.3, we only need to show that the following variance

$$\mathbb{V}ar\left[\frac{1}{K_n}\sum_{i\in S^{(n)}} \psi(x^t_i, i)\right] \longrightarrow 0 \tag{51}$$

converges to zero. The proof of this convergence is similar to the proof of [21, Proposition 17] and thus is be omitted.

The proof of Theorem 3.3 follows then from Proposition 4.8 and the convergence in (51).

5 AMP with General Activation Functions and Nonzero Diagonal Matrix

5.1 AMP for General Activation Functions

Now that we have proved the AMP convergence result for polynomial activation functions in Theorem 3.1, we can generalize this result for non-polynomial activation functions by approximation arguments. In other words we complete the proof of our main Theorem 2.1 still assuming that the matrix model has a zero diagonal ($X_{ii} = 0$).

We start this section with an approximation of the activation function h by polynomials in order to use the convergence results of polynomial AMP.

Lemma 5.1 *Let h be an activation function satisfying A-5 and let $\left(Z^1, \cdots, Z^t\right) \sim \mathrm{DE}\left(S, h, \boldsymbol{x}^0, t\right)$. Let $e > 0$ be a (small) real number, then there exists a set of functions $(p_e(\cdot,\cdot,t))_{t=1}^{t_{\max}}$ such that for each $\eta \in \mathcal{Q}_\eta$, $p_e(.,\eta,t)$ is a polynomial and*

$$\mathbb{E}\left(h(Z_i^t, \eta_i, t) - p_e(Z_i^t, \eta_i, t)\right)^2 \leq e \quad \text{and} \quad \left|\mathbb{E}\left(\partial h(Z_i^t, \eta_i, t) - \partial p_e(Z_i^t, \eta_i, t)\right)\right| \leq e,$$

for $t = 0, \cdots, t_{\max}$ with the convention that $Z^0 = \boldsymbol{x}^0$ deterministic. In addition, let $\breve{R}_i^{t_{\max}}$ be the covariance matrix of the i-th row of $\left(\breve{\boldsymbol{Z}}^1, \cdots, \breve{\boldsymbol{Z}}^t\right) \sim \mathrm{DE}\left(p_e, \boldsymbol{x}^0, S, t\right)$, then there exists $\delta(e)$ such that $\delta(e) \to 0$ when $e \to 0$ and

$$\| R_i^{t_{\max}} - \breve{R}_i^{t_{\max}} \| \leq \delta(e), \quad \forall i \in [n].$$

In order to prove this lemma, we need to show that the variances of Z_i^t are bounded away from zero. To that end, we use Assumptions A-4, A-5 and A-6.

Lemma 5.2 *Let S be a matrix satisfying A-2, $\boldsymbol{x}^0$ an n-dimensional vector satisfying A-4, h a function satisfying A-5 and A-6. Following the notations of Definition 1.3 let $\left(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t\right) \sim \mathrm{DE}\left(h, \boldsymbol{x}^0, S, t\right)$ and recall the definition of the covariance matrix $R_i^t \in \mathbb{R}^{t\times t}$. Then for every $t \in \mathbb{N}$ there exist two constant $C = C(t) > 0$ and $c = c(t) > 0$ such that*

(1) The spectral norms of the covariance matrices are bounded

$$\forall n \in \mathbb{N},\ \forall i \in [n], \quad \|R_i^t\| \leq C.$$

(2) The variances of Z_i^t are bounded away from zero

$$\forall n \in \mathbb{N},\ \forall i \in [n], \quad R_i^t(t,t) \geq c.$$

The proof of this technical lemma is given in Appendix 1. The proof of the first part of Lemma 5.1 relies on the polynomial density Lemma C.1 and the fact that the variances of Z_i^t are bounded from above and also bounded away from zero which is detailed in Lemma 5.2. The second part uses the same proof technique described in the proof of

Lemma 5.6. An immediate consequence of this approximation is that the covariance matrices $\check{R}_i^{t_{\max}}$ are also bounded.

Let $(\check{\boldsymbol{x}}^t)$ the AMP sequence considered in Theorem 3.1. The following lemma allows us to replace the "random" formulation of the Onsager term by a deterministic equivalent, i.e.

$$\operatorname{diag}\Big(W \odot W^\top \partial p_e(\check{\boldsymbol{x}}^t, \cdot, t)\Big) \quad \text{with} \quad \operatorname{diag}\Big(V \partial p_e(\check{\boldsymbol{x}}^t, \cdot, t)\Big).$$

Lemma 5.3 *For each $t \in \mathbb{N}$ there exists a constant C that does not depend on n such that:*

$$\mathbb{E}\left[\left(\sum_{j\in[n]} \big(W_{ij}W_{ji} - V_{ij}\big)\, \partial p_e(\check{x}_j^t, \eta_j, t)\right)^4\right] \le C/K_n^2 \quad \textit{for all } i \in [n].$$

where $V_{ij} = \tau_{ij}\sqrt{s_{ij}s_{ji}} = \mathbb{E}\big[W_{ij}W_{ji}\big]$.

The proof of this lemma is provided in Appendix 1.

The following lemma gives the desired comparison of two sequences $(\boldsymbol{x}^t)$ and $(\check{\boldsymbol{x}}^t)$ defined by

$$(\boldsymbol{x}^t) = \text{AMP-Z}\Big(X, S, h, \boldsymbol{x}^0, \boldsymbol{\eta}\Big) \quad \text{and} \quad (\check{\boldsymbol{x}}^t) = \text{AMP-W}\Big(X, S, p_e, \boldsymbol{x}^0, \boldsymbol{\eta}\Big), \quad (52)$$

where p_e is the polynomial approximation of the function h by an error margin e in the sense of Lemma 5.1.

Lemma 5.4 *Fix $t_{\max} > 0$. Let $(\boldsymbol{x}^t)$ and $(\check{\boldsymbol{x}}^t)$ be two AMP sequences defined as in Eq. (52), then there exists $\delta(e) \to 0$ as $e \to 0$ such that the following holds for each $t = 1, \cdots, t_{\max}$,*

$$\|\boldsymbol{x}^t - \check{\boldsymbol{x}}^t\|_n \le \delta(e) + o_{\mathbb{P}}(1) \quad \textit{and} \quad \|h(\boldsymbol{x}^t) - p_e(\check{\boldsymbol{x}}^t)\|_n \le \delta(e) + o_{\mathbb{P}}(1),$$

where $o_{\mathbb{P}}(1) \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Using this Lemma, we are now able to prove the AMP convergence result for general activation functions.

Proof of Theorem 2.1 in the zero-diagonal case Let $\varphi : \mathbb{R}^{t_{\max}} \to \mathbb{R}$ be a pseudo-Lipschitz function and denote $\boldsymbol{x}_i = \big(x_i^1, \cdots, x_i^{t_{\max}}\big)^\top$ and $\check{\boldsymbol{x}}_i = \big(\check{x}_i^1, \cdots, \check{x}_i^{t_{\max}}\big)^\top$, without loss of generality we omit the scalars β_i and the parameters η_i by considering that φ depends also on the index i. We have

$$\frac{1}{n}\sum_{i\in[n]} \varphi(\boldsymbol{x}_i) = \frac{1}{n}\sum_{i\in[n]} \big(\varphi(\boldsymbol{x}_i) - \varphi(\check{\boldsymbol{x}}_i)\big) + \frac{1}{n}\sum_{i\in[n]} (\varphi(\check{\boldsymbol{x}}_i) - \varphi(\check{\mathbf{Z}}_i)) + \frac{1}{n}\sum_{i\in[n]} (\varphi(\check{\mathbf{Z}}_i) - \varphi(\mathbf{Z}_i)).$$

The pseudo-Lipschitz property of φ implies that

$$\frac{1}{n}\left|\sum_{i\in[n]}\varphi(\boldsymbol{x}_i)-\varphi(\check{\boldsymbol{x}}_i)\right| \le \frac{C}{n}\sum_{i\in[n]}\|\boldsymbol{x}_i-\check{\boldsymbol{x}}_i\|\left(1+\|\boldsymbol{x}_i\|+\|\check{\boldsymbol{x}}_i\|\right)$$
$$\le C\left(\sum_{t=1}^{t_{\max}}\|\boldsymbol{x}^t-\check{\boldsymbol{x}}^t\|_n\right)\left(1+\sum_{t=1}^{t_{\max}}\|\boldsymbol{x}^t\|_n+\sum_{t=1}^{t_{\max}}\|\check{\boldsymbol{x}}^t\|_n\right).$$

By Lemma 5.4 we have $\sum_{t=1}^{t_{\max}}\|\boldsymbol{x}^t-\check{\boldsymbol{x}}^t\|_n \le \delta(e)+o_{\mathbb{P}}(1)$, and by Theorem 3.1 applied to the test function $x\mapsto x^2$ we get $\|\check{\boldsymbol{x}}^t\|_n \le C+o_{\mathbb{P}}(1)$ which also implies that $\|\boldsymbol{x}^t\|_n \le C+o_{\mathbb{P}}(1)$, finally we have

$$\frac{1}{n}\left|\sum_{i\in[n]}\varphi(\boldsymbol{x}_i)-\varphi(\check{\boldsymbol{x}}_i)\right| \le \delta(e)+o_{\mathbb{P}}(1).$$

By Theorem 3.1, we have that

$$\frac{1}{n}\sum_{i\in[n]}(\varphi(\check{\boldsymbol{x}}_i)-\varphi(\check{\boldsymbol{Z}}_i)) = o_{\mathbb{P}}(1)\,.$$

And finally by using Lemma 5.1 we get

$$\frac{1}{n}\left|\sum_{i\in[n]}\varphi(\boldsymbol{Z}_i)-\varphi(\check{\boldsymbol{Z}}_i)\right| \le \delta(e),$$

which concludes the proof of our main theorem.

In order to provide a comparison between the two AMP sequences in (52), we need the boundedness of the spectral norm of W, a technical yet very important condition. This condition is enforced by A-3 that controls the sparsity level of the random matrix.

Proposition 5.5 *Let A-1, A-2 and A-3 hold true. Then the following bound holds true with probability one,*

$$\sup_{n\ge 1}\|W\| < \infty.$$

The proof of this proposition is due to a result of [4] and is provided in Appendix 1. In the following paragraph we give the sketch of proof of Lemma 5.4.

Proof of Lemma 5.4 The proof is basically an induction argument in which we use Lemma 5.1, Lemma 5.3 and the AMP convergence result for polynomial activation functions. The base case ($t = 1$) is easy. Suppose now that the result is valid for

all $s = 1, \cdots, t$ and let us prove that it also holds for $s = t+1$. By the triangular inequality, we can write

$$\|\boldsymbol{x}^{t+1} - \check{\boldsymbol{x}}^{t+1}\|_n \leq \|W\| \|h(\boldsymbol{x}^t) - p_e(\check{\boldsymbol{x}}^t)\|_n + \|\mathrm{diag}\left(V\mathbb{E}\partial h(Z^t)\right) h(\boldsymbol{x}^{t-1}) - \mathrm{diag}\left(W \odot W^\top \partial p_e(\check{\boldsymbol{x}}^t)\right) p_e(\check{\boldsymbol{x}}^{t-1})\|_n.$$

The first term is directly handled by the induction hypothesis as well as the bound on the spectral norm of W (see Proposition 5.5). Let us now show that the second term, which corresponds to the normalized distance between the two Onsager terms, can also be bounded by $\delta(e) + o_{\mathbb{P}}(1)$. Using the triangular inequality, this term is less than $\|\Delta^{(1)}\|_n + \|\Delta^{(2)}\|_n + \|\Delta^{(3)}\|_n + \|\Delta^{(4)}\|_n$, where

$$\begin{aligned}
\Delta^{(1)} &= \mathrm{diag}\left(V\left(\mathbb{E}\partial h(Z^t) - \mathbb{E}\partial p_e(\check{\boldsymbol{Z}}^t)\right)\right) h(\boldsymbol{x}^{t-1}),\\
\Delta^{(2)} &= \mathrm{diag}\left(V\mathbb{E}\partial p_e(\check{\boldsymbol{Z}}^t)\right)\left(h(\boldsymbol{x}^{t-1}) - p_e(\check{\boldsymbol{x}}^{t-1})\right),\\
\Delta^{(3)} &= \mathrm{diag}\left(V\left(\mathbb{E}\partial p_e(\check{\boldsymbol{Z}}^t) - \partial p_e(\check{\boldsymbol{x}}^t)\right)\right) p_e(\check{\boldsymbol{x}}^{t-1}),\\
\Delta^{(4)} &= \mathrm{diag}\left((V - W \odot W^\top)\partial p_e(\check{\boldsymbol{x}}^t)\right) p_e(\check{\boldsymbol{x}}^{t-1}).
\end{aligned}$$

For $\|\Delta^{(1)}\|_n$. We bound $|[V(\mathbb{E}\partial h(Z^t) - \mathbb{E}\partial p_e(\check{\boldsymbol{Z}}^t))]_i|$ by

$$\left|[V\mathbb{E}\partial h(Z^t) - V\mathbb{E}\partial h(\check{\boldsymbol{Z}}^t)]_i\right| + \left|[V\mathbb{E}\partial h(\check{\boldsymbol{Z}}^t) - V\mathbb{E}\partial p_e(\check{\boldsymbol{Z}}^t)]_i\right| \quad \leq \quad Ce + \delta(e), \tag{53}$$

where the last inequality is due to Lemma 5.1. The normalized norm of $h(\boldsymbol{x}^{t-1})$ can be controlled using the Lipschitz property of h and the result of Lemma 3.2.

For $\|\Delta^{(2)}\|_n$. We bound the real numbers $[V\mathbb{E}\partial p_e(\check{\boldsymbol{Z}}^t)]_i$ using inequality (53) and we conclude using the induction hypothesis.

For $\|\Delta^{(3)}\|_n$. We use Theorem 3.1-(20a) to show that $[V(\mathbb{E}\partial p_e(\check{\boldsymbol{Z}}^t) - \partial p_e(\check{\boldsymbol{x}}^t))]_i \xrightarrow[n\to\infty]{\mathbb{P}} 0$ for any sequence (i) less than (n). We then use the bounds (19) to show that $\mathbb{E}\|\Delta^{(3)}\|_n \xrightarrow[n\to\infty]{} 0$.

For $\|\Delta^{(4)}\|_n$. Finally, we use Lemma 5.3 to show that $\|\Delta^{(4)}\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Using all these bounds we finally get

$$\|\boldsymbol{x}^{t+1} - \check{\boldsymbol{x}}^{t+1}\|_n \quad \leq \quad \delta(e) + o_{\mathbb{P}}(1). \tag{54}$$

Now, it remains to show that

$$\|h(\boldsymbol{x}^{t+1}) - p_e(\check{\boldsymbol{x}}^{t+1})\|_n \quad \leq \quad \delta(e) + o_{\mathbb{P}}(1).$$

Using Lipschitz property of h as well as the bound (54), we get

$$\|h(\boldsymbol{x}^{t+1}) - p_e(\check{\boldsymbol{x}}^{t+1})\|_n \quad \leq \quad \delta(e) + o_{\mathbb{P}}(1) + \|h(\check{\boldsymbol{x}}^{t+1}) - p_e(\check{\boldsymbol{x}}^{t+1})\|_n \ .$$

Let $\varphi(x) = (h(x) - p_e(x))^2$ a continuous function with at most polynomial growth at infinity, we write

$$\|h(\check{\boldsymbol{x}}^{t+1}) - p_e(\check{\boldsymbol{x}}^{t+1})\|_n^2 = \frac{1}{n}\sum_{i\in[n]}\left(\varphi(\check{x}_i^{t+1}) - \mathbb{E}\varphi(\check{Z}_i^{t+1})\right) + \mathbb{E}\|h(\check{\boldsymbol{Z}}^{t+1}) - p_e(\check{\boldsymbol{Z}}^{t+1})\|_n^2,$$

by Lemma 3.2 the first term converges to 0 in probability, and by Lemma 5.1 the second term is bounded by e.

5.2 The Nonzero Diagonal Matrix Model

We have been working so far with a matrix S with vanishing diagonal ($S_{ii} = 0$), under A-7. In [21] and [7], this assumption simplifies the combinatorial derivations since it prevents the appearance of loops in the combinatorial structures.

In this section, we lift Assumption A-7 and prove that Theorem 2.1 holds for random matrices with nonzero diagonal elements. We proceed with a perturbation argument.

Consider a matrix X that satisfies A-1. Let $S = (s_{ij})_{1\le i,j\le n}$ be the variance profile matrix satisfying A-2 where the diagonal entries s_{ii} are non-necessarily zero. Finally, define the matrix W as in Eq. 3, i.e.

$$W_{ij} = \sqrt{s_{ij}}X_{ij}\,.$$

Let $\boldsymbol{x}^0$ and $\boldsymbol{\eta}$ two n dimensional vectors satisfying A-4, and h a function satisfying A-5 and A-6. Consider the sequence defined by

$$\left(\boldsymbol{x}^t\right)_{t\in\mathbb{N}} := \text{AMP-Z}\left(X, S, h, \boldsymbol{x}^0, \boldsymbol{\eta}\right)\,.$$

We remind below the iteration expression:

$$\boldsymbol{x}^{t+1} = Wh\left(\boldsymbol{x}^t, \boldsymbol{\eta}, t\right) - \text{diag}\left(V\mathbb{E}\left[\partial h(\boldsymbol{Z}^t, \boldsymbol{\eta}, t)\right]\right)h(\boldsymbol{x}^{t-1}, \boldsymbol{\eta}, t-1)\,,$$

where $V = (v_{ij}) = (\tau_{ij}\sqrt{s_{ij}s_{ji}})$ and $(Z^1, \cdots, Z^t) \sim \text{DE}\left(h, \boldsymbol{x}^0, S, t\right)$.

In order to proceed, define $\tilde{S}$ to be equal to S except the diagonal elements that we set to zero;

$$\tilde{s}_{ij} = (1-\delta_{ij})s_{ij}\,.$$

Define matrix $\widetilde{W}$ by $\widetilde{W}_{ij} = \sqrt{\tilde{s}_{ij}}X_{ij}$, and the $\mathbb{R}^n$-valued sequences $\left(\tilde{\boldsymbol{x}}^t\right)_{t\in\mathbb{N}}$ by

$$\left(\tilde{\boldsymbol{x}}^t\right)_{t\in\mathbb{N}} := \text{AMP-Z}\left(X, \widetilde{S}, h, \boldsymbol{x}^0, \boldsymbol{\eta}\right),$$

where the iterations are given by

$$\tilde{\boldsymbol{x}}^{t+1} = \widetilde{W}h\left(\tilde{\boldsymbol{x}}^t, \boldsymbol{\eta}, t\right) - \text{diag}\left(\widetilde{V}\mathbb{E}\left[\partial h(\tilde{Z}^t, \boldsymbol{\eta}, t)\right]\right)h(\tilde{\boldsymbol{x}}^{t-1}, \boldsymbol{\eta}, t-1)\,.$$

Here $\widetilde{V} = \left(\widetilde{S} \odot \widetilde{S}^\top\right)^{\odot 1/2} \odot T = \left((1-\delta_{ij})v_{ij}\right)$ and $(\tilde{Z}^1, \cdots, \tilde{Z}^t) \sim \mathrm{DE}\left(h, \boldsymbol{x}^0, \tilde{S}, t\right)$.

Since this sequence is generated using a matrix model with vanishing diagonal, we can apply the AMP result proven so far, i.e. for every uniformly bounded sequence $(\beta_i)_{i\in[n]}$ and every PL test function $\varphi : \mathbb{R}^{t_{\max}+1} \to \mathbb{R}$, we have

$$\frac{1}{n}\sum_{i\in[n]} \beta_i \varphi(\eta_i, \tilde{x}_i^1, \cdots, \tilde{x}_i^{t_{\max}}) - \beta_i \varphi(\eta_i, \tilde{Z}_i^1, \cdots, \tilde{Z}_i^{t_{\max}}) \quad \xrightarrow[n\to\infty]{\mathbb{P}} \quad 0\,.$$

In order to prove the same convergence result for $(\boldsymbol{x}^t)_{t\in\mathbb{N}}$, we prove that $\boldsymbol{x}^t$ is a small perturbation of $\tilde{\boldsymbol{x}}^t$ as n grows to infinity.

Lemma 5.6 *For each $i \in [n]$ and $t \le t_{\max}$ recall that R_i^t (respectively $\tilde{R}_i^t$) is the covariance matrix of $\vec{Z}_i^t := [Z_i^1, \cdots, Z_i^t]^\top$ (respectively $\vec{\tilde{Z}}_i^t$). Then $\|R_i^t - \tilde{R}_i^t\|$ converges to 0 as n grows to infinity.*

Proof We prove this result by induction on t. For $t = 1$ we write:

$$R_i^1 - \tilde{R}_i^1 = \sum_{\ell\in[n]} s_{i\ell}\left(h(x_\ell^0, \eta_\ell, 0)\right)^2 - \sum_{\ell\in[n]\,:\,\ell\neq i} s_{i\ell}\left(h(x_\ell^0, \eta_\ell, 0)\right)^2 = s_{ii}\left(h(x_i^0, \eta_i, 0)\right)^2 .$$

Hence

$$\left|R_i^1 - \tilde{R}_i^1\right| \;\le\; \frac{C}{K_n} \quad \xrightarrow[n\to\infty]{} \quad 0\,.$$

Suppose now that for all $s \le t$ the quantity $\|R_i^s - \tilde{R}_i^s\|$ converges to zero and let us now prove that this convergence also holds at iteration step $t+1$. To this end, we must study the $(t+1, s+1)$-th entry of the $(t+1)\times(t+1)$ of the covariance matrices R_i^{t+1} and $\tilde{R}_i^{t+1}$. We have

$$\begin{aligned} &R_i^{t+1}(t+1, s+1) - \tilde{R}_i^{t+1}(t+1, s+1) \\ &\quad= \sum_{\ell\in[n]\,:\,\ell\neq i} s_{i\ell}\left(\mathbb{E}\left[h(Z_\ell^t, \eta_\ell, t)h(Z_\ell^s, \eta_\ell, s)\right] - \mathbb{E}\left[h(\tilde{Z}_\ell^t, \eta_\ell, t)h(\tilde{Z}_\ell^s, \eta_\ell, s)\right]\right) \\ &\quad+ s_{ii}\mathbb{E}\left[h(Z_i^t, \eta_i, t)h(Z_i^s, \eta_i, s)\right]\,. \end{aligned} \tag{55}$$

Using the fact that $\mathbb{E}\left[h(Z_i^s, \eta_i, s)^2\right]$ is bounded by a constant that depends only on t and using Cauchy-Schwartz inequality, we have

$$\left|s_{ii}\mathbb{E}\left[h(Z_i^t, \eta_i, t)h(Z_i^s, \eta_i, s)\right]\right| \le \frac{C}{K_n}\,.$$

In order to bound the first term of the right-hand side of Eq. (55), first notice that since h is Lipschitz then $H : (x_1, x_2) \mapsto h(x_1)h(x_2)$ is PL, i.e. there exists $C > 0$ such that

$$\forall x, y \in \mathbb{R}^2 \quad |H(x) - H(y)| \le C\|x - y\|_2\,(1 + \|x\|_2 + \|y\|_2)\,.$$

Let $\Sigma^2 \in \mathbb{R}^{2\times 2}$ and $\tilde{\Sigma}^2 \in \mathbb{R}^{2\times 2}$ be the covariance matrices of the vectors $Z_\ell^{t,s} = (Z_\ell^t, Z_\ell^s)$ and $\tilde{Z}_\ell^{t,s} = (\tilde{Z}_\ell^t, \tilde{Z}_\ell^s)$, respectively. Then given $\xi \sim \mathcal{N}_2(0, I_2)$ we can write

$$\left|\mathbb{E}\left[H(Z_\ell^{t,s}) - H(\tilde{Z}_\ell^{t,s})\right]\right| = \left|\mathbb{E}\left[H(\Sigma\xi) - H(\tilde{\Sigma}\xi)\right]\right| \\ \leq C\|\Sigma - \tilde{\Sigma}\|\mathbb{E}\left[\|\xi\|_2\left(1 + \|Z_\ell^{t,s}\|_2 + \|\tilde{Z}_\ell^{t,s}\|_2\right)\right].$$

Using Lemma 5.2 it is easy to see that the factor

$$\mathbb{E}\left[\|\xi\|_2\left(1 + \|Z_\ell^{t,s}\|_2 + \|\tilde{Z}_\ell^{t,s}\|_2\right)\right].$$

is bounded by a constant depending only on $t_{\max}$. Now using the induction hypothesis we obtain the following inequality:

$$\|\Sigma - \tilde{\Sigma}\| \leq \|\Sigma^2 - \tilde{\Sigma}^2\|^{1/2} \leq \|R_\ell^t - \tilde{R}_\ell^t\|^{1/2} \xrightarrow[n\to\infty]{} 0$$

Here we used the fact that the matrix squared root is $1/2$-Hölder continuous on the set of symmetric positive matrices, the proof in in Appendix 1. Note that by A-2 we have $s_{ij} \leq C_S K_n^{-1}$, plugging this into (55) gives the desired result. □

Remark 5.1 Notice that we can also specify the convergence rate of $\|R_i^t - \tilde{R}_i^t\|$ to 0. In fact we can show that

$$\|R_i^t - \tilde{R}_i^t\| \leq \frac{C}{K_n^{1/2^t}}.$$

Proof of Theorem 2.1 in the General Case

We begin by proving the following convergence by induction on t,

$$\|\boldsymbol{x}^t - \tilde{\boldsymbol{x}}^t\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0\,. \tag{56}$$

For $t = 1$, knowing that the x_i^0's live on a compact $\mathcal{Q}_x$ we get

$$\|\boldsymbol{x}^1 - \tilde{\boldsymbol{x}}^1\|_n^2 = \|(W - \tilde{W})h(\boldsymbol{x}^0)\|_n^2 = \frac{1}{n}\sum_{i=1}^n s_{ii} X_{ii}^2 h(x_i^0)^2 \leq \frac{C}{K_n}\left(\frac{\sum_{i=1}^n X_{ii}^2}{n}\right), \tag{57}$$

thus $\|\boldsymbol{x}^1 - \tilde{\boldsymbol{x}}^1\|_n^2 \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Now assume that this holds for all $s \in \{1, \cdots, t\}$ and let us show that it is also satisfied for $t + 1$, i.e.

$$\|\boldsymbol{x}^{t+1} - \tilde{\boldsymbol{x}}^{t+1}\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0\,.$$

Let us write the difference between $\boldsymbol{x}^{t+1}$ and $\tilde{\boldsymbol{x}}^{t+1}$,

$$\begin{aligned}\boldsymbol{x}^{t+1} - \tilde{\boldsymbol{x}}^{t+1} &= Wh(\boldsymbol{x}^t) - \tilde{W}h(\tilde{\boldsymbol{x}}^t) \\ &\quad + \operatorname{diag}\left(V\mathbb{E}\left[\partial h(\boldsymbol{Z}^t)\right]\right) h(\boldsymbol{x}^{t-1}) - \operatorname{diag}\left(\tilde{V}\mathbb{E}\left[\partial h(\tilde{\boldsymbol{Z}}^t)\right]\right) h(\tilde{\boldsymbol{x}}^{t-1})\,,\end{aligned}$$

We first show that

$$\|Wh(\boldsymbol{x}^t) - \tilde{W}h(\tilde{\boldsymbol{x}}^t)\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0\,.$$

We have

$$\|Wh(\boldsymbol{x}^t) - \tilde{W}h(\tilde{\boldsymbol{x}}^t)\|_n \leq \|(W - \tilde{W})h(\tilde{\boldsymbol{x}}^t)\|_n + \|W(h(\boldsymbol{x}^t) - h(\tilde{\boldsymbol{x}}^t))\|_n \tag{58}$$

Using the fact that the $\tilde{x}_i^t$ are bounded by a constant $C = C(t)$ independent of n we can directly see that the first term of (58) converges to zero. For the second term, we use the bound on $\|W\|$ (see Proposition 5.5) as well as the Lipschitz property of h and the induction hypothesis.

Now let us study the term

$$\operatorname{diag}\left(V\mathbb{E}\left[\partial h(\boldsymbol{Z}^t)\right]\right) h(\boldsymbol{x}^{t-1}) - \operatorname{diag}\left(\tilde{V}\mathbb{E}\left[\partial h(\tilde{\boldsymbol{Z}}^t)\right]\right) h(\tilde{\boldsymbol{x}}^{t-1})\,. \tag{59}$$

This term can be decomposed as follows

$$\begin{aligned}&\operatorname{diag}\left((V - \tilde{V})\mathbb{E}\left[\partial h(\boldsymbol{Z}^t)\right]\right) h(\boldsymbol{x}^{t-1}) \\ &+ \operatorname{diag}\left(\tilde{V}\mathbb{E}\left[\partial h(\boldsymbol{Z}^t) - \partial h(\tilde{\boldsymbol{Z}}^t)\right]\right) h(\boldsymbol{x}^{t-1}) \\ &+ \operatorname{diag}\left(\tilde{V}\mathbb{E}\left[\partial h(\tilde{\boldsymbol{Z}}^t)\right]\right)\left(h(\boldsymbol{x}^{t-1}) - h(\tilde{\boldsymbol{x}}^{t-1})\right) \\ &:= \Delta_1 + \Delta_2 + \Delta_3\,.\end{aligned}$$

Using the Lipschitz property of h we can bound $\|\Delta_3\|_n^2$ as follows:

$$\begin{aligned}\|\Delta_3\|_n &= \left\|\operatorname{diag}\left(\tilde{V}\mathbb{E}\left[\partial h(\tilde{\boldsymbol{Z}}^t)\right]\right)\left(h(\boldsymbol{x}^{t-1}) - h(\tilde{\boldsymbol{x}}^{t-1})\right)\right\|_n \\ &\leq \left\|\operatorname{diag}\left(\tilde{V}\mathbb{E}\left[\partial h(\tilde{\boldsymbol{Z}}^t)\right]\right)\right\| \|h(\boldsymbol{x}^{t-1}) - h(\tilde{\boldsymbol{x}}^{t-1})\|_n \\ &\leq C \max_{j\in[n]}\left\{\mathbb{E}\left|\partial h(\tilde{\boldsymbol{Z}}_j^t)\right|\right\} \|\boldsymbol{x}^{t-1} - \tilde{\boldsymbol{x}}^{t-1}\|_n\,.\end{aligned}$$

Recall that $\max_{j\in[n]}\left\{\mathbb{E}\left|\partial h(\tilde{\boldsymbol{Z}}_j^t)\right|\right\}$ is bounded by $C = C(t)$, using the induction hypothesis we prove that $\|\Delta_3\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

In order to bound the first term $\|\Delta_1\|_n$, notice that $V-\tilde{V}$ is a diagonal matrix whose entries are bounded by C/K_n, thus

$$\left\|\operatorname{diag}\left((V-\tilde{V})\mathbb{E}\left[\partial h(\boldsymbol{Z}^t)\right]\right)\right\| \leq \frac{C}{K_n}\max_{i\in[n]}\left\{\mathbb{E}\left|\partial h(\boldsymbol{Z}_i^t)\right|\right\} = \mathcal{O}\left(\frac{1}{K_n}\right),$$

where the last equality is by the boundness of $\max_{i\in[n]}\left\{\mathbb{E}\left|\partial h(\boldsymbol{Z}_i^t)\right|\right\}$. Now write

$$h(\boldsymbol{x}^{t-1}) = \left(h(\boldsymbol{x}^{t-1}) - h(\tilde{\boldsymbol{x}}^{t-1})\right) + h(\tilde{\boldsymbol{x}}^{t-1}),$$

by the induction hypothesis we clearly see that $\|h(\boldsymbol{x}^{t-1}) - h(\tilde{\boldsymbol{x}}^{t-1})\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$, in addition we know that $\|h(\tilde{\boldsymbol{x}}^{t-1})\|_n^2 - \mathbb{E}\left\|h(\tilde{Z}^{t-1})\right\|_n^2 \xrightarrow[n\to\infty]{\mathbb{P}} 0$ so by bounding $\mathbb{E}\left\|h(\tilde{Z}^{t-1})\right\|_n^2$ we get that the probability of $\|h(\boldsymbol{x}^{t-1})\|_n$ not being bounded converges to 0. Finally $\|\Delta_1\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

For $\|\Delta_2\|_n$, we use Lemma 5.6 to bound $\left\|\operatorname{diag}\left(\tilde{V}\mathbb{E}\left[\partial h(\boldsymbol{Z}^t) - \partial h(\tilde{\boldsymbol{Z}}^t)\right]\right)\right\|$ by C/K_n and finally get $\|\Delta_2\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$. To sum up, we have proved that the difference between the two Onsager terms (59) has a normalized norm converging to 0. Finally, we have proved (57) by induction, i.e. $\tilde{\boldsymbol{x}}^t$ asymptotically approximates $\boldsymbol{x}^t$ in terms of normalized norm. Now we are able to use the convergence result of the sequence $(\tilde{\boldsymbol{x}}^t)_t$ to prove the convergence of $\tilde{\boldsymbol{x}}^t$ as n grows to ∞. Let $\varphi : \mathbb{R}^{t_{\max}} \to \mathbb{R}$ be a pseudo-Lipschitz function and denote $\boldsymbol{x}_i = \left(x_i^1, \cdots, x_i^{t_{\max}}\right)^\top$ and $\tilde{\boldsymbol{x}}_i = \left(\tilde{x}_i^1, \cdots, \tilde{x}_i^{t_{\max}}\right)^\top$, and without loss of generality we omit the scalars β_i and the parameters η_i by considering that φ depends also on the index i. We have

$$\begin{aligned}
&\frac{1}{n}\left|\sum_{i=1}^n \varphi(\boldsymbol{x}_i) - \varphi(\vec{Z}_i^t)\right| \\
&\quad\leq \frac{1}{n}\left|\sum_{i=1}^n \varphi(\boldsymbol{x}_i) - \varphi(\tilde{\boldsymbol{x}}_i)\right| \\
&\quad\quad + \frac{1}{n}\left|\sum_{i=1}^n \varphi(\tilde{\boldsymbol{x}}_i) - \varphi(\vec{\tilde{Z}}_i^t)\right| \\
&\quad\quad + \frac{1}{n}\left|\sum_{i=1}^n \varphi(\vec{\tilde{Z}}_i^t) - \varphi(\vec{Z}_i^t)\right| \\
&\quad=: \Theta_1 + \Theta_2 + \Theta_3\,.
\end{aligned}$$

Using the pseudo-Lipschitz property of φ we get the following

$$\begin{aligned}\Theta_1 &\leq \frac{C}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i-\tilde{\boldsymbol{x}}_i\|\,(1+\|\boldsymbol{x}_i\|+\|\tilde{\boldsymbol{x}}_i\|)\\ &\leq \frac{C}{n}\left(\sum_{t=1}^{t_{\max}}\|\boldsymbol{x}^t-\tilde{\boldsymbol{x}}^t\|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n}(1+\|\boldsymbol{x}_i\|+\|\tilde{\boldsymbol{x}}_i\|)^2\right)^{\frac{1}{2}}\\ &\leq \frac{C}{n}\left(\sum_{t=1}^{t_{\max}}\|\boldsymbol{x}^t-\tilde{\boldsymbol{x}}^t\|\right)\left(n+\sum_{t=1}^{t_{\max}}\|\boldsymbol{x}^t\|^2+\|\tilde{\boldsymbol{x}}^t\|^2\right)^{\frac{1}{2}}\\ &\leq C\left(\sum_{t=1}^{t_{\max}}\|\boldsymbol{x}^t-\tilde{\boldsymbol{x}}^t\|_n\right)\left(1+\sum_{t=1}^{t_{\max}}\|\boldsymbol{x}^t\|_n+\sum_{t=1}^{t_{\max}}\|\tilde{\boldsymbol{x}}^t\|_n\right).\end{aligned}$$

Then, by using (57) we get $\Theta_1 \xrightarrow[n\to\infty]{\mathbb{P}} 0$. The term Θ_2 converges to 0 in probability by Theorem 2.1 applied with zero diagonal matrix model. As for Θ_3 we use the pseudo-Lipschitz property of φ as well as Lemma 5.6. This ends the proof for Theorem 2.1.

Appendix A. Proof of Theorem 2.3

We prove here the AMP result for non-centered matrices described in Theorem 2.3.

We follow the general idea described in [18], which is to reduce the problem to an AMP with centered random matrix model and apply Theorem 2.1. To this end, write the following,

$$\begin{aligned}\boldsymbol{x}^{t+1} &= \lambda\left\langle \boldsymbol{v}, h_t(\boldsymbol{x}^t,\boldsymbol{\eta})\right\rangle \boldsymbol{u} + W h_t(\boldsymbol{x}^t,\boldsymbol{\eta}) - \operatorname{diag}\left(V\mathbb{E}\partial h_t(Z^t+\mu_t\boldsymbol{u},\boldsymbol{\eta})\right) h_{t-1}(\boldsymbol{x}^{t-1},\boldsymbol{\eta})\\ &= \mu_{t+1}\boldsymbol{u} + W h_t(\boldsymbol{x}^t,\boldsymbol{\eta}) - \operatorname{diag}\left(V\mathbb{E}\partial h_t(Z^t+\mu_t\boldsymbol{u},\boldsymbol{\eta})\right) h_{t-1}(\boldsymbol{x}^{t-1},\boldsymbol{\eta}) + \delta_{t+1}\boldsymbol{u},\end{aligned}$$

where $\delta_t := \lambda\left\langle \boldsymbol{v}, h_{t-1}(\boldsymbol{x}^{t-1},\boldsymbol{\eta})\right\rangle - \mu_t$. One should think of $\delta_{t+1}\boldsymbol{u}$ as an error term, we will show later that this term has a negligible effect. Define now the following sequence $\left(\tilde{\boldsymbol{y}}^t\right)_{t\in\mathbb{N}}$ as follows,

$$\tilde{\boldsymbol{y}}^0 = \boldsymbol{x}^0 \quad \text{and} \quad \tilde{\boldsymbol{y}}^t := \boldsymbol{x}^t - \mu_t\boldsymbol{u} \quad \text{for } t\geq 1,$$

this sequence satisfies the following recursion,

$$\tilde{\boldsymbol{y}}^{t+1} = W g_t(\tilde{\boldsymbol{y}}^t,\boldsymbol{v},\boldsymbol{\eta}) - \operatorname{diag}\left(V\mathbb{E}\partial g_t(\boldsymbol{Z}^t,\boldsymbol{v},\boldsymbol{\eta})\right) g_{t-1}(\tilde{\boldsymbol{y}}^{t-1},\boldsymbol{v},\boldsymbol{\eta}) + \delta_{t+1}\boldsymbol{v}, \quad (60)$$

where the function $g_t(x,v,\eta)$ with parameters v and η is given by,

$$g_t(x,v,\eta) := h_t(x+\lambda v,\eta) \quad \forall x\in\mathbb{R}.$$

One can clearly see that this function satisfies the same assumptions as h_t. Now define the following AMP algorithm $(\boldsymbol{y}^t)_{t\in\mathbb{N}}$ by

$$\begin{cases} \boldsymbol{y}^0 & = \boldsymbol{x}^0, \\ \boldsymbol{y}^{t+1} & = W g_t(\boldsymbol{y}^t, \boldsymbol{v}, \boldsymbol{\eta}) - \operatorname{diag}\left(V\mathbb{E}\partial g_t(\boldsymbol{Z}^t, \boldsymbol{v}, \boldsymbol{\eta})\right) g_{t-1}(\boldsymbol{y}^{t-1}, \boldsymbol{v}, \boldsymbol{\eta}) \,, \end{cases} \tag{61}$$

where

$$\left(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t\right) \sim \widetilde{\mathrm{DE}}\left(h, \boldsymbol{x}^0, S, t, \boldsymbol{u}, \boldsymbol{v}\right) \,,$$

in the sense of Definition 2.11. A key observation is that

$$\left(\boldsymbol{Z}^1, \cdots, \boldsymbol{Z}^t\right) \sim \mathrm{DE}\left(g, \boldsymbol{x}^0, S, t\right) \,.$$

Hence Theorem 2.1 applies for the recursion (61) and yields that for any pseudo-Lipschitz test function $\varphi : \mathbb{R}^{t+1} \to \mathbb{R}$ it holds that

$$\frac{1}{n}\sum_{i=1}^{n} \beta_i \varphi\left(\eta_i, y_i^1, \cdots, y_i^t\right) - \beta_i \mathbb{E}\left[\varphi\left(\eta_i, Z_i^1, \cdots, Z_i^t\right)\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0 \,. \tag{62}$$

In order to prove our result, it suffices to show that the error term $\delta_{t+1}\boldsymbol{u}$ in Eq. (60) is negligible and that for all t one has $\boldsymbol{y}^t \approx \tilde{\boldsymbol{y}}^t$. To this end, we want to prove by induction on t that,

$$\delta_t \xrightarrow[n\to\infty]{\mathbb{P}} 0 \quad \text{and} \quad \|\tilde{\boldsymbol{y}}^t - \boldsymbol{y}^t\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \text{for all } t \geq 1. \tag{63}$$

For $t = 1$, we have $\delta_1 = 0$ and $\tilde{\boldsymbol{y}}^1 = \boldsymbol{y}^1$. Suppose that (63) is true for t, and let us prove that this remains true for $t+1$ as well. Let us begin with δ_{t+1}. We have the following

$$\begin{aligned} \delta_{t+1} &= \lambda \sum_{i\in[n]} v_i \left(g_t(\tilde{y}_i^t) - \mathbb{E}g_t(Z_i^t)\right) \\ &= \lambda \sum_{i\in[n]} v_i \left(g_t(\tilde{y}_i^t) - g_t(y_i^t)\right) + \lambda \sum_{i\in[n]} v_i \left(g_t(y_i^t) - \mathbb{E}g_t(Z_i^t)\right) \\ &:= T_1 + T_2 \,. \end{aligned}$$

Using the Lipschitz property of the function g_t as well as the induction hypothesis, namely, $\|\tilde{\boldsymbol{y}}^t - \boldsymbol{y}^t\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$ we directly get that $T_1 \xrightarrow[n\to\infty]{\mathbb{P}} 0$. As for the second term, $T_2 \xrightarrow[n\to\infty]{\mathbb{P}} 0$ is a direct application of Theorem 2.1, i.e. Eq. (62).

It remains to show that $\|\tilde{\boldsymbol{y}}^{\boldsymbol{t+1}} - \boldsymbol{y}^{t+1}\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Using the recursive definition of $\left(\tilde{\boldsymbol{y}}^t\right)_t$ and $\left(\boldsymbol{y}^t\right)_t$ in (60) and (61) we can write the following;

$$\tilde{\boldsymbol{y}}^{\boldsymbol{t+1}} - \boldsymbol{y}^{t+1} = W\left(g_t(\tilde{\boldsymbol{y}}) - g_t(\boldsymbol{y}^t)\right) - \operatorname{diag}\left(V\mathbb{E}\partial g_t(\boldsymbol{Z}^t)\right)$$
$$\left(g_{t-1}(\tilde{\boldsymbol{y}}^{\boldsymbol{t-1}}) - g_{t-1}(\boldsymbol{y}^{\boldsymbol{t-1}})\right) + \delta_{t+1}\boldsymbol{u}\,.$$

The normalized norm of the first term can be easily handled using the Lipschitz property of the function g_t as well as the induction hypothesis, we also use Proposition 5.5 which ensures the boundness of the spectral norm $\|W\|$. As for the second term, we similarly show that the quantity $\|g_{t-1}(\tilde{\boldsymbol{y}}^{t-1}) - g_{t-1}(\boldsymbol{y}^{t-1})\|_n$ vanishes, in probability. It remains to show that $\|\operatorname{diag}\left(V\mathbb{E}\partial g_t(\boldsymbol{Z}^t)\right)\|$ is bounded as n goes to infinity, this clearly holds as ∂g_t is the derivative of a Lipschitz function and thus is bounded.

Finally, we have proved that $\|\tilde{\boldsymbol{y}}^{\boldsymbol{t+1}} - \boldsymbol{y}^{t+1}\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$ which ends the induction argument. Using (63) and the AMP result of the sequence $\left(\boldsymbol{y}^t\right)_t$ we directly deduce an AMP result of the sequence $\left(\tilde{\boldsymbol{y}}^t\right)_t$.

Appendix B. Elements of Proof of Lemma 3.2

Lemma B.1 *Let (m_n) and (σ_n^2) be two bounded sequences and let (ν_n) be the sequence of Gaussian measures with means m_n and variances σ_n^2. Let (μ_n) be any sequence of probability measures such that the following holds for each $k \in \mathbb{N}$,*

$$\int x^k d\mu_n - \int x^k d\nu_n \xrightarrow[n\to\infty]{} 0\,. \tag{64}$$

Then for any continuous function $\psi : \mathbb{R} \to \mathbb{R}$ such that $|\varphi(x)| \leq C(1 + |x|^m)$ for some constant $C > 0$ and some integer m we have

$$\int \psi(x) d\mu_n - \int \psi(x) d\nu_n \xrightarrow[n\to\infty]{} 0\,. \tag{65}$$

Proof First, it is sufficient to show that from any subsequence of (n) we can extract a further subsequence such that the convergence in (65) holds along this subsequence. So without loss of generality we only prove that if (64) holds along the sequence (n) then there exists a subsequence of (n) along which (65) holds.

The sequence of probability measures (ν_n) is tight because (m_n) and (σ_n^2) are bounded, thus we can extract a subsequence of (n), which also be denoted as (n), such that (ν_n) converges weakly to a probability measure ν. Consider now the moment generating function Φ_{ν_n} of ν_n defined on $\mathbb{R}$ as follows,

$$\Phi_{\nu_n}(t) = \int e^{tx} d\nu_n(x) = \exp(m_n t + \sigma_n^2 t^2/2), \quad t \in \mathbb{R}.$$

This function can be viewed as a restriction to the real line of the following holomorphic function

$$\Phi_{\nu_n}(z) = \int e^{zx} d\nu_n(x) = \exp(m_n z + \sigma_n^2 z^2/2), \quad z \in \mathbb{C}.$$

Notice that the sequence (Φ_{ν_n}) is uniformly bounded on compact sets of $\mathbb{C}$, thus there exists a holomorphic function Φ and a subsequence of (n) such that (Φ_{ν_n}) converges uniformly to Φ on compact sets. This implies the pointwise convergence of the moment generating function $(\Phi_{\nu_n}(t))$ to $\Phi(t)$ so by a convergence result in [14, Theorem 3] and the uniqueness of the weak limit, we get $\Phi(t) = \Phi_\nu(t)$. The convergence of $(\Phi_{\nu_n}(t))$ to $\Phi_\nu(t)$ implies the convergence of the moments, and by (64) we get

$$\int x^k d\mu_n \xrightarrow[n\to\infty]{} \int x^k d\nu, \tag{66}$$

we also know that Φ_ν characterizes ν [14, Theorem 1], thus ν is determined by its moments, so (μ_n) converges weakly to ν. Let ψ be a function as in the lemma and let X_n and X be random variables with distributions μ_n and ν, respectively, we want to prove that $\mathbb{E}[\psi(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[\psi(X)]$, this follows from the convergence in distribution of $(\psi(X_n))$ to $\psi(X)$ and the uniform integrability of $(\psi(X_n))$. The latter is due the following observation

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[(\psi(X_n))^2\right] \le C^2 \sup_{n\in\mathbb{N}} \mathbb{E}\left[(1+|X_n|^m)^2\right] = C^2 \sup_{n\in\mathbb{N}} \int (1+|x|^m)^2 d\mu_n(x) < \infty .$$

The last inequality is due to the convergence of the moments (66). □

Remark B.1 Results of Lemma B.1 can be extended to probability measures μ on $\mathbb{R}^d$ by Cramér-Wold theorem, i.e. considering the push-forward probability measure μ_t by the map $x \mapsto \langle x, t\rangle$ for each $t \in \mathbb{R}^d$.

Remark B.2 We can also extend Lemma B.1 to the case where (μ_n) and (ν_n) are sequences of random probability measure and where we replace both two convergence statements by convergence in probability formulations. The proof follows from the subsequence criterion [24, Lemma 3.2].

Appendix C. Polynomial Approximation

The following lemma states a basic density result of polynomial functions in the Hilbert space $L^2(\mu)$ where μ is a Gaussian measure. The polynomial approximation is shown to hold uniformly on certain sets of Gaussian measures $(\mu_\sigma)_{\sigma\in S}$.

Lemma C.1 *[21] Let $\mathcal{Q} \subset \mathbb{R}$ a compact set and $h : \mathbb{R}\times\mathcal{Q} \to \mathbb{R}$ a function satisfying the following properties. (i) There exists a fixed number $L > 0$ such that uniformly in $\eta \in \mathcal{Q}$,*

$$|h(x,\eta) - h(y,\eta)| \le L|x-y|, \qquad \forall (x,y) \in \mathbb{R}^2 .$$

(ii) There exists a continuous non-decreasing function $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ with $\kappa(0) = 0$ such that

$$|h(x,\eta) - h(x,\eta')| \le \kappa(|\eta-\eta'|)\,(1+|x|), \qquad \forall x \in \mathbb{R},\ \forall (\eta,\eta') \in \mathcal{Q}^2 .$$

Let $0 < \sigma_{\min} \leq \sigma_{\max}$ *and* $\varepsilon > 0$ *be fixed, and* $\xi \sim \mathcal{N}(0,1)$.

There exists a function $g_\varepsilon : \mathbb{R} \times \mathcal{Q} \to \mathbb{R}$ *such that for every* $\eta \in \mathcal{Q}$, $x \mapsto g_\varepsilon(x,\eta)$ *is a polynomial, and uniformly in* $\eta \in \mathcal{Q}$ *and* $\sigma \in [\sigma_{\min}, \sigma_{\max}]$,

$$\mathbb{E}\left(h(\sigma\xi,\eta) - g_\varepsilon(\sigma\xi,\eta)\right)^2 \leq \varepsilon \quad \textit{and} \quad \left|\mathbb{E}\,\partial_x h(\sigma\xi,\eta) - \mathbb{E}\,\partial_x g_\varepsilon(\sigma\xi,\eta)\right| \leq \varepsilon\,.$$

Proof Let $\delta > 0$ and consider a δ-covering of the compact set $\mathcal{Q}$ with balls centered in $\{\eta_k\}_{k\in[K]}$. Fix $k \in [K]$ and consider the function $x \mapsto h(x,\eta_k)$. By the density of polynomials in the space $L^2(\mathcal{N}(0,\sigma_{\max}^2))$, there exists a polynomial $x \mapsto g_\varepsilon(x,\eta_k)$ such that

$$\mathbb{E}\left(h(\sigma_{\max}\xi,\eta_k) - g_\varepsilon(\sigma_{\max}\xi,\eta_k)\right)^2 \quad \leq \quad \frac{\varepsilon}{4}\,.$$

Let $\eta \in \mathcal{Q}$ and η_k such that $|\eta - \eta_k| \leq \delta$ and put $g_\varepsilon(x,\eta) := g_\varepsilon(x,\eta_k)$ for such η. By the properties of function h, we have

$$\begin{aligned}\mathbb{E}\left(h(\sigma_{\max}\xi,\eta) - g_\varepsilon(\sigma_{\max}\xi,\eta)\right)^2 &\leq 2\mathbb{E}\left(h(\sigma_{\max}\xi,\eta) - h(\sigma_{\max}\xi,\eta_k)\right)^2 \\ &\quad + 2\mathbb{E}\left(h(\sigma_{\max}\xi,\eta_k) - g_\varepsilon(\sigma_{\max}\xi,\eta_k)\right)^2, \\ &\leq 2L^2\kappa(\delta)^2\mathbb{E}\left(1 + \sigma_{\max}|\xi|\right)^2 + \frac{\varepsilon}{2}.\end{aligned}$$

Using the properties of κ we can choose $\delta > 0$ small enough so that

$$\mathbb{E}\left(h(\sigma_{\max}\xi,\eta) - g_\varepsilon(\sigma_{\max}\xi,\eta)\right)^2 \leq \varepsilon\,.$$

Let $\sigma \in [\sigma_{\min}, \sigma_{\max}]$, denote $\varphi(x) := h(x,\eta) - g_\varepsilon(x,\eta)$. A change of variable yields

$$\mathbb{E}\varphi(\sigma\xi)^2 \quad \leq \quad \frac{\sigma_{\max}}{\sigma_{\min}}\mathbb{E}\varphi(\sigma_{\max}\xi)^2 \quad \leq \quad \frac{\sigma_{\max}}{\sigma_{\min}}\varepsilon\,.$$

By Stein's integration by parts lemma we also have

$$\left|\mathbb{E}\varphi'(\sigma\xi)\right| \quad = \quad \frac{1}{\sigma}\mathbb{E}[\xi\varphi(\sigma\xi)] \quad \leq \quad \frac{1}{\sigma_{\min}}\sqrt{\mathbb{E}\varphi(\sigma\xi)^2} \quad \leq \quad \sqrt{\frac{\sigma_{\max}}{(\sigma_{\min})^3}}\sqrt{\varepsilon}\,,$$

which concludes the proof. □

Appendix D Proof of Lemma 5.3

Proof of Lemma 5.3 In this proof, we use the framework introduced in Sect. 4.2. Let us put $p_j := \partial p(\check{x}_j^t, \eta_j, t)$ as a simplification of the notations, the expectation can be developed as follows,

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{j\in[n]}\left(W_{ij}W_{ji}-V_{ij}\right)p_j\right)^4\right]\\
&\quad=\sum_{j_1,j_2,j_3,j_4\in[n]}\mathbb{E}\left[\left(\prod_{\ell=1}^{4}\left(W_{ij_\ell}W_{j_\ell i}-V_{ij_\ell}\right)\right)p_{j_1}p_{j_2}p_{j_3}p_{j_4}\right]\\
&\quad:=\sum_{j_1,j_2,j_3,j_4\in[n]}\mathbb{E}\varphi(j_1,j_2,j_3,j_4)\,,
\end{aligned}$$

with p_j having the following form

$$p_j=\sum_{\ell=0}^{d-1}(1+\ell)\alpha_\ell(j,t)\left(\check{x}_j^t\right)^\ell,$$

notice now that by using Lemma 4.6, we can easily see p_j as a sum over unmarked trees with root type j, with depth at most t and with each vertex having at most $d-1$ children, the weight of the trees (i.e. the terms $W(T)$, $\tilde{\Gamma}(T)$ and $x(T)$) are the same as in Lemma 4.6.

$$p_j=\sum_{T\in\bar{\mathcal{U}}_j^t}W(T)\tilde{\Gamma}(T)x(T)\,.$$

Thus, the quantity $\varphi(j_1,j_2,j_3,j_4)$ above can be written as a sum over trees as follows:

$$\begin{aligned}
\varphi(j_1,j_2,j_3,j_4)&=\sum_{\substack{(T_1,T_2,T_3,T_4)\in\\ \bar{\mathcal{U}}_{j_1}^t\times\bar{\mathcal{U}}_{j_2}^t\times\bar{\mathcal{U}}_{j_3}^t\times\bar{\mathcal{U}}_{j_4}^t}}\psi(T_1,T_2,T_3,T_4),\\
\psi(T_1,T_2,T_3,T_4)&:=\prod_{\ell=1}^{4}\left(W_{ij_\ell}W_{j_\ell i}-V_{ij_\ell}\right)W(T_\ell)\tilde{\Gamma}(T_\ell)x(T_\ell)\,.
\end{aligned}\tag{67}$$

In the case where j_1, j_2, j_3 and j_4 are distinct, the above sum can interpreted as a sum over trees having the structure described in Figure 4.

these are trees having a root of type i, this root has four children of types j_1, j_2, j_3 and j_4, each one of these four vertices has a child of type i and is also the planted root of a tree of length $t-1$. Let us denote by $\mathcal{S}_i$ the set of all these trees. Let $T\in\mathcal{S}_i$ a tree parameterized by $(T_1,T_2,T_3,T_4)\in\bar{\mathcal{U}}_{j_1}^t\times\bar{\mathcal{U}}_{j_2}^t\times\bar{\mathcal{U}}_{j_3}^t\times\bar{\mathcal{U}}_{j_4}^t$ and let μ be the number of edges of T, i.e.

$$\mu=8+\sum_{\ell=1}^{4}|E(T_\ell)|.$$

Following the proof of Proposition 4.2, we know that

$$|\mathbb{E}\psi(T_1,T_2,T_3,T_4)|\le CK_n^{-\mu/2},\tag{68}$$

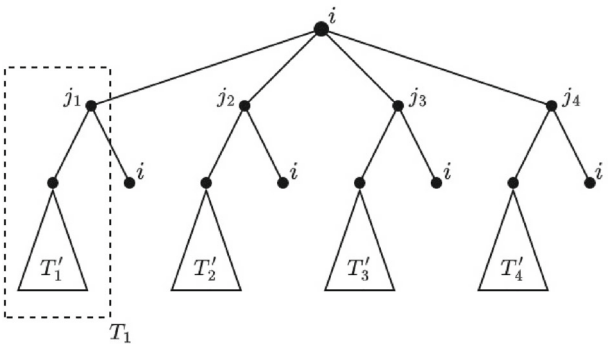


Fig. 4 Tree structure.

Let us now compute the number of non-vanishing contributions in $\varphi(j_1, j_2, j_3, j_4)$. A term $\mathbb{E}\psi(T_1, T_2, T_3, T_4)$ vanishes if there exists an $\ell = 1, 2, 3, 4$ such that neither the edge $(i \to j_\ell)$ nor $(j_\ell \to i)$ belongs to set of edges of the trees $T_1, \cdots, T_4$ or if there exists another edge in $T_1, \cdots, T_4$ which occurs once, in other words, if we consider the graph G obtained by identifying the vertices of the same type in T then T has a non-vanishing contribution if all the edges are covered in G at least twice and the edges $\{(i, j_\ell) \mid \ell = 1, \cdots, 4\}$ at least three times, then:

$$\mu \geq 2\,(|E(G)| - 4) + 3 \times 4 = 2|E(G)| + 4.$$

Notice that G is a connected graph (there exists a path from any vertex of G to i), then

$$|V(G)| \leq |E(G)| + 1 \leq \mu/2 - 1.$$

The vertices except $\{i, j_1, j_2, j_3, j_4\}$ can have arbitrary types from a set of at most CK_n types, so we get

$$|\mathbb{E}\varphi(j_1, j_2, j_3, j_4)| \leq CK_n^{-\mu/2} K_n^{\mu/2-1-5} = CK_n^{-6},$$

In addition, we have $\binom{K_n}{4} \leq CK_n^4$ choices for quadruples (j_1, j_2, j_3, j_4) with distinct elements, this means that

$$\sum_{\substack{j_1, j_2, j_3, j_4 \in [n] \\ \text{distinct}}} |\mathbb{E}\varphi(j_1, j_2, j_3, j_4)| \leq CK_n^{-2}.$$

A similar argument can be used to analyze the other cases where j_1, j_2, j_3, j_4 are not necessarily distinct.

Appendix E. Proof of Proposition 5.5

We begin by decoupling the entries of our random matrix W using triangular inequality twice

$$\left(\mathbb{E}\|W\|^p\right)^{1/p} \leq \left(\mathbb{E}(\|U\| + \|L\|)^p\right)^{1/p} \leq \left(\mathbb{E}\|U\|^p\right)^{1/p} + \left(\mathbb{E}\|L\|^p\right)^{1/p},$$

where U and L are $n \times n$ triangular matrices corresponding to the upper part (including diagonal) and lower part of W, respectively. Notice that U can be seen as an $n \times n$ random matrix with independent entries having the following variance profile

$$s^u_{ij} = \begin{cases} s_{ij} \text{ if } i \leq j \\ 0 \quad \text{otherwise.} \end{cases}$$

Following the notations of [4] we define

$$\sigma_1 = \max_i \left(\sum_{j \geq i} s_{ij}\right)^{1/2}, \quad \sigma_2 = \max_j \left(\sum_{i \leq j} s_{ij}\right)^{1/2}, \quad \sigma_* = \max_{i \leq j} \sqrt{s_{ij}}.$$

Now using the results of [4] we get

$$\begin{aligned} \left(\mathbb{E}\|U\|^{2\log(n)}\right)^{1/2\log(n)} &\lesssim \sigma_1 + \sigma_2 + \sigma_*(\log(n))^{(\rho \vee 1)/2} \\ &\lesssim 1 + \sqrt{\frac{(\log(n))^{\rho \vee 1}}{K_n}}. \end{aligned}$$

Using assumption A-2 we get $\left(\mathbb{E}\|U\|^{2\log(n)}\right)^{1/2\log(n)} \leq C$ and with a similar treatment to L we finally get $\left(\mathbb{E}\|W\|^{2\log(n)}\right)^{1/2\log(n)} \leq C$. Using Markov's inequality,

$$\mathbb{P}\left[\|W\| \geq Ce\right] \leq \frac{1}{n^2}.$$

Finally, using Borel-Cantelli's lemma we get

$$\mathbb{P}\left[\sup_n \|W\| < \infty\right] = 1.$$

Appendix F. Proof of Lemma 5.2

We prove both results by induction on t. The proof of the first item is very similar to [21, Lemma 1] and thus will be omitted. Let us now prove the second item. For $t = 1$ we have $R^1_i(1,1) = \sum_{\ell=1}^n s_{i\ell} \left(h(x^0_\ell, \eta_\ell, 0)\right)^2 \geq \inf_{n \in \mathbb{N}} \inf_{i \in [n]} \left(h(x^0_i, \eta_i, 0)\right)^2 \sum_{\ell=1}^n s_{i\ell}$,

using assumptions A-2, A-4 and A-6-(1) we get the result. Suppose now that that exists $c > 0$ such that

$$\forall n \in \mathbb{N}, \forall i \in [n],\ \sigma_i := \sqrt{R_i^t(t,t)} \geq c.$$

Let $\xi \sim \mathcal{N}(0,1)$, we can write

$$R_i^{t+1}(t+1,t+1) = \sum_{\ell=1}^{n} s_{i\ell}\mathbb{E}\left(h(Z_\ell^t, \eta_\ell, t)\right)^2 = \sum_{\ell=1}^{n} s_{i\ell}\mathbb{E}\left(h(\sigma_\ell\xi, \eta_\ell, t)\right)^2$$
$$\geq \mathbb{E}\left(h(\sigma_\star\xi, \eta_\star, t)\right)^2 \sum_{\ell=1}^{n} s_{i\ell},$$

where $(\sigma_\star, \eta_\star)$ is such that $\mathbb{E}\left(h(\sigma_\star\xi, \eta_\star, t)\right)^2 = \min_{\ell\in[n]} \mathbb{E}\left(h(\sigma_\ell\xi, \eta_\ell, t)\right)^2$. Let $D > 0$ be as in A-6-(2), using the induction hypothesis and the previous result we can see that $0 < c \leq \sigma_\star \leq C$, using this gives the following

$$\begin{aligned}
\mathbb{E}\left(h(\sigma_\star\xi, \eta_\star, t)\right)^2 &= \frac{1}{\sigma_\star\sqrt{2\pi}} \int_{\mathbb{R}} \left(h(x, \eta_\star, t)\right)^2 \exp(-x^2/2\sigma_\star^2)dx \\
&\geq \frac{1}{C\sqrt{2\pi}} \int_{[-D,D]} \left(h(x, \eta_\star, t)\right)^2 \exp(-x^2/2\sigma_\star^2)dx \\
&\geq \frac{\exp(-D^2/2\sigma_\star^2)}{C\sqrt{2\pi}} \int_{[-D,D]} \left(h(x, \eta_\star, t)\right)^2 dx \\
&\geq \frac{\exp(-D^2/2c^2)}{C\sqrt{2\pi}} \inf_{\eta\in\mathcal{Q}_\eta} \int_{[-D,D]} \left(h(x, \eta, t)\right)^2 dx\,.
\end{aligned}$$

Finally assumption A-6-(2) gives the result.

Appendix G. Hölder Continuity of the Squared Root

Lemma G.1 *The function $X \mapsto X^{1/2}$ is $\frac{1}{2}$-Hölder continuous on $\mathcal{S}_+^n$ (the set of symmetric positive matrices).*

Proof Let $A, B \in \mathcal{S}_+^n$, it suffices to show the following inequality,

$$\|A - B\|^2 \leq \|A^2 - B^2\|.$$

Let λ be an eigenvalue of $A - B$ such that $|\lambda| = \|A - B\|$, then there exists $u \in \mathbb{R}^n$ of norm 1 such that

$$(A - B)u = \lambda u.$$

We can write the following

$$A^2 - B^2 = (A - B)^2 + B(A - B) + (A - B)B,$$

taking the quadratic form of this matrix at u gives

$$\|A^2 - B^2\| \geq u^\top (A^2 - B^2)u = \lambda^2 + 2\lambda u^\top Bu\,.$$

We can assume without loss of generality that $\lambda \geq 0$, having that $u^\top Bu \geq 0$ gives

$$\|A^2 - B^2\| \geq \lambda^2 + 2\lambda u^\top Bu \geq \lambda^2 = \|A - B\|^2.$$

□

This result is used in the proof of Lemma 5.6.

Appendix H. Proof of Corollary 2.2

Consider the two AMP sequences $(\boldsymbol{x}^t)$ and $(\tilde{\boldsymbol{x}}^t)$ defined recursively by

$$\begin{aligned}\boldsymbol{x}^{t+1} &= Wh(\boldsymbol{x}^t) - \rho\,\langle \partial h(\boldsymbol{x}^t)\rangle_n\, h(\boldsymbol{x}^{t-1}),\\ \tilde{\boldsymbol{x}}^{t+1} &= Wh(\tilde{\boldsymbol{x}}^t) - \rho\,\mathbb{E}\,\partial h(Z^t)\, h(\tilde{\boldsymbol{x}}^{t-1}),\end{aligned}$$

with $\boldsymbol{x}^0 = \tilde{\boldsymbol{x}}^0$, and where $(Z_1, \dots, Z_t)$ satisfies the "asymptotic" Density Evolution equations (13).

We present here only the proof of the convergence of the empirical measure of the last iterate $\mu^{\boldsymbol{x}^t}$. The proof of the convergence of the empirical measure of the first t iterates $(\boldsymbol{x}^1, \dots, \boldsymbol{x}^t)$ follows by the same argument. To prove Corollary 2.2, it therefore suffices to show, by induction, that for all $t \geq 0$,

$$\|\boldsymbol{x}^t - \tilde{\boldsymbol{x}}^t\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0, \tag{69}$$

$$\mu^{\boldsymbol{x}^t} \xrightarrow[n\to\infty]{\mathbb{P},\,\mathcal{W}_2} \mu_t, \tag{70}$$

where we denote by μ_t the law of Z_t and we use the notation of Wasserstein convergence $\mathcal{W}_2$. The statements clearly hold for $t = 0, 1$. Assume now that (69) and (70) hold for t and $t-1$, and let us prove that they also hold for $t+1$.

Step 1. Proof of (69) for $t+1$. We decompose

$$\begin{aligned}\boldsymbol{x}^{t+1} - \tilde{\boldsymbol{x}}^{t+1} &= W\big(h(\boldsymbol{x}^t) - h(\tilde{\boldsymbol{x}}^t)\big)\\ &\quad - \rho\, h(\boldsymbol{x}^{t-1})\big[\langle \partial h(\boldsymbol{x}^t)\rangle_n - \mathbb{E}\,\partial h(Z^t)\big]\\ &\quad - \rho\,\mathbb{E}\,\partial h(Z^t)\,\big[h(\boldsymbol{x}^{t-1}) - h(\tilde{\boldsymbol{x}}^{t-1})\big]\\ &=: \Delta_1 + \Delta_2 + \Delta_3.\end{aligned}$$

Term Δ_1. We have $\|\Delta_1\|_n \leq L\|W\|\,\|\boldsymbol{x}^t - \tilde{\boldsymbol{x}}^t\|_n$. Using the spectral norm bound on W (see Proposition 5.5) and the induction hypothesis (69), we conclude that $\|\Delta_1\|_n \to 0$ in probability.

Term Δ_2. By the induction hypothesis (70), $\mu^{\boldsymbol{x}^t} \Rightarrow \mu_t$ in probability. Let (n) be a subsequence along which this convergence holds almost surely. Since ∂h is bounded (as h is Lipschitz) and continuous μ_t-a.e., we have by [18, Lemma 7.14] that $\langle \partial h(\boldsymbol{x}^t)\rangle_n - \mathbb{E}\,\partial h(Z^t) \to 0$ a.s. along (n). It remains to show that $\|h(\boldsymbol{x}^{t-1})\|_n$ is bounded. We can write

$$\|h(\boldsymbol{x}^{t-1})\|_n \le \|h(\boldsymbol{x}^{t-1}) - h(\tilde{\boldsymbol{x}}^{t-1})\|_n + \|h(\tilde{\boldsymbol{x}}^{t-1})\|_n.$$

The first term converges to zero in probability by (69), and the second is bounded almost surely (see Theorem 2.1). By the subsequence criterion [24, Lemma 3.2], we conclude that $\|\Delta_2\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Term Δ_3. This term vanishes in probability directly from the induction hypothesis (69) for $t-1$.

Step 2. Proof of (70) for $t+1$. Let φ be a bounded Lipschitz test function. Then

$$\left|\int \varphi\, d\mu^{\boldsymbol{x}^{t+1}} - \int \varphi\, d\mu_{t+1}\right| \le \left|\int \varphi\, d\mu^{\boldsymbol{x}^{t+1}} - \int \varphi\, d\mu^{\tilde{\boldsymbol{x}}^{t+1}}\right| + \left|\int \varphi\, d\mu^{\tilde{\boldsymbol{x}}^{t+1}} - \int \varphi\, d\mu_{t+1}\right|.$$

The second term vanishes in probability by Theorem 2.1, and the first is bounded by $\|\boldsymbol{x}^{t+1} - \tilde{\boldsymbol{x}}^{t+1}\|_n$ (up multiplication by the Lipschitz constant), which vanishes in probability as established above.

To obtain convergence in $\mathcal{W}_2$, it remains to show that the difference between the second moments of $\mu^{\boldsymbol{x}^{t+1}}$ and μ_{t+1} vanishes in probability:

$$\begin{aligned}&\left|\int x^2\, d\mu^{\boldsymbol{x}^{t+1}} - \int x^2\, d\mu_{t+1}\right| \\ &\quad\le \left|\int x^2\, d\mu^{\boldsymbol{x}^{t+1}} - \int x^2\, d\mu^{\tilde{\boldsymbol{x}}^{t+1}}\right| + \left|\int x^2\, d\mu^{\tilde{\boldsymbol{x}}^{t+1}} - \int x^2\, d\mu_{t+1}\right|.\end{aligned}$$

The second term again vanishes by Theorem 2.1. For the first term,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n (x_i^{t+1})^2 - (\tilde{x}_i^{t+1})^2 &= \frac{1}{n}\sum_{i=1}^n (x_i^{t+1} - \tilde{x}_i^{t+1})(x_i^{t+1} + \tilde{x}_i^{t+1}) \\ &\le \|\boldsymbol{x}^{t+1} - \tilde{\boldsymbol{x}}^{t+1}\|_n \|\boldsymbol{x}^{t+1} + \tilde{\boldsymbol{x}}^{t+1}\|_n \\ &\le \|\boldsymbol{x}^{t+1} - \tilde{\boldsymbol{x}}^{t+1}\|_n \big(\|\boldsymbol{x}^{t+1} - \tilde{\boldsymbol{x}}^{t+1}\|_n + 2\|\tilde{\boldsymbol{x}}^{t+1}\|_n\big).\end{aligned}$$

The quantity $\|\tilde{\boldsymbol{x}}^{t+1}\|_n$ is bounded (Theorem 2.1), while $\|\boldsymbol{x}^{t+1} - \tilde{\boldsymbol{x}}^{t+1}\|_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$, completing the induction.

Acknowledgements The authors would like to thank Mylène Maïda, François Massol and Viet-Chi Tran for useful discussions.

Author Contributions All authors have equally contributed to the present article.

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Akjouj, I., Barbier, M., Clenet, M., Hachem, W., Maïda, M., Massol, F., Najim, J., Tran, V.-C.: Complex systems in ecology: a guided tour with large lotka-volterra models and random matrices. Proc. Royal Soc. A **480**(2285), 20230284 (2024)
2. Akjouj, I., Hachem, W., Maïda, M., Najim, J.: Equilibria of large random lotka-volterra systems with vanishing species: a mathematical approach. J. Math. Biol. **89**(6), 61 (2024)
3. Allesina, S., Tang, S.: The stability-complexity relationship at age 40: a random matrix perspective. Popul. Ecol. **57**(1), 63–75 (2015)
4. Bandeira, A.S., Van Handel, R.: Sharp nonasymptotic bounds on the norm of random matrices with independent entries. Ann. Prob. **44**(4), 2479–2506 (2016)
5. Bao, Z., Han, Q., Xu, X.: A leave-one-out approach to approximate message passing. Ann. Appl. Probab. **35**(4), 2716–2766 (2025)
6. Barbier, J., Krzakala, F.: Approximate Message-Passing decoder and capacity achieving sparse superposition codes. IEEE Trans. Inf. Theory **63**(8), 4894–4927 (2017)
7. Bayati, M., Lelarge, M., Montanari, A.: Universality in polytope phase transitions and message passing algorithms. Ann. Appl. Prob. **25**(2), 753–822 (2015)
8. Bayati, M., Montanari, A.: The dynamics of message passing on dense graphs, with applications to compressed sensing. IEEE Trans. Inf. Theory **57**(2), 764–785 (2011)
9. Behne, J. K., Reeves, G.: Fundamental limits for rank-one matrix estimation with groupwise heteroskedasticity. In: Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, vol. 151 of Proceedings of Machine Learning Research, pp. 8650–8672, PMLR, 28–30 Mar 2022
10. Busiello, D.M., Suweis, S., Hidalgo, J., Maritan, A.: Explorability and the origin of network sparsity in living systems. Sci. Rep. **7**(1), 12323 (2017)
11. Bunin, G.: Ecological communities with lotka-volterra dynamics. Phys. Rev. E **95**, 042414 (2017)
12. Clenet, M., Massol, F., Najim, J.: Impact of a block structure on the lotka-volterra model. Peer Commun. J. **4**, e86 (2024)
13. Cure, S., Neri, I.: Antagonistic interactions can stabilise fixed points in heterogeneous linear dynamical systems. SciPost Phys. **14**(5), 093 (2023)
14. Curtiss, J.H.: A note on the theory of moment generating functions. Ann. Math. Stat. **13**(4), 430–433 (1942)
15. Deshpande, Y., Abbe, E., Montanari, A.: Asymptotic mutual information for the balanced binary stochastic block model. Inf. Inference: J. IMA **6**(2), 125–170 (2017)
16. Deshpande, Y., Montanari, A.: Information-theoretically optimal sparse pca. In: 2014 IEEE International Symposium on Information Theory, pp. 2197–2201. IEEE, 2014
17. Donoho, D.L., Maleki, A., Montanari, A.: Message-passing algorithms for compressed sensing. Proc. Natl. Acad. Sci. **106**(45), 18914–18919 (2009)
18. Feng, O.Y., Venkataramanan, R., Rush, C., Samworth, R.J.: A unifying tutorial on approximate message passing. Found. Trends Mach. Learn. **15**(4), 335–536 (2022)
19. Gueddari, M-Y., Hachem, W., Najim, J.: Elliptic approximate message passing and an application to theoretical ecology. Random Matrices: Theory Appl. 14(4), (2025)
20. Guionnet, A., Ko, J., Krzakala, F., Zdeborová, L.: Low-rank matrix estimation with inhomogeneous noise. Inf. Inference: J IMA **14**(2), iaaf010 (2025)
21. Hachem, W.: Approximate message passing for sparse matrices with application to the equilibria of large ecological lotka–volterra systems. Stoch. Process. Appl. **170**, 104276 (2024)
22. Han, Q.: Entrywise dynamics and universality of general first order methods. Ann. Statist. **53**(4), 1783–1807 (2025)

23. Javanmard, A., Montanari, A.: State evolution for general approximate message passing algorithms, with applications to spatial coupling. Inf. Inference: J. IMA **2**(2), 115–144 (2013)
24. Kallenberg, O.: Foundations of modern probability. Probability and its applications, 2nd edn. Springer-Verlag, New York (2002)
25. Lelarge, M., Miolane, L.: Fundamental limits of symmetric low-rank matrix estimation. Probab. Theory Relat. Fields **173**, 859–929 (2019)
26. Montanari, A.: Optimization of the Sherrington-Kirkpatrick hamiltonian. SIAM J. Comput. (2021). https://doi.org/10.1137/20M132016X
27. Montanari, A., Venkataramanan, R.: Estimation of low-rank matrices via approximate message passing. Ann. Stat. **49**(1), 321–345 (2021)
28. Pak, A., Ko, J., Krzakala, F.: Optimal Algorithms for the Inhomogeneous Spiked Wigner Model. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in neural information processing systems, pp. 76409–76424. Curran Associates Inc, USA (2023)
29. Rush, C., Greig, A., Venkataramanan, R.: Capacity-achieving sparse superposition codes via approximate message passing decoding. IEEE Trans. Inf. Theory **63**(3), 1476–1500 (2017)