Approximate Message Passing for sparse matrices with application to the equilibria of large ecological Lotka-Volterra systems

Walid Hachem

CNRS, Laboratoire d'informatique Gaspard Monge (LIGM / UMR 8049),
Université Gustave Eiffel, ESIEE Paris, France

June 11, 2024 (revised version)

Abstract

This paper is divided into two parts. The first part is devoted to the study of a class of Approximate Message Passing (AMP) algorithms which are widely used in the fields of statistical physics, machine learning, or communication theory. The AMP algorithms studied in this part are those where the measurement matrix has independent elements, up to the symmetry constraint when this matrix is symmetric, with a variance profile that can be sparse. The AMP problem is solved by adapting the approach of Bayati, Lelarge, and Montanari (2015) to this matrix model.

The Lotka-Volterra (LV) model is the standard model for studying the dynamical behavior of large dimensional ecological food chains. The second part of this paper is focused on the study of the statistical distribution of the globally stable equilibrium vector of a LV system in the situation where the random symmetric interaction matrix among the living species is sparse, and in the regime of large dimensions. This equilibirium vector is the solution of a Linear Complementarity Problem, which distribution is shown to be characterized through the AMP approach developed in the first part. In the large dimensional regime, this distribution is close to a mixture of a large number of truncated Gaussians.

Keywords: Approximate Message Passing, Equilibria of ecological systems, Lotka-Volterra Ordinary Differential Equations, Sparse random matrices.

1 Introduction

An ecosystem can be seen as a multi-dimensional dynamical system that represents the time evolution of the abundances of the interacting species. The behavior of such systems is governed by the intrinsic population dynamics and by the strengths of the interactions among these species. Given a system model, it is of interest to evaluate the distribution of the species at the equilibrium when this equilibrium exists and is unique. The present paper is motivated by this general problem.

An archetypal model for an ecological dynamical system is provided by the so called Lotka-Volterra (LV) multi-dimensional Ordinary Differential Equation (ODE). The dynamics of a LV ODE with n species take the form

$$\dot{u}(t) = u(t) \odot (r + (\Sigma - I_n)\, u(t)) \,, \quad t \geqslant 0,$$

where the vector function $u : \mathbb{R}_+ \to \mathbb{R}_+^n$ represents the abundances of the n coexisting species after a proper normalization, $\odot$ is the element-wise product, r is the so-called vector of intrinsic growth rates of the species, and Σ is a $n \times n$ matrix which $(i, j)^{\text{th}}$ element reflects the interaction effect of Species j on the growth of Species i [Tak96].

Denoting as $\|\cdot\|$ the spectral norm, it known that under a condition such as $\|(\Sigma+\Sigma^\top)/2\| < 1$, the ODE solution is well-defined, and this ODE has an unique globally stable equilibrium $u_\star = \left[u_{\star,i}\right]_{i=1}^n$ in the classical sense of the Lyapounov theory [Tak96, LJM09]. It is of interest to study the distribution of the elements of this vector, which quantifies the relative abundances of the species at the equilibrium. It is useful to note here that $u_\star$ frequently lies at the boundary of the first quadrant of $\mathbb{R}^n$, and therefore, it is of particular interest to evaluate the proportion of surviving species at these equilibria.

Usually, the interaction matrix Σ is difficult to measure or to evaluate, and all the more so as the ecosystem's dimension n gets large. To circumvent this difficulty, a whole line of research in theoretical ecology considers that the matrix Σ is a random matrix, and focuses on the dynamics of the LV system in the regime where $n \to \infty$ [AT15, ABC$^+$22]. The idea is to predict some essential aspects of the dynamical behavior of the ecosystem on the basis of a few "phenomenological" statistical features of the interaction matrix, rather than on its fine structure. The application of large random matrix tools to the dynamical behavior study of ecological systems dates back to the work of May [May72]. Among the most widely studied statistical models for the interaction matrix from the standpoint of the large random matrix theory are the Gaussian Orthogonal Ensemble (GOE) model, the Gaussian model with i.i.d. elements (sometimes called the Ginibre model), or the so-called elliptical model, which can be seen as an "interpolation" these two with possibly a non-zero mean [AT12].

Given a statistical model for Σ, and assuming the existence of $u_\star$ which is now a random vector, the problem amounts in our context to evaluating the asymptotic behavior of the random probability measure

$$\mu^{u_\star} = \frac{1}{n}\sum_{i\in[n]} \delta_{u_{\star,i}}$$

as $n \to \infty$. In the literature, this has been mostly done with tools issued from the physics. In [Bun17], Bunin obtained the asymptotic distribution of the equilibrium for the elliptical model by using the so-called dynamical cavity method. A similar result was obtained by Galla in [Gal18] with the help of generating functionals techniques. Older results in the same vein can be found in, *e.g.*, [OD92, Tok04]. Heuristic evaluations of the asymptotic behavior of $\mu^{u_\star}$ were proposed in [CEN22, CNM22], most generally in the elliptical non-centered case.

In this paper, we consider a symmetric model for Σ, which can be used to represent the competitive and the mutualistic interactions [BBC18, ABC$^+$22]. In this framework, we assume that the coefficients of Σ are not necessarily Gaussian, are independent up to the symmetry constraint, and are subjected to a variance profile that can be sparse. The two main features of our model are thus the inhomogeneity of the interaction strengths between the species, and their sparsity. Regarding this last assumption, it is indeed commonly observed that a species interacts with a very small proportion of the other species coexisting within the ecosystem [BSHM17].

Our approach is mathematically rigorous, contrary to the references we just mentioned. To obtain our results, we generalize the technique of our recent preprint [AHMN23] devoted to the GOE case. The idea goes as follows: The vector $u_\star$ can be identified as the solution of a Linear Complementarity Problem (LCP), a class of problems studied in the field of linear programming (see [Tak96, CPS09]). In [AHMN23], it is shown that in the GOE case, the asymptotic behavior of $\mu^{x_\star}$, where $u_\star$ is now the LCP solution, can be evaluated with the help of an Approximate Message Passing (AMP) technique. Such techniques have recently aroused an intense and growing research effort in the fields of statistical physics, communication theory, or statistical Machine Learning [FVRS22]. In a word, given a function $h : \mathbb{R}\times\mathbb{R}\times\mathbb{N} \to \mathbb{R}$ and a random symmetric $n \times n$ so-called measurement matrix W, a standard AMP algorithm is an iterative algorithm of the form

$$\hat{x}^{t+1} = W h(\hat{x}^t, \eta, t) \ + \ \text{a "correction" term},$$

where $\eta = [\eta_i]_{i=1}^n \in \mathbb{R}^n$ is a parameter vector, and where $h(\hat{x}^t, \eta, t) = \left[h(\hat{x}_i^t, \eta_i, t)\right]_{i=1}^n$. By properly designing the correction term, one is able to control the joint distribution of the $(t+1)$–uple $(\eta, \hat{x}^1, \dots, \hat{x}^t)$ for each $t \geqslant 0$ and for $n \to \infty$, as will developed more precisely below. We shall be able to make use of such a result to evaluate the large–n distribution of our equilibrium vector, and show that this distribution is close to a mixture of Gaussians.

Thus, from the standpoint of theoretical ecology, the present paper is a generalization of [AHMN23] to the case where Σ is non-necessarily Gaussian and is subjected to a variance profile that can be sparse. To that end, a version of the AMP algorithm well-suited to these kind of matrices is developed below, and might have its own interest due to its potential applications in other fields than in ecology.

Let us provide a quick review of the AMP literature in order to better position our contribution in this respect. Many of the original ideas lying behing the AMP algorithms come from the fields of statistical physics and communication theory. The first rigorous AMP results in a framework close to this paper were developed by Bolthausen [Bol14] and Bayati and Montanari [BM11] both for GOE matrices and for rectangular matrices with i.i.d. Gaussian elements. Since then, the AMP approach has been generalized in many directions. Let us cite the Generalized AMP of [Ran11], the contributions [JM13], [BR22] and [GKKZ22, PKK23] where block variance profiles are considered, [Fan22] devoted to rotationally invariant matrices, or the graph-based approach of [GB21]. Universality results in terms of the distribution of the elements of W were proposed in the recent papers [CL21], [DLS22], and [WZF22]. The closest to our paper among these is [WZF22], which considers among others the case where W is a centered symmetric matrix with independent elements satisfying $\mathbb{E}W_{ij}^2 \lesssim 1/n$ and $\lim_n \max_i |(\sum_j \mathbb{E}W_{ij}^2) - 1| = 0$, and shows that W can be replaced with a GOE matrix as regards the AMP problem. This constraint on the variance profile is alleviated in our context, leading to a more involved expression of the asymptotic joint distribution of the algorithm iterates.

In [BLM15], Bayati, Lelarge, and Montanari were among the first to establish a universal AMP result. Starting with polynomial activation functions and using a combinatorial approach, these authors propose a moment computation of the elements of the vector iterates, where these elements are expressed as sums of monomials in the matrix entries which are indexed by labelled trees. As regards the AMP part, the present paper is essentially an adaptation of the approach of [BLM15] to the case of a sparse variance profile.

Section 2 is devoted to our general AMP results for symmetric and non-symmetric measurement matrices, independently of the ecological application. Our LV problem is then stated in Section 3, along with the result on the large–n distribution of the equilibrium. The proofs for Sections 2 and 3 are provided in Sections 4 and 5 respectively.

2 Sparse AMP with a variance profile: problem statement and results

We start with our assumptions. Let (n) be a sequence of integers in the set $\{2, 3, \ldots\}$ that converges to infinity. For each n, let $\{X_{ij}^{(n)}\}_{1 \leqslant i < j \leqslant n}$ be a set of real random variables such that:

Assumption 1. The following facts hold true.

- The $n(n-1)/2$ random variables $X_{ij}^{(n)}$ for $1 \leqslant i < j \leqslant n$ are independent.
- $\mathbb{E}X_{i,j}^{(n)} = 0$ and $\mathbb{E}(X_{ij}^{(n)})^2 = 1$.
- For each integer $k > 2$, there exists a constant $C_{\mathrm{mom}}(k) > 0$ such that

$$\sup_n \max_{1 \leqslant i < j \leqslant n} \left(\mathbb{E} \left| X_{ij}^{(n)} \right|^k \right)^{1/k} \leqslant C_{\mathrm{mom}}(k).$$

Let us write $X_{ji}^{(n)} = X_{ij}^{(n)}$ for $1 \leqslant i < j \leqslant n$, and $X_{ii}^{(n)} = 0$ for $i \in [n]$, and let us consider the $n \times n$ random symmetric matrix $X^{(n)} = \left[X_{ij}^{(n)} \right]_{i,j=1}^n$.

For each n, let $\{s_{ij}^{(n)}\}_{1\leqslant i<j\leqslant n}$ be a set of deterministic non-negative numbers. Write $s_{ji}^{(n)} = s_{ij}^{(n)}$ for $1 \leqslant i < j \leqslant n$, and $s_{ii}^{(n)} = 0$ for $i \in [n]$, and consider the $n \times n$ symmetric matrix with non-negative elements $S^{(n)} = \left[s_{ij}^{(n)}\right]_{i,j=1}^{n}$. Define the random symmetric matrix $W^{(n)}$ as

$$W^{(n)} = \left[W_{ij}^{(n)}\right]_{i,j=1}^{n} = \left(S^{(n)}\right)^{\odot 1/2} \odot X^{(n)}, \tag{1}$$

where $\odot$ is the Hadamard product, and where $A^{\odot 1/2}$ is the element-wise square root of the matrix A.

Letting (K_n) be a sequence of positive integers indexed by n such that $K_n \leqslant n$, the variance profile matrix $S^{(n)}$ of $W^{(n)}$ complies with the following assumption:

Assumption 2. The following facts hold true.

- $K_n \to \infty$.
- There exists a constant $C_{\text{card}} > 0$ such that

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

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

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

Beyond these constraints, we do not put any structural assumption on the variance profile matrix $S^{(n)}$. In this paper, we shall be mostly interested in the situations where $K_n/n \to_n 0$, making the matrices $W^{(n)}$ sparse.

For each n, let $x^{(n),0} = \left[x_1^{(n),0}, \ldots, x_n^{(n),0}\right]^\top \in \mathbb{R}^n$ be a deterministic vector that will represent the initial value of our AMP sequence, and let $\eta^{(n)} = \left[\eta_1^{(n)}, \ldots, \eta_n^{(n)}\right]^\top \in \mathbb{R}^n$ be a deterministic parameter vector. These vectors obey the following assumptions:

Assumption 3. There exists a compact set $\mathcal{Q}_x \subset \mathbb{R}$ such that $x_i^{(n),0} \in \mathcal{Q}_x$ for all n and all $i \in [n]$.

Assumption 4. There exists a compact set $\mathcal{Q}_\eta \subset \mathbb{R}$ such that $\eta_i^{(n)} \in \mathcal{Q}_\eta$ for all n and all $i \in [n]$.

Consider a measurable function $h : \mathbb{R} \times \mathcal{Q}_\eta \times \mathbb{N} \to \mathbb{R}$, called activation function, that satisfies one of the two following assumptions:

Assumption 5 (Lipschitz activation function in the AMP parameter)**.** For each $t \in \mathbb{N}$, there exists a constant $C > 0$ and a continuous non-decreasing function $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ with $\kappa(0) = 0$, such that:

$$\forall x, x' \in \mathbb{R}, \quad |h(x, \eta, t) - h(x', \eta, t)| \leqslant C|x - x'|,$$

and

$$\forall \eta, \eta' \in \mathcal{Q}_\eta, \quad |h(x, \eta, t) - h(x, \eta', t)| \leqslant \kappa(|\eta - \eta'|)(1 + |x|).$$

Assumption 6 (AC activation function with polynomial growth in the AMP parameter)**.** The following properties hold true:

- For each $\eta \in \mathcal{Q}_\eta$ and each $t \in \mathbb{N}$, the function $h(\cdot, \eta, t)$ is absolutely continuous with a locally integrable derivative.
- For each $t \in \mathbb{N}$, there exists two constants $m, C > 0$ and a continuous non-decreasing function $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ with $\kappa(0) = 0$, such that:

$$|h(x, \eta, t)| \leqslant C(1 + |x|^m),$$

  and

$$\forall \eta, \eta' \in \mathcal{Q}_\eta, \quad |h(x, \eta, t) - h(x, \eta', t)| \leqslant \kappa(|\eta - \eta'|)(1 + |x|^m).$$

At the heart of the AMP algorithm that we shall study is a Gaussian $\mathbb{R}^n$–valued sequence of random vectors that we denote for each n as $(\boldsymbol{Z}^{(n),t})_{t\in\mathbb{N}_*}$, where $\mathbb{N}_* = \mathbb{N}\backslash\{0\}$. The probability distribution of this sequence is defined recursively in t as follows. Let $\boldsymbol{Z}^{(n),t} = \left[Z_1^{(n),t}, \ldots, Z_n^{(n),t}\right]^\top$. Writing $Z_i^{(n)} = (Z_i^{(n),1}, Z_i^{(n),2}, \ldots)$, the sequences $\{Z_i^{(n)}\}_{i=1}^n$ are centered, Gaussian, and independent. Denote as $R_i^{(n),t}$ the covariance matrix of the vector $\vec{Z}_i^{(n),t} = [Z_i^{(n),1}, \ldots, Z_i^{(n),t}]^\top$. These matrices are constructed recursively in the parameter t as follows. Defining the variances $\{\Xi_i^{(n),0}\}_{i=1}^n$ as $\Xi_i^{(n),0} = h(x_i^{(n),0}, \eta_i^{(n)}, 0)^2$, we set

$$R_i^{(n),1} = \sum_{l\in[n]} s_{il}^{(n)} \Xi_l^{(n),0}.$$

Given the matrices $\{R_i^{(n),t}\}_{i=1}^n$, the covariance matrices

$$\Xi_i^{(n),t} = \mathbb{E} \begin{bmatrix} h(x_i^{(n),0}, \eta_i^{(n)}, 0) \\ h(Z_i^{(n),1}, \eta_i^{(n)}, 1) \\ \vdots \\ h(Z_i^{(n),t}, \eta_i^{(n)}, t) \end{bmatrix} \begin{bmatrix} h(x_i^{(n),0}, \eta_i^{(n)}, 0) & h(Z_i^{(n),1}, \eta_i^{(n)}, 1) & \cdots & h(Z_i^{(n),t}, \eta_i^{(n)}, t) \end{bmatrix}, \quad (2)$$

are well-defined for all $i \in [n]$, whether Assumption 5 or Assumption 6 is used. With this at hand, we set

$$R_i^{(n),t+1} = \sum_{l\in[n]} s_{il}^{(n)} \Xi_l^{(n),t}. \quad (3)$$

It is clear that $R_i^{(n),t}$ is the principal matrix of $R_i^{(n),t+1}$ consisting in its first t rows and columns.

The distribution of the sequence $(\boldsymbol{Z}^{(n),t})_t$ is determined by the initial value $x^{(n),0}$, the parameter vector $\eta^{(n)}$, the variance profile matrix $S^{(n)}$, and the activation function h. We shall say from now on that this distribution is determined by the $(S^{(n)}, h, \eta^{(n)}, x^{(n),0})$-state evolution equations, which provide the covariance matrices $R_i^{(n),t}$. The spectral norms $\|R_i^{(n),t}\|$ are bounded as specified by the following lemma:

Lemma 1. Let Assumptions 2–4, and either Assumption 5 or Assumption 6 hold true. Then for each integer $t > 0$, it holds that

$$\sup_n \max_{i\in[n]} \|R_i^{(n),t}\| < \infty. \quad (4)$$

Sketch of proof. By recurrence on t. For $t = 1$, the result is a consequence of Assumption 2 and the easily shown continuity of $h(\cdot, \cdot, 0)$ on the compact $\mathcal{Q}_x \times \mathcal{Q}_\eta$. Assume the result is true for t. Then, using Assumption 4 and either Assumption 5 or Assumption 6 again, standard Gaussian derivations show that $\sup_n \max_i \mathbb{E} h(Z_i^{(n),t}, \eta_i, t)^2 < \infty$. From this, we obtain by using the Cauchy-Schwarz inequality that $\sup_n \max_i \|\Xi_i^{(n),t}\| < \infty$. By Assumption 2, we then obtain that $\sup_n \max_i \|R_i^{(n),t+1}\| < \infty$. □

Our next assumption will be used to ensure that the variances of the random variables $Z_i^{(n),t}$ are bounded away from zero:

Assumption 7. There exists a constant $c_S > 0$ such that

$$\inf_n \min_{i\in[n]} \sum_{j\in[n]} s_{ij}^{(n)} \geqslant c_S.$$

Recalling Assumption 2, let

$$\alpha_S = \frac{c_S}{2C_S - c_S/C_{\text{card}}}.$$

There is a sequence of positive constants $(c_{\rm h}(0), c_{\rm h}(1), \ldots)$ such that for each set $\mathcal{S}^{(n)} \subset [n]$ with $|\mathcal{S}^{(n)}| = \lfloor \alpha_S K_n \rfloor$, it holds that

$$\frac{1}{K_n} \sum_{l \in \mathcal{S}^{(n)}} h(x_l^{(n),0}, \eta_l^{(n)}, 0)^2 \geqslant c_{\rm h}(0),$$

and for each $t \geqslant 1$,

$$\frac{1}{K_n} \sum_{l \in \mathcal{S}^{(n)}} \mathbb{E} h(\underline{\xi}, \eta_l^{(n)}, t)^2 \geqslant c_{\rm h}(t),$$

where $\underline{\xi} \sim \mathcal{N}(0,1)$.

Non-degeneracy conditions similar to this assumption are usually not constraining and are invoked in the vast majority of the contributions dealing with AMP algorithms, see, *e.g.*, [JM13, FVRS22]. If needed, it is certainly possible to lighten this assumption by properly perturbing the matrix $S^{(n)}$ or the function h in the course of the proof of Theorems 2 and 4 below. A perturbation argument of this sort was considered in, *e.g.*, [BMN19].

In all this paper, we follow the following notational convention. Given a function $f : \mathbb{R}^k \to \mathbb{R}$ and a k–uple of $\mathbb{R}^n$–valued vectors $(a^\ell = [a_i^\ell]_{i=1}^n)_{\ell \in [k]}$, we denote as $f(a^1, \ldots, a^k)$ the $\mathbb{R}^n$–valued vector $[f(a_i^1, \ldots, a_i^k)]_{i=1}^n$. Similarly, if f is a $\mathbb{R}^k \times \mathbb{N} \to \mathbb{R}$ function, $f(a^1, \ldots, a^k, t)$ is the $\mathbb{R}^n$–valued vector $[f(a_i^1, \ldots, a_i^k, t)]_{i=1}^n$.

With either Assumption 5 or Assumption 6 at hand, we denote as $\partial h(x, \eta, t)$ a measurable function that coincides almost everywhere on $\mathbb{R}$ with the partial derivative $\partial h(\cdot, \eta, t)$ at x. Note that under Assumption 5, the existence of this function follows from Rademacher's theorem, and it is furthermore obvious that $\mathbb{E} \partial h(\boldsymbol{Z}^{(n),t}, \eta^{(n)}, t)$ exists for each $t \in \mathbb{N}_*$. Under Assumption 6, Stein's lemma says that $(\mathbb{E}\underline{\xi}^2) \mathbb{E} \partial h(\underline{\xi}, \eta, t) = \mathbb{E} \underline{\xi} h(\underline{\xi}, \eta, t)$ for each Gaussian centered random variable $\underline{\xi}$, see, *e.g.*, [FSW18, Th. 2.1]. From this identity, we can deduce that under Assumption 6, $\mathbb{E} \partial h(\boldsymbol{Z}^{(n),t}, \eta^{(n)}, t)$ also exists for each $t \in \mathbb{N}_*$.

We are now ready to state the results of this section. The AMP algorithm built around the matrix $W^{(n)}$ and the activation function h that we shall study here takes the following form: Starting with the vector $x^{(n),0}$, the vector $x^{(n),t+1} = [x_i^{(n),t+1}]_{i=1}^n \in \mathbb{R}^n$ delivered by this algorithm at Iteration $t+1$ is given as

$$x^{(n),t+1} = W^{(n)} h(x^{(n),t}, \eta^{(n)}, t) - \operatorname{diag}\Big(S^{(n)} \mathbb{E} \partial h(\boldsymbol{Z}^{(n),t}, \eta^{(n)}, t) \Big) h(x^{(n),t-1}, \eta^{(n)}, t-1), \quad (5)$$

with the term $\operatorname{diag}(\cdots) h(\cdots)$ being equal to zero for $t = 0$. Variants for this algorithm could be considered by replacing the diagonal matrix

$$\operatorname{diag}(S^{(n)} \mathbb{E} \partial h(\boldsymbol{Z}^{(n),t}, \eta^{(n)}, t))$$

above with $\operatorname{diag}(S^{(n)} \partial h(x^{(n),t}, \eta^{(n)}, t))$, or with $\operatorname{diag}((W^{(n)})^{\odot 2} \partial h(x^{(n),t}, \eta^{(n)}, t))$ where $A^{\odot 2}$ is the square of the matrix A in the Hadamard product, as is done in the literature in similar contexts. Algorithm (5) (or its variants) generalizes the algorithms studied in, *e.g.*, [JM13] or [BLM15].

Given an integer $k > 0$, a function $\varphi : \mathbb{R}^k \to \mathbb{R}$ is said pseudo-Lipschitz of degree 2 (notation: $\varphi \in \mathrm{PL}_2(\mathbb{R}^k)$) if there exists a constant $L > 0$ such that

$$\forall u, v \in \mathbb{R}^k, \quad |\varphi(u) - \varphi(v)| \leqslant L \|u - v\| (1 + \|u\| + \|v\|),$$

where $\|u\|$ is the Euclidean norm of the vector u. Our first result is devoted to the case where $h(\cdot, \eta, t)$ is Lipschitz:

Theorem 2. Let Assumptions 1–5 and 7 hold true. Let $C_W > 0$ be an arbitrary constant. Fix an arbitrary integer $t_{\max} > 0$. For each positive integer n, let $(\beta_1^{(n)}, \ldots, \beta_n^{(n)})$ a n–uple of real numbers such that $\sup_n \max_{i\in[n]} |\beta_i^{(n)}| < \infty$. Let $\varphi : \mathbb{R}^{t_{\max}+1} \to \mathbb{R}$ be a function in $\mathrm{PL}_2(\mathbb{R}^{t_{\max}+1})$. Then, for each $\varepsilon > 0$, it holds that

$$\mathbb{P}\left[\left[\left|\frac{1}{n}\sum_{i\in[n]} \beta_i^{(n)}\varphi(\eta_i^{(n)}, x_i^{(n),1}, \dots, x_i^{(n),t_{\max}}) - \beta_i^{(n)}\mathbb{E}\varphi(\eta_i^{(n)}, Z_i^{(n),1}, \dots, Z_i^{(n),t_{\max}})\right| \geqslant \varepsilon\right] \cap \left[\|W^{(n)}\| \leqslant C_W\right]\right] \xrightarrow[n\to\infty]{} 0. \quad (6)$$

Before interpreting this theorem, we need to bound the operator norm $\|W^{(n)}\|$. With the help of the results of Bandeira and van Handel in [BvH16], this can be done in the framework of the following assumption:

Assumption 8. There exists a constant $C > 0$ such that the moment bounds from Assumption 1 satisfy

$$C_{\mathrm{mom}}(k) \leqslant C k^{\rho/2},$$

for some $\rho \geqslant 0$.

As is well known, the cases $\rho = 1$ and $\rho = 2$ correspond to the sub-Gaussian and the sub-exponential distributions respectively.

Proposition 3. Let Assumptions 1, 2 and 8 hold true, and assume furthermore that $K_n \gtrsim (\log n)^{\rho \vee 1}$. Then, there exists a constant $C > 0$ such that $\limsup_n \|W^{(n)}\| \leqslant C$ with probability one.

Thus, the heavier the tail of the probability distributions of the $X_{ij}^{(n)}$, the faster we need to make K_n converge to infinity in order to ensure that $\|W^{(n)}\|$ is bounded. This proposition is proven in Appendix A.

We now comment Theorem 2.

It is useful to interpret the result of this theorem in terms of the empirical distribution of the particles $\{(\eta_i^{(n)}, x_i^{(n),1}, \dots, x_i^{(n),t_{\max}})\}_{i=1}^n$ as an element of a Wasserstein space of probability measures. For an integer $k > 0$, denote as $\mathcal{P}(\mathbb{R}^k)$ the space of probability measures on $\mathbb{R}^k$, and by $\mathcal{P}_2(\mathbb{R}^k)$ the Wasserstein space of probability measures on $\mathbb{R}^k$ with finite second moment, equipped with the Wasserstein distance $\boldsymbol{d}_2$ [Vil09]. Given a sequence of probability measures (ν_n) in $\mathcal{P}_2(\mathbb{R}^k)$ and a probability measure $\nu \in \mathcal{P}_2(\mathbb{R}^k)$, it is well-known that the three following assertions are equivalent [Vil09, FVRS22]:

i) $\boldsymbol{d}_2(\nu_n, \nu) \to 0$.

ii) For each continuous function $\varphi : \mathbb{R}^k \to \mathbb{R}$ such that $|\varphi(x_1, \dots, x_k)| \leqslant C(1 + x_1^2 + \cdots + x_k^2)$ for some $C > 0$, it holds that

$$\int \varphi d\nu_n \xrightarrow[n\to\infty]{} \int \varphi d\nu. \quad (7)$$

iii) The convergence (7) holds true for each function $\varphi \in \mathrm{PL}_2(\mathbb{R}^k)$.

Furthermore, if the sequence (ν_n) is random, then, the convergence $\boldsymbol{d}_2(\nu_n, \nu) \xrightarrow{\mathcal{P}} 0$ is equivalent to the convergence $\int \varphi d\nu_n \xrightarrow{\mathcal{P}} \int \varphi d\nu$ for each continuous function $\varphi : \mathbb{R}^k \to \mathbb{R}$ either with quadratic growth as in ii), or belonging to the class $\mathrm{PL}_2(\mathbb{R}^k)$.

We are interested here in the asymptotic behavior of the $\mathcal{P}_2(\mathbb{R}^{t_{\max}+1})$–valued random probability measure $\mu^{\eta^{(n)}, x^{(n),1}, \dots, x^{(n),t_{\max}}}$, defined as

$$\mu^{\eta^{(n)}, x^{(n),1}, \dots, x^{(n),t_{\max}}} = \frac{1}{n}\sum_{i=1}^n \delta_{(\eta_i^{(n)}, x_i^{(n),1}, \dots, x_i^{(n),t_{\max}})}.$$

In general, this random measure does not converge to any fixed probability measure, because we did not put any assumption on the variance profile of the elements of $W^{(n)}$ beyond Assumption 2. However, let us define the deterministic measure $\boldsymbol{\mu}^{(n),t_{\max}} \in \mathcal{P}(\mathbb{R}^{t_{\max}+1})$ as follows: Letting $\theta^{(n)}$ be a discrete random variable uniformly distributed on the set $[n]$ and independent of the sequence $(\boldsymbol{Z}^{(n),t})_t$, we put $\boldsymbol{\mu}^{(n),t_{\max}} = \mathscr{L}((\eta^{(n)}_{\theta^{(n)}}, \vec{Z}^{(n),t_{\max}}_{\theta^{(n)}}))$. In particular, $\boldsymbol{\mu}^{(n),t_{\max}}(\mathbb{R} \times \cdot)$ is a mixture of n multivariate Gaussians. By Assumption 4 and Lemma 1, one can readily show that the sequence $(\boldsymbol{\mu}^{(n),t_{\max}})_{n\geqslant 2}$ is pre-compact in the space $\mathcal{P}_2(\mathbb{R}^{t_{\max}+1})$. Moreover, taking $(\beta_1^{(n)}, \ldots, \beta_n^{(n)}) = (1, \ldots, 1)$ in the statement of Theorem 2 and assuming that $W^{(n)}$ is constructed in such a way that $\limsup_n \|W^{(n)}\| < \infty$ with probability one (see Proposition 3), Theorem 2 shows that

$$\forall \varphi \in \mathrm{PL}_2(\mathbb{R}^{t_{\max}+1}), \quad \int \varphi \, d\mu^{\eta^{(n)}, x^{(n),1}, \ldots, x^{(n),t_{\max}}} - \int \varphi \, d\boldsymbol{\mu}^{(n),t_{\max}} \xrightarrow[n\to\infty]{\mathcal{P}} 0, \tag{8}$$

which is equivalent to $\boldsymbol{d}_2(\mu^{\eta^{(n)}, x^{(n),1}, \ldots, x^{(n),t_{\max}}}, \boldsymbol{\mu}^{(n),t_{\max}}) \xrightarrow{\mathcal{P}} 0$, and also equivalent to replacing the statement "$\forall \varphi \in \mathrm{PL}_2(\mathbb{R}^{t_{\max}+1})$" above with "$\forall \varphi : \mathbb{R}^{t_{\max}+1} \to \mathbb{R}$ continuous with $|\varphi(u_1, \ldots, u_{t_{\max}+1})| \leqslant C(1 + u_1^2 + \cdots u_{t_{\max}+1}^2)$" [FVRS22].

Let us now describe a case where the convergence (8) boils down to the convergence of the empirical measure $\mu^{\eta^{(n)}, x^{(n),1}, \ldots, x^{(n),t_{\max}}}$ to a fixed probability measure. Consider the situation where $K_n = n$ and where the variances satisfy $s_{ij}^{(n)} = 1/n$ for all $i, j \in [n]$ (without more details, we neglect here the constraint $s_{ii}^{(n)} = 0$ that will be used in the proof of Theorem 2). In this case, it is easy to check that the Gaussian vectors $\{\vec{Z}_i^{(n),t}\}_{i\in[n]}$ are i.i.d., and furthermore, writing $\vec{Z}_i^{(n),t} \stackrel{\mathcal{L}}{=} \vec{Z}^{(n),t} = [Z^{(n),1}, \ldots, Z^{(n),t}]^\top$ with the covariance matrix $R^{(n),t}$, it holds that

$$R^{(n),t+1} = \\ \frac{1}{n}\sum_{i=1}^n \mathbb{E} \begin{bmatrix} h(x_i^{(n),0}, \eta_i^{(n)}, 0) \\ h(Z^{(n),1}, \eta_i^{(n)}, 1) \\ \vdots \\ h(Z^{(n),t}, \eta_i^{(n)}, t) \end{bmatrix} \begin{bmatrix} h(x^{(n),0}, \eta_i^{(n)}, 0) & h(Z^{(n),1}, \eta_i^{(n)}, 1) & \cdots & h(Z^{(n),t}, \eta_i^{(n)}, t) \end{bmatrix}.$$

Assume now that the probability measure $\nu^{(n)} = n^{-1} \sum_i \delta_{(x_i^{(n),0}, \eta_i^{(n)})}$ converges narrowly to a deterministic probability measure ν^∞ as $n \to \infty$, as it is frequently done in the literature, often under a randomness assumption on $(x^{(n),0}, \eta^{(n)})$ coupled with an independence assumption from $W^{(n)}$. Then, given a couple of real random variables $(\overline{x}, \overline{\eta})$ with the distribution ν^∞, we can readily replace $R^{(n),t+1}$ with the matrix $R^{\infty,t+1}$ that does not depend on n and that is written as

$$R^{\infty,t+1} = \mathbb{E}_{(\overline{x},\overline{\eta})} \mathbb{E}_{Z^{\infty,t}} \begin{bmatrix} h(\overline{x}, \overline{\eta}, 0) \\ h(Z^{\infty,1}, \overline{\eta}, 1) \\ \vdots \\ h(Z^{\infty,t}, \overline{\eta}, t) \end{bmatrix} \begin{bmatrix} h(\overline{x}, \overline{\eta}, 0) & h(Z^{\infty,1}, \overline{\eta}, 1) & \cdots & h(Z^{\infty,t}, \overline{\eta}, t) \end{bmatrix},$$

with $Z^{\infty,t} \sim \mathcal{N}(0, R^{\infty,t})$ being independent of $(\overline{x}, \overline{\eta})$. One consequence of this replacement is that the measure $\boldsymbol{\mu}^{(n),t_{\max}}$ in (8) can be replaced with the distribution $\boldsymbol{\mu}^{\infty,t_{\max}} = \mathscr{L}((\overline{\eta}, Z^{\infty,t_{\max}})) \in \mathcal{P}_2(\mathbb{R}^{t_{\max}+1})$, and the convergence (8) amounts to

$$\mu^{\eta^{(n)}, x^{(n),1}, \ldots, x^{(n),t_{\max}}} \xrightarrow[n\to\infty]{\mathcal{P}} \boldsymbol{\mu}^{\infty,t_{\max}} \quad \text{in } \mathcal{P}_2(\mathbb{R}^{t_{\max}+1}).$$

We recover here a well-known result given in, *e.g.*, [BM11, FVRS22].

In the general setting of this paper, only the convergence (8) is available. Note that in general, putting the assumptions on $x^{(n),0}, \eta^{(n)}$ mentioned above does not suffice to transform this convergence to a convergence to a fixed deterministic measure. Once again, this is due to the fact that we did not put any assumption on the asymptotic behavior of the variance profile matrix $S^{(n)}$ beyond Assumption 2.

Our aim now is to explore a generalization of Theorem 2 into two directions: We replace Assumption 5 on the activation function with the more general assumption 6, and we replace the test function $\varphi \in \mathrm{PL}_2(\mathbb{R}^{t_{\max}+1})$ (or, equivalently, a test function φ satisfying the statement ii) above) with a continuous function $\varphi : \mathcal{Q}_\eta \times \mathbb{R}^{t_{\max}} \to \mathbb{R}$ such that $|\varphi(\alpha, u_1, \ldots, u_{t_{\max}})| \leqslant C(1 + |u_1|^m + \cdots + |u_{t_{\max}}|^m)$ for a given arbitrarily integer $m > 0$. In this case, we obtain the following partial result that generalizes [BLM15]. For a vector u, we write hereinafter $\|u\|_n = \|u\|/\sqrt{n}$.

Theorem 4. Let Assumptions 1–4 and 6, 7 hold true. Let $C_W > 0$ be an arbitrary constant. Fix an arbitrary integer $t_{\max} > 0$. For each positive integer n, let $(\beta_1^{(n)}, \ldots, \beta_n^{(n)})$ be a n–uple of real numbers such that $\sup_n \max_{i\in[n]} |\beta_i^{(n)}| < \infty$. Let $\varphi : \mathcal{Q}_\eta \times \mathbb{R}^{t_{\max}} \to \mathbb{R}$ be a continuous function satisfying $|\varphi(\alpha, u_1, \ldots, u_{t_{\max}})| \leqslant C(1 + |u_1|^m + \cdots + |u_{t_{\max}}|^m)$ for some $C, m > 0$. Then, there exists a sequence of matrices $(\widetilde{\boldsymbol{X}}^{(n)})_{n\geqslant 2}$, with $\widetilde{\boldsymbol{X}}^{(n)} = \left[\tilde{x}^{(n),1}, \ldots, \tilde{x}^{(n),t_{\max}}\right] \in \mathbb{R}^{n\times t_{\max}}$, such that for every $\varepsilon > 0$,

$$\mathbb{P}\left[\left[\left\|\tilde{x}^{(n),t+1} - \bar{x}^{(n),t+1}\right\|_n > \varepsilon\right)\right] \cap \left[\|W^{(n)}\| \leqslant C_W\right]\right] \xrightarrow[n\to\infty]{} 0$$

for each $t = 0, 1, \ldots, t_{\max} - 1$, with

$$\bar{x}^{(n),t+1} = W^{(n)} h(\tilde{x}^{(n),t}, \eta^{(n)}, t) - \operatorname{diag}\left(S^{(n)} \mathbb{E}\partial h(\boldsymbol{Z}^{(n),t}, \eta^{(n)}, t)\right) h(\tilde{x}^{(n),t-1}, \eta^{(n)}, t-1)$$

and with $\tilde{x}^{(n),0} = x^{(n),0}$ and $\operatorname{diag}(\cdots)h(\cdots) = 0$ for $t = 0$, and furthermore,

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

Observe that the sequence $(\widetilde{\boldsymbol{X}}^{(n)})$ does not exactly describe an AMP sequence, furthermore, its construction depends on the function φ and the n–uples $(\beta_i^{(n)})_{i\in[n]}$. Despite these limitations, Theorem 4 is useful in most of the practical situations where the activation functions are not necessarily Lipschitz in the AMP parameter or where the test functions grow faster than quadratically at infinity.

The AMP algorithms studied in the literature for symmetric measurement matrices have non-symmetric analogues. For the sake of generality, we now succinctly deal with this case. The results of the next paragraph will not be used elsewhere in this paper.

The non-symmetric variant (sketch)

Let $(p(n))$ be a sequence of positive integers such that $0 < \inf p(n)/n \leqslant \sup p(n)/n < \infty$. For a given $p = p(n)$, let $G^{(n)} = [G_{ij}^{(n)}]_{i,j=1}^{p,n}$ be a random rectangular $p \times n$ matrix, where the random variables $G_{ij}^{(n)}$ for $(i,j) \in [p] \times [n]$ are independent, centered, and unit-variance random variables that share with the elements of $X^{(n)}$ the moment bounds given by Assumption 1. Let $B^{(n)} = [b_{ij}^{(n)}]_{i,j=1}^{p,n} \in \mathbb{R}^{p\times n}$ be a deterministic matrix with non-negative elements, and assume that the $(n+p) \times (n+p)$ symmetric matrix $\begin{bmatrix} 0 & B^{(n)} \\ (B^{(n)})^\top & 0 \end{bmatrix}$ satisfies Assumption 2 in the role of $S^{(n)}$ there. The rectangular $p \times n$ matrix of interest here is the random matrix $Y^{(n)}$ with independent elements and a variance profile defined as

$$Y^{(n)} = (B^{(n)})^{\odot 1/2} \odot G^{(n)}.$$

Let $\tilde{x}^{(n),0} = [\tilde{x}_j^{(n),0}]_{i\in[n]} \in \mathbb{R}^n$ be a deterministic vector which elements belong to a compact set as in Assumption 3. Let $\eta^{(n)} = [\eta_i^{(n)}]_{i\in[p]} \in \mathbb{R}^p$ and $\tilde{\eta}^{(n)} = [\tilde{\eta}_j^{(n)}]_{j\in[n]} \in \mathbb{R}^n$ be deterministic parameter vectors which elements belong to a compact set as in Assumption 4. Define two functions $h(x, \eta, t)$ and $\tilde{h}(\tilde{x}, \tilde{\eta}, t)$ that both satisfy Assumption 5. Assume furthermore that these functions satisfy non-degeneracy assumptions of the type of Assumption 7, and that all the row and column sums of $B^{(n)}$ are lower bounded by a positive bound independent of n.

With these objects, we define a centered Gaussian $\mathbb{R}^p$–valued process $(\boldsymbol{Z}^{(n),t})_{t\in\mathbb{N}}$, and a centered Gaussian $\mathbb{R}^n$–valued process $(\widetilde{\boldsymbol{Z}}^{(n),t})_{t\in\mathbb{N}_*}$ as follows. Write $\boldsymbol{Z}^{(n),t} = \left[Z_1^{(n),t}, \ldots, Z_p^{(n),t}\right]^\top$, and $\widetilde{\boldsymbol{Z}}^{(n),t} = \left[\widetilde{Z}_1^{(n),t}, \ldots, \widetilde{Z}_n^{(n),t}\right]^\top$, and define the sequences $Z_i^{(n)} = (Z_i^{(n),0}, Z_i^{(n),1}, \ldots)$ for $i \in [p]$, and $\widetilde{Z}_j^{(n)} = (\widetilde{Z}_j^{(n),1}, \widetilde{Z}_j^{(n),2}, \ldots)$ for $j \in [n]$. Then, the sequences $\{Z_i^{(n)}\}_{i=1}^p$ are independent, and so is the case of the sequences $\{\widetilde{Z}_j^{(n)}\}_{j=1}^n$. The covariance matrices $R_i^{(n),t} = \mathbb{C}\text{ov}(Z_i^{(n),0}, \ldots, Z_i^{(n),t})$ and $\widetilde{R}_j^{(n),t} = \mathbb{C}\text{ov}(\widetilde{Z}_j^{(n),1}, \ldots, \widetilde{Z}_j^{(n),t})$ are recursively defined in t as follows.

Writing $\widetilde{\Xi}_j^{(n),0} = \tilde{h}(\tilde{x}_j^0, \tilde{\eta}_j^{(n)}, 0)^2$ for $j \in [n]$, we set

$$R_i^{(n),0} = \sum_{j\in[n]} b_{ij}^{(n)} \widetilde{\Xi}_j^{(n),0}, \quad i \in [p].$$

With these variances at hand, we can construct the variances $\Xi_i^{(n),0} = \mathbb{E} h(Z_i^{(n),0}, \eta_i^{(n)}, 0)^2$ for $i \in [p]$, and then, we set

$$\widetilde{R}_j^{(n),1} = \sum_{i\in[p]} b_{ij}^{(n)} \Xi_i^{(n),0}, \quad j \in [n].$$

Let $t > 0$, and assume that the matrices $\widetilde{R}_j^{(n),t}$ are available. Then we are able to construct the n covariance matrices

$$\widetilde{\Xi}_j^{(n),t} = \mathbb{E} \begin{bmatrix} \tilde{h}(\tilde{x}_j^{(n),0}, \tilde{\eta}_j^{(n)}, 0) \\ \tilde{h}(\widetilde{Z}_j^{(n),1}, \tilde{\eta}_j^{(n)}, 1) \\ \vdots \\ \tilde{h}(\widetilde{Z}_j^{(n),t}, \tilde{\eta}_j^{(n)}, t) \end{bmatrix} \begin{bmatrix} \tilde{h}(\tilde{x}_j^{(n),0}, \tilde{\eta}_j^{(n)}, 0) & \cdots & \tilde{h}(\widetilde{Z}_j^{(n),t}, \tilde{\eta}_j^{(n)}, t) \end{bmatrix},$$

and we can then construct the p covariance matrices

$$R_i^{(n),t} = \sum_{j\in[n]} b_{ij}^{(n)} \widetilde{\Xi}_j^{(n),t}.$$

With these matrices at hand, we are able to write

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

and we set

$$\widetilde{R}_j^{(n),t+1} = \sum_{i\in[p]} b_{ij}^{(n)} \Xi_i^{(n),t}, \quad j \in [n].$$

With these equations, the AMP non-symmetric equations take the form

$$\begin{aligned} x^{(n),t} &= Y^{(n)} \tilde{h}(\tilde{x}^{(n),t}, \tilde{\eta}^{(n)}, t) - \operatorname{diag}\big(B^{(n)} \mathbb{E} \partial \tilde{h}(\widetilde{Z}^{(n),t}, \tilde{\eta}^{(n)}, t)\big) h(x^{(n),t-1}, \eta^{(n)}, t-1) \\ \tilde{x}^{(n),t+1} &= (Y^{(n)})^\top h(x^{(n),t}, \eta^{(n)}, t) - \operatorname{diag}\big((B^{(n)})^\top \mathbb{E} \partial h(Z^{(n),t}, \eta^{(n)}, t)\big) \tilde{h}(\tilde{x}^{(n),t}, \tilde{\eta}^{(n)}, t), \end{aligned}$$

where the vectors $\tilde{h}(\tilde{x}^t, \tilde{\eta}, t) \in \mathbb{R}^n$ and $h(x^t, \eta, t) \in \mathbb{R}^p$ are defined in the obvious way, and where $h(x^{-1}, \eta, -1)$ is zero.

Furthermore, considering an arbitrary integer $t_{\max} \geqslant 0$, two test functions $\varphi, \tilde\varphi \in \mathrm{PL}_2(\mathbb{R}^{t_{\max}+2})$, and for each n, a p–uple $(\beta_i^{(n)})_{i\in[p]}$ and a n–uple $(\tilde\beta_j^{(n)})_{j\in[n]}$ which elements are uniformly bounded as in the statement of Theorem 2, it holds that

$$\mathbb{P}\left[\left[\left|\frac{1}{n}\sum_{i\in[p]} \beta_i^{(n)}\varphi(\eta_i^{(n)}, x_i^{(n),0}, \ldots, x_i^{(n),t_{\max}}) - \beta_i^{(n)}\mathbb{E}\varphi(\eta_i^{(n)}, Z_i^{(n),0}, \ldots, Z_i^{(n),t_{\max}})\right| \geqslant \varepsilon\right] \cap \left[\|Y^{(n)}\| \leqslant C_Y\right]\right] \xrightarrow[n\to\infty]{} 0,$$

and

$$\mathbb{P}\left[\left[\left|\frac{1}{n}\sum_{j\in[n]} \tilde\beta_j^{(n)}\tilde\varphi(\tilde\eta_j^{(n)}, \tilde x_j^{(n),1}, \ldots, \tilde x_j^{(n),t_{\max}+1}) - \tilde\beta_j^{(n)}\mathbb{E}\tilde\varphi(\tilde\eta_j^{(n)}, \tilde Z_j^{(n),1}, \ldots, \tilde Z_j^{(n),t_{\max}+1})\right| \geqslant \varepsilon\right] \cap \left[\|Y^{(n)}\| \leqslant C_Y\right]\right] \xrightarrow[n\to\infty]{} 0$$

for each $\varepsilon > 0$ and each $C_Y > 0$.

These expressions can be deduced from the symmetric case described by Theorem 2 by replacing the matrix W there with the $(n+p)\times(n+p)$ symmetric matrix $\begin{bmatrix} 0 & Y^{(n)} \\ (Y^{(n)})^\top & 0 \end{bmatrix}$, and by choosing the activation function adequately. More details on this passage from the symmetric to the non-symmetric case are provided in [JM13].

An analogue of Theorem 4 can also be devised for the non-symmetric case. We omit the related details.

3 Application to the equilibria of LV systems

We now apply the previous results to the study of the distribution of the globally stable equilibria of large Lotka-Volterra ecological systems.

Let $\mathbb{R}_{*+} = (0,\infty)$. Keeping the sequence (n) of integers introduced in Section 2, let $(r^{(n)})$ be a sequence of vectors such that $r^{(n)} \in \mathbb{R}^n_{*+}$, and let $(\Sigma^{(n)})$ be a sequence of random symmetric matrices such that $\Sigma^{(n)} \in \mathbb{R}^{n\times n}$. For a given n, starting with a vector $u^{(n)}(0) \in \mathbb{R}^n_{*+}$, our LV ODE is given as

$$\dot u^{(n)}(t) = u^{(n)}(t) \odot \left(r^{(n)} + \left(\Sigma^{(n)} - I_n\right) u^{(n)}(t)\right), \quad t \geqslant 0. \tag{9}$$

We recall that $r^{(n)}$ is the so-called vector of intrinsic growth rates of the species, and the matrix $\Sigma^{(n)}$ is the species interaction matrix. When this ODE has a unique solution $u^{(n)} : \mathbb{R}_+ \to \mathbb{R}^n_+$ for each $u^{(n)}(0) \in \mathbb{R}^n_{*+}$, and when the image of this function is pre-compact, we say that the ODE is well-defined. A sufficient condition for well-definiteness is $\|\Sigma^{(n)}\| < 1$ [LJM09]. It is also well-known that under this condition, the ODE (9) has a globally stable equilibrium $u_\star^{(n)} \in \mathbb{R}^n_+$ in the classical sense of the Lyapounov theory [Tak96, Chap. 3]. We shall focus herein on the distribution $\mu^{u_\star^{(n)}} \in \mathcal{P}(\mathbb{R}_+)$ of the elements of $u_\star^{(n)}$ for the large values of n, which reflects the relative abundances of the species at the equilibrium.

We now state our assumptions regarding the vectors $r^{(n)}$ and the matrices $\Sigma^{(n)}$. Let us call "hypotheses" the assumptions relative to our LV system. Our first hypothesis reads as follows:

Hypothesis 1. Let $X^{(n)}$ be the random symmetric $n \times n$ matrix constructed after Assumption 1 above. Let $V^{(n)} = \left[v_{ij}^{(n)}\right]_{i,j=1}^n$ be a deterministic symmetric matrix with the same structure as

$S^{(n)}$ in Section 2, and that complies with Assumption 2 above, with the $s_{ij}^{(n)}$ in this assumption being replaced with $v_{ij}^{(n)}$. Furthermore, assume that there is a constant $c_V > 0$ such that

$$\min_{i\in[n]} \sum_{j\in[n]} v_{ij}^{(n)} \geqslant c_V$$

(see Assumption 7). Then, the matrices $\Sigma^{(n)}$ are written as

$$\Sigma^{(n)} = (V^{(n)})^{\odot 1/2} \odot X^{(n)}.$$

According to our model, Species i interacts only with the species in the set $\mathcal{V}_i^{(n)} = \{j \in [n],\ v_{ij}^{(n)} > 0\}$. The relative cardinalities of these sets satisfy $\min_i(|\mathcal{V}_i^{(n)}|/n) \sim K_n/n$ and $\max_i(|\mathcal{V}_i^{(n)}|/n) \sim K_n/n$. This ratio is the "degree of sparsity" of our model,

Since $\Sigma^{(n)}$ is random, we need to guarantee that $\limsup_n \|\Sigma^{(n)}\| < 1$ with probability one to ensure that the LV ODE is well-defined and has a global equilibrium. We consider this condition as a hypothesis:

Hypothesis 2. $\limsup_n \|\Sigma^{(n)}\| < 1$ with probability one.

Let us provide some comments on this hypothesis. The spectral norms of random matrices with structures close to W were studied in [BvH16] in the framework of Assumption 8. In this setting, explicit bounds on the spectral norm of $\Sigma^{(n)}$ can be obtained in some cases, among which the Gaussian case plays a prominent role. We note that in general, these bounds are not tight. Let $|||A|||$ and $\|A\|_\infty$ be respectively the maximum row sum norm and the max norm of the matrix A respectively. When $X^{(n)}$ is a Gaussian matrix, we know from [BvH16, Th. 1.1] that

$$\mathbb{E}\|\Sigma^{(n)}\| \leqslant T_{\text{Gauss}}^{(n)},$$

where

$$T_{\text{Gauss}}^{(n)} = (1+\varepsilon)\left(2|||V^{(n)}|||^{1/2} + \frac{6}{\sqrt{\log(1+\varepsilon)}}(\|V^{(n)}\|_\infty \log n)^{1/2}\right)$$

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

$$\mathbb{P}\left[\|\Sigma^{(n)}\| \geqslant T_{\text{Gauss}}^{(n)} + t\right] \leqslant \exp(-t^2/(2\|V^{(n)}\|_\infty)^2), \tag{10}$$

for $\varepsilon \in (0, 1/2]$, as given by [BvH16, Cor. 3.9]. By consequence, it holds that in the Gaussian case, $\limsup_n \|\Sigma^{(n)}\| < 1$ with probability one if $\limsup_n T_{\text{Gauss}}^{(n)} < 1$ for some choice of $\varepsilon \in (0, 1/2]$, and if the sequence (K_n) satisfies $K_n \gtrsim \log n$, owing to the constraint on $\|V^{(n)}\|_\infty$ provided in Assumption 2. In the Gaussian case, we can therefore admit a degree of sparsity of $\log n/n$. Under our hypothesis 1, this rate turns out to be the optimal rate for Hypothesis 2 to be verified, as a consequence of [BvH16, Cor. 3.15].

Similar explicit bounds on $\limsup_n \|\Sigma^{(n)}\|$ are available in the case where the random variables $X_{ij}^{(n)}$ are bounded, see [BvH16, Cor. 3.12] for more details. Here also, Hypothesis 2 is ensured when $|||V^{(n)}|||$ and $\|V^{(n)}\|_\infty \log n$ are bounded properly.

Under Hypothesis 1 without more specific information on the distributions of the X_{ij}, Hypothesis 2 is verified if both $\limsup_n |||V^{(n)}|||$ and $\limsup_n \|V^{(n)}\|_\infty (\log n)^{\rho\vee 1}$ are small enough, as shown by Proposition 3. We thus need that $K_n \gtrsim (\log n)^{\rho\vee 1}$ with a large enough factor.

In any case, we shall also need

Hypothesis 3. $\limsup_n |||V^{(n)}||| < 1/4$.

In the classical Wigner case where $V^{(n)} = \alpha n^{-1} 1_n 1_n^\top$ for some $\alpha > 0$, this condition is necessary and sufficient to ensure that $\limsup_n \|\Sigma^{(n)}\| < 1$ with probability one, by the well-known result on the almost sure convergence of $\|\Sigma^{(n)}\|$ to the edge of the semi-circle law. Sparse cases that behave similarly to the Wigner case in this respect, and thus, that require Hypothesis 3, were treated in the literature. These include the models dealt with in [BvH16, Sec. 4]. Similar results can be found in [Kho08, Sod10, BGP14].

Regarding the intrinsic growth rate vector, we consider the following hypothesis:

Hypothesis 4. The vector $r^{(n)}$ is a deterministic vector which elements belong to a compact $\mathcal{Q}_r \subset \mathbb{R}_{*+}$.

In the practical ecological settings where the intrinsic growth rates within a given ecosystem are positive, it is not unrealistic to assume that these rates are lower and upper bounded by positive constants.

Let $x, y \in \mathbb{R}^n$. In the statement and in the proof of the following theorem, we denote as $xy, x/y, \sqrt{x}$, etc., the $\mathbb{R}^n$–valued vector obtained by performing these operations elementwise. Given a $\mathbb{R}^n$–valued random vector $U = [U_i]_{i=1}^n$, the notations $\mathbb{E}U$ and $\mathbb{P}[U \geqslant 0]$ will respectively refer in the remainder to the vectors $[\mathbb{E}U_i]_{i=1}^n$ and $[\mathbb{P}[U_i \geqslant 0]]_{i=1}^n$.

Theorem 5. Let Hypotheses 1 to 4 hold true. Then, for each n for which $\|\Sigma^{(n)}\| < 1$, the ODE (9) is well-defined, and it has a globally stable equilibrium $u_\star^{(n)}$. For the other values of n, put $u_\star^{(n)} = 0$. The distribution $\mu^{u_\star^{(n)}}$ is a $\mathcal{P}_2(\mathbb{R})$–valued random variable on the probability space where $\Sigma^{(n)}$ is defined.

Let $\xi^{(n)} \sim \mathcal{N}(0, I_n)$. Then, for each n large enough, the system of equations

$$p = V^{(n)} \operatorname{diag}(1+\zeta)^2 \, \mathbb{E} \left(\sqrt{p}\xi^{(n)} + r^{(n)} \right)_+^2 \tag{11a}$$

$$\zeta = \operatorname{diag}(1+\zeta) V^{(n)} \operatorname{diag}(1+\zeta) \; \mathbb{P}\left[\sqrt{p}\xi^{(n)} + r^{(n)} \geqslant 0 \right]. \tag{11b}$$

admits an unique solution $(p, \zeta) = (p^{(n)}, \zeta^{(n)}) \in \mathbb{R}_+^n \times [0,1]^n$. This solution satisfies

$$\sup_n \|p^{(n)}\|_\infty < \infty.$$

Let $Y^{(n)}$ be the Gaussian vector

$$Y^{(n)} = \left[Y_i^{(n)} \right]_{i=1}^n = \left(1 + \zeta^{(n)} \right) \left(\sqrt{p^{(n)}} \xi^{(n)} + r^{(n)} \right),$$

and define the deterministic measure $\boldsymbol{\mu}^{(n)} \in \mathcal{P}_2(\mathbb{R})$ as $\boldsymbol{\mu}^{(n)} = \mathscr{L}((Y_{\theta^{(n)}}^{(n)})_+)$, where $\theta^{(n)}$ is a uniformly distributed random variable on the set $[n]$, which is independent of $\xi^{(n)}$. Then

$$\boldsymbol{d}_2 \left(\mu^{u_\star^{(n)}}, \boldsymbol{\mu}^{(n)} \right) \xrightarrow[n\to\infty]{\mathcal{P}} 0.$$

Remark 1. Since $\sup_n \|p^{(n)}\|_\infty < \infty$, the sequence of measures $(\boldsymbol{\mu}^{(n)})$, which are mixtures of truncated Gaussians with bounded means and variances, is a pre-compact sequence in the space $\mathcal{P}_2(\mathbb{R})$. Therefore, for each sub-sequence of (n), there is a further sub-sequence (n') and a measure $\boldsymbol{\nu} \in \mathcal{P}_2(\mathbb{R})$ such that $\mu^{u_\star^{(n')}} \to \boldsymbol{\nu}$ in probability in $\mathcal{P}_2(\mathbb{R})$.

Remark 2. Write $u_\star^{(n)} = [u_{\star,i}^{(n)}]_{i=1}^n$. By Theorem 5,

$$\frac{1}{n} \sum_{i\in[n]} \varphi(u_{\star,i}^{(n)}) - \mathbb{E}\varphi((Y_{\theta^{(n)}}^{(n)})_+) \xrightarrow[n\to\infty]{\mathcal{P}} 0$$

for each continuous function $\varphi : \mathbb{R} \to [0,1]$ such that $\varphi(0) = 0$. With this at hand, we can consider the positive number $\gamma^{(n)} = \mathbb{P}[Y_{\theta^{(n)}}^{(n)} > 0]$ as an approximation of the proportion of surviving species at the equilibrium, in the sense that $\gamma^{(n)} \simeq n^{-1} \sum_{i\in[n]} \varphi(u_{\star,i}^{(n)})$ when φ is chosen as a "continuous approximation" of the real function $\psi(x) = \mathbf{1}_{x>0}$. Of course, this does not imply that $\gamma^{(n)} - n^{-1}\sum_i \mathbf{1}_{u^{(n)}_{\star,i}>0}$ converges to zero in probability.

Note that $\liminf_n \gamma^{(n)} > 1/2$ by Hypothesis 4 and by the bound $\sup \|p^{(n)}\|_\infty < \infty$, which implies that the proportion of species that survive at the equilibrium for large n is lower bounded by $1/2$ with a positive gap.

The following corollary to Theorem 5 shows that when the variance profile matrix is $V^{(n)} = \alpha n^{-1} 1_n 1_n^\top$ with $\alpha \in (0, 1/4)$, then, the large–n behavior of the empirical measure $\mu^{u^{(n)}_\star}$ is the one predicted by [Bun17, Gal18, ABC⁺22]. Remember that we are not restricted in the present paper to the Gaussian case.

Corollary 6. In the setting of the previous theorem, assume that $V^{(n)} = \alpha n^{-1} 1_n 1_n^\top$ with $\alpha \in (0, 1/4)$. Let $\underline{\xi} \sim \mathcal{N}(0,1)$, and let $\theta^{(n)}$ be independent of $\underline{\xi}$ and uniformly distributed on the set $[n]$. Then, there is an unique couple $(\underline{p}^{(n)}, \underline{\zeta}^{(n)}) \in \mathbb{R}_+ \times [0,1]$ such that, dropping the superscript $^{(n)}$,

$$\begin{aligned}\underline{p} &= \alpha(1+\underline{\zeta})^2 \mathbb{E}_\theta \mathbb{E}_{\underline{\xi}}\big[\big(\sqrt{\underline{p}}\,\underline{\xi} + r_\theta\big)_+^2\big], \\ \underline{\zeta} &= \alpha(1+\underline{\zeta})^2 \mathbb{E}_\theta \mathbb{P}_{\underline{\xi}}\big[\sqrt{\underline{p}}\,\underline{\xi} + r_\theta \geqslant 0\big].\end{aligned}$$

Moreover,

$$\boldsymbol{d}_2\left(\mu^{u_\star}, \mathscr{L}\left((1+\underline{\zeta})\big(\sqrt{\underline{p}}\,\underline{\xi} + r_\theta\big)_+\right)\right) \xrightarrow[n\to\infty]{\mathcal{P}} 0.$$

To prove this corollary, we notice that due to the structure of V, the solution of the system (11) is of the form $(p^{(n)}, \zeta^{(n)}) = (\underline{p}^{(n)} 1_n, \underline{\zeta}^{(n)} 1_n)$ where $(p^{(n)}, \zeta^{(n)})$ is the solution of the system shown in the statement. The result follows.

We now turn to our proofs. In the remainder of this paper, the notations $C, c > 0$ refer to constants that can change from line to line. Some of the statements hold true for n large enough instead of holding for each n. This will not be specified.

4 Proofs of Theorems 2 and 4

The first and most important part of this section is common to the proofs of Theorems 2 and 4. The idea is to consider an AMP sequence where the function $h(\cdot, \eta, t)$ is replaced with a polynomial, and to apply to this sequence the combinatorial technique of Bayati, Lelarge and Montanari in [BLM15].

In the remainder, $c > 0$ and $C > 0$ denote absolute constants that can change from a display to another.

4.1 AMP with polynomial activation functions

4.1.1 The setting

Let $d > 0$ be a fixed integer. Consider a function $p^{(n)} : \mathbb{R} \times [n] \times \mathbb{N} \to \mathbb{R}$ with the property that $p^{(n)}(\cdot, i, t)$ is a polynomial with a degree bounded by d, and is thus written as

$$p^{(n)}(u, i, t) = \sum_{\ell=0}^{d} \alpha_\ell^{(n)}(i, t) u^\ell.$$

Starting with the deterministic vector $\check{x}^{(n),0} = x^{(n),0}$, our purpose here is to study the AMP recursion in the vectors $\check{x}^{(n),1}, \check{x}^{(n),2}, \ldots$ given as

$$\check{x}^{(n),t+1} = W^{(n)} p^{(n)}(\check{x}^{(n),t}, \cdot, t) - \operatorname{diag}\Big((W^{(n)})^{\odot 2} \partial p^{(n)}(\check{x}^{(n),t}, \cdot, t)\Big) p(\check{x}^{(n),t-1}, \cdot, t-1), \tag{12}$$

where, writing $x = [x_i]_{i=1}^n$, we set $p^{(n)}(x, \cdot, t) = [p^{(n)}(x_i, i, t)]_{i=1}^n$ and $\partial p^{(n)}(x, \cdot, t) = [\partial p^{(n)}(\cdot, i, t)|_{x_i}]_{i=1}^n$, and we consider that $p^{(n)}(\cdot, \cdot, -1) \equiv 0$.

The general idea goes as follows. With the polynomial activation function $p^{(n)}(\cdot, i, t)$, the elements of the vectors $\breve{x}^{(n),t}$ are random variables which are sums of monomials in the elements of $W^{(n)}$ with indices indexed by labelled trees. Our purpose will be to compute the moments of these random variables. Due to the presence of the "correction" so-called Onsager term $\operatorname{diag}(\cdot)p(\cdot)$, the effect of the paths with the so-called backtracking edges will be cancelled, rendering these moments close to the Gaussian moments provided by the state evolution equations below. A nice intuitive explanation of the main proof idea is provided in [BLM15, § 2].

Let us provide the expressions of the state evolution equations for Algorithm (12). Consider the Gaussian sequence $(\breve{\boldsymbol{Z}}^{(n),t})_{t\in\mathbb{N}_*}$ described as follows. Writing $\breve{\boldsymbol{Z}}^{(n),t} = \left[\breve{Z}_1^{(n),t}, \ldots, \breve{Z}_n^{(n),t}\right]^\top$ and $\breve{Z}_i^{(n)} = (\breve{Z}_i^{(n),1}, \breve{Z}_i^{(n),2}, \ldots)$, the sequences $\{\breve{Z}_i^{(n)}\}_{i=1}^n$ are centered, Gaussian, and independent, and the covariance matrix $\breve{R}_i^{(n),t}$ of the vector $\vec{\breve{Z}}_i^{(n),t} = [\breve{Z}_i^{(n),1}, \ldots, \breve{Z}_i^{(n),t}]^\top$ is constructed recursively in t as follows.

$$\breve{R}_i^{(n),1} = \sum_{l\in[n]} s_{il}^{(n)} p^{(n)}(x_l^{(n),0}, l, 0)^2,$$

and

$$\breve{R}_i^{(n),t+1} = \sum_{l\in[n]} s_{il}^{(n)} \mathbb{E} \begin{bmatrix} p^{(n)}(x_l^{(n),0}, l, 0) \\ p^{(n)}(\breve{Z}_l^{(n),1}, l, 1) \\ \vdots \\ p^{(n)}(\breve{Z}_l^{(n),t}, l, t) \end{bmatrix} \begin{bmatrix} p^{(n)}(x_l^{(n),0}, l, 0) & p^{(n)}(\breve{Z}_l^{(n),1}, l, 1) & \cdots & p^{(n)}(\breve{Z}_l^{(n),t}, l, t) \end{bmatrix}.$$

With this at hand, the main result of this subsection reads as follows.

Proposition 7. Let Assumptions 1–3 hold true, and consider the AMP algorithm (12). Assume that for each integer $t \geqslant 0$,

$$\sup_n \max_{\ell\leqslant d} \max_{i\in[n]} |\alpha_\ell^{(n)}(i, t)| < \infty. \tag{13}$$

Then,

$$\forall t > 0, \quad \sup_n \max_{i\in[n]} \|\breve{R}_i^{(n),t}\| < \infty.$$

Moreover,

$$\forall t > 0, \ \forall m \in \mathbb{N}_*, \quad \sup_n \max_{i\in[n]} \mathbb{E}|\breve{x}_i^{(n),t}|^m < \infty. \tag{14}$$

Given $t \in \mathbb{N}_*$, let $\psi^{(n)} : \mathbb{R}^t \times [n] \to \mathbb{R}$ be such that $\psi^{(n)}(\cdot, l)$ is a multivariate polynomial with a bounded degree and bounded coefficients as functions of (l, n). Let $\mathcal{S}^{(n)} \subset [n]$ be such that $|\mathcal{S}^{(n)}| \leqslant CK^{(n)}$. Then,

$$\frac{1}{|\mathcal{S}^{(n)}|} \sum_{i\in\mathcal{S}^{(n)}} \psi^{(n)}(\breve{x}_i^{(n),1}, \ldots, \breve{x}_i^{(n),t}, i) - \mathbb{E}\psi^{(n)}(\breve{Z}_i^{(n),1}, \ldots, \breve{Z}_i^{(n),t}, i) \xrightarrow[n\to\infty]{\mathcal{P}} 0, \quad \text{and} \tag{15a}$$

$$\frac{1}{n} \sum_{i\in[n]} \psi^{(n)}(\breve{x}_i^{(n),1}, \ldots, \breve{x}_i^{(n),t}, i) - \mathbb{E}\psi^{(n)}(\breve{Z}_i^{(n),1}, \ldots, \breve{Z}_i^{(n),t}, i) \xrightarrow[n\to\infty]{\mathcal{P}} 0. \tag{15b}$$

The proof of this proposition will revolve around a version of our AMP sequence where the $\mathbb{R}^n$–valued vectors $\breve{x}^{(n),t}$ will be replaced with $\mathbb{R}^{n\times q}$–valued matrices, with $q > 0$ being a given integer. At the same time, the test function $\psi^{(n)}(\cdot, i)$ acting on the iterates $\breve{x}_i^{(n),1}, \ldots, \breve{x}_i^{(n),t}$ above will be replaced with a test function acting on the t^{th} iterate only. In all the remainder, a vector

$\boldsymbol{x} \in \mathbb{R}^q$ will be written as $\boldsymbol{x} = [x(1), \ldots, x(q)]^\top$. Consider the function

$$f^{(n)} \; : \; \mathbb{R}^q \times [n] \times \mathbb{N} \longrightarrow \mathbb{R}^q$$
$$(\boldsymbol{u}, l, t) \longmapsto \begin{bmatrix} f_1^{(n)}(\boldsymbol{u}, l, t) \\ \vdots \\ f_q^{(n)}(\boldsymbol{u}, l, t) \end{bmatrix},$$

which is a polynomial in $\boldsymbol{u}$, with degree bounded by d, written as

$$f_r^{(n)}(\boldsymbol{u}, l, t) = \sum_{i_1+\cdots+i_q=0}^{d} \alpha_{i_1,\ldots,i_q}^{(n)}(r, l, t) \prod_{s=1}^{q} u(s)^{i_s}.$$

Starting with the deterministic n-uple $(\boldsymbol{x}_1^{(n),0}, \ldots, \boldsymbol{x}_n^{(n),0})$ with $\boldsymbol{x}_i^{(n),0} \in \mathbb{R}^q$, the AMP recursion in t will provide us at Iteration $t+1$ with the n-uple $(\boldsymbol{x}_1^{(n),t+1}, \ldots, \boldsymbol{x}_n^{(n),t+1})$, where $\boldsymbol{x}_i^{(n),t+1} = [x_i^{(n),t+1}(1), \ldots, x_i^{(n),t+1}(q)]^\top$ is given as

$$x_i^{(n),t+1}(r) = \sum_{l\in[n]} W_{i,l}^{(n)} f_r^{(n)}(\boldsymbol{x}_l^{(n),t}, l, t) - \sum_{s\in[q]} f_s^{(n)}(\boldsymbol{x}_i^{(n),t-1}, i, t-1) \sum_{l\in[n]} (W_{i,l}^{(n)})^2 \frac{\partial f_r^{(n)}}{\partial x(s)}(\boldsymbol{x}_l^{(n),t}, l, t), \quad r \in [q], \tag{16}$$

with $f(\cdot, \cdot, -1) \equiv 0$.

We also introduce the following sequence of Gaussian $\mathbb{R}^{nq}$–valued random vectors $(U^{(n),t})_{t\in\mathbb{N}_*}$. Writing

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

the $\{U_i^{(n),t}\}_{i\in[n]}$ are $\mathbb{R}^q$–valued independent Gaussian random vectors, $U_i^{(n),t} \sim \mathcal{N}\left(0, Q_i^{(n),t}\right)$, and the covariance matrices $Q_i^{(n),t}$ are recursively defined in t as follows. Starting with $t = 1$, we set

$$Q_i^{(n),1} = \sum_{l\in[n]} s_{il}^{(n)} f^{(n)}(\boldsymbol{x}_l^{(n),0}, l, 0) f^{(n)}(\boldsymbol{x}_l^{(n),0}, l, 0)^\top \quad \text{for } i \in [n].$$

Given $\{Q_i^{(n),t}\}_{i\in[n]}$, we set

$$Q_i^{(n),t+1} = \sum_{l\in[n]} s_{il}^{(n)} \mathbb{E} f^{(n)}(U_l^{(n),t}, l, t) f^{(n)}(U_l^{(n),t}, l, t)^\top \quad \text{for } i \in [n].$$

The correlations between the elements of $U^{(n),t}$ and $U^{(n),s}$ for $t \neq s$ are irrelevant to our purpose.

Given $\boldsymbol{x} \in \mathbb{R}^q$ and a multi-index $\boldsymbol{m} = \left[m(1), \ldots, m(q)\right] \in \mathbb{N}^q$, we write in the sequel

$$\boldsymbol{x}^{\boldsymbol{m}} = \prod_{r\in[q]} x(r)^{m(r)}.$$

Proposition 8. Let Assumptions 1 and 2 hold true, and consider the iterative algorithm(16). Assume that for each $t > 0$, there is a constant $C > 0$ such that

$$\left|\alpha_{i_1,\ldots,i_q}^{(n)}(r, l, t)\right| \leqslant C, \tag{17}$$

and furthermore,

$$\sup_n \max_{i\in[n]} \left\| \boldsymbol{x}_i^{(n),0} \right\| < \infty. \tag{18}$$

Then,

$$\forall t > 0, \quad \sup_n \max_{i\in[n]} \|Q_i^{(n),t}\| < \infty. \tag{19}$$

Moreover,

$$\forall t > 0,\ \forall \boldsymbol{m} \in \mathbb{N}^q, \quad \sup_n \max_{i\in[n]} \mathbb{E}|(\boldsymbol{x}_i^t)^{\boldsymbol{m}}| < \infty. \tag{20}$$

Let $\psi^{(n)} : \mathbb{R}^q \times [n] \to \mathbb{R}$ be such that $\psi^{(n)}(\cdot, l)$ 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 $|\mathcal{S}^{(n)}| \leqslant CK^{(n)}$. Then,

$$\frac{1}{K^{(n)}} \sum_{i\in\mathcal{S}^{(n)}} \psi(\boldsymbol{x}_i^{(n),t}, i) - \mathbb{E}\psi(U_i^{(n),t}, i) \xrightarrow[n\to\infty]{\mathcal{P}} 0, \quad \text{and} \tag{21a}$$

$$\frac{1}{n} \sum_{i\in[n]} \psi(\boldsymbol{x}_i^{(n),t}, i) - \mathbb{E}\psi(U_i^{(n),t}, i) \xrightarrow[n\to\infty]{\mathcal{P}} 0. \tag{21b}$$

The proof of Proposition 8 is an adaptation of the approach of [BLM15] to the structure of the variance profile at interest in this paper. For self-containedness, we reproduce large parts of the proof of [BLM15], putting the focus on the parts of this proof where these adaptations are necessary.

In the remainder, the superscript $^{(n)}$ will be often omitted for notational simplicity.

4.1.2 Proof of Proposition 8

Let us quickly prove that $\sup_n \max_{i\in[n]} \|Q_i^t\| < \infty$ for each integer $t > 0$. For $t = 1$, this is a consequence of Assumption 2 and the bounds (17) and (18). Assume the result is true for t. Then, using the bound (17) again, standard Gaussian derivations show that $\sup_n \max_i \left\|\mathbb{E} f(U_i^t, i, t) f(U_i^t, i, t)^\top\right\| < \infty$. Then, using Assumption 2, we obtain from the expression of Q_i^{t+1} above that $\sup_n \max_i \left\|Q_i^{t+1}\right\| < \infty$.

Of importance in the proof of Proposition 8 are the sequences issued from the so-called "non-backtracking" iterations. Given any $i, j \in [n]$ with $i \neq j$, define the set of $\mathbb{R}^q$–valued vectors $\{\boldsymbol{z}_{i\to j}^0,\ i, j \in [n],\ i \neq j\}$ as $\boldsymbol{z}_{i\to j}^0 = \boldsymbol{x}_i^0$. Assuming that the $\mathbb{R}^q$–valued vectors $\{\boldsymbol{z}_{i\to j}^t,\ i, j \in [n],\ i \neq j\}$ are defined, the vectors $\boldsymbol{z}_{i\to j}^{t+1}$ for $i \neq j$ are given as

$$z_{i\to j}^{t+1}(r) = \sum_{l\in[n]\setminus\{j\}} W_{il} f_r(\boldsymbol{z}_{l\to i}^t, l, t) \tag{22}$$

(here, we implicitly consider that in the summation over l, the case $l = i$ is excluded because $W_{ii} = 0$). Having the vectors $\{\boldsymbol{z}_{i\to j}^t,\ i, j \in [n],\ i \neq j\}$ at hand, we define the $\mathbb{R}^q$–valued vectors $\{\boldsymbol{z}_i^{t+1},\ i \in [n]\}$ as

$$z_i^{t+1}(r) = \sum_{l\in[n]} W_{il} f_r(\boldsymbol{z}_{l\to i}^t, l, t). \tag{23}$$

For each n, let us now consider an i.i.d. sequence $(W^{(n),t})_{t=0,1,\ldots}$ of symmetric $n \times n$ matrices such that $W^{(n),t} \stackrel{\mathcal{L}}{=} W^{(n)}$. 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 $\{\boldsymbol{y}_{i\to j}^t,\ i, j \in [n],\ i \neq j\}$, we set

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

Also,

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

The tree structure. The pivotal object in the proof is a rooted and labelled tree T with the following structure (we paraphrase here [BLM15, §4.2] with the same notations). Write $T = (V(T), E(T))$ where $V(T)$ and $E(T)$ are the sets of vertices and edges of T respectively. We also denote as $L(T)$ the set of the leaves of T. Denote as $\circ \in V(T)$ the root of the tree, and let $|u|$ be the distance of a vertex u to $\circ$. We consider that the edges are oriented towards the root. Thus, considering the edge $(u \to v) \in E(T)$, we have that $v = \pi(u)$, where $\pi(u)$ the parent of the node u. We assume that the root has only one child (thus, T is a so called planted tree). Each vertex other than the root and the leaves can have up to d children, thus, its degree is bounded by $d+1$. By definition, the degree of a leaf is one, its only neighbor being its parent.

The tree T is a labelled tree. We now describe this labelling. The label of the root is an integer $\ell(\circ) \in [n]$. The label of a vertex v which is neither the root nor a leaf is a couple $(\ell(v), r(v)) \in [n] \times [q]$. The label of a leaf v is $(\ell(v), r(v), v[1], \ldots, v[q]) \in [n] \times [q] \times \{(a_1, \ldots, a_q) \in \mathbb{N}^q,\ a_1 + \cdots + a_q \leqslant d\}$. If $v \in L(T)$ and $|v| \leqslant t-1$, then we set $v[1] = \cdots = v[q]) = 0$.

The integer $\ell(v)$ is denoted as the "type" of the vertex v, and $r(v)$ is the "mark" of this vertex when it exists.

For a vertex $u \in V(T) \backslash L(T)$, we denote as $u[r]$ the number of children of u with the mark r. The children of $u \in V(T) \backslash L(T)$ are ordered with respect to their mark. Specifically, the labels of the children of u are $(\ell^1, 1), \ldots, (\ell^{u[1]}, 1), (\ell^{u[1]+1}, 2), \ldots, (\ell^{u[1]+\cdots+u[q]}, q)$

Specific classes of these trees will be of importance. Following the definitions and notations of [BLM15], we introduce the following families of trees:

- $\overline{\mathcal{T}}^t$ is the set of labelled trees as above, with depth t at most.
- $\mathcal{T}^t \subset \overline{\mathcal{T}}^t$ is the subset that additionally satisfies the following so-called non-backtracking condition: if $v_1 = \circ, v_2, \ldots, v_k$ is a path starting from $\circ$, *i.e.*, $v_i = \pi(v_{i+1})$, then the corresponding sequence of types is non-backtracking. This means that for each $i \in [k-2]$, the three types $\ell(v_i)$, $\ell(v_{i+1})$ and $\ell(v_{i+2})$ are distinct.
- $\mathcal{T}^t_{i \to j}(r) \subset \mathcal{T}^t$ is 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 $r(v) = r$.
- $\mathcal{T}^t_i(r) \subset \mathcal{T}^t$ is 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 $r(v) = r$.

Given a tree T, we also set

$$W(T) = \prod_{(u \to v) \in E(T)} W_{\ell(u)\ell(v)},$$

$$\Gamma(T, \boldsymbol{\alpha}, t) = \prod_{(u \to v) \in E(T)} \alpha_{u[1], \ldots, u[q]}(r(u), \ell(u), t - |u|),$$

$$x(T) = \prod_{v \in L(T)} \prod_{s \in [q]} \left(x^0_{\ell(v)}(s) \right)^{v[s]}.$$

Lemma 9 (Lemma 1 of [BLM15])**.**

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

$$z^t_i(r) = \sum_{T \in \mathcal{T}^t_i(r)} W(T)\Gamma(T, \boldsymbol{\alpha}, t) x(T).$$

This lemma is a structural lemma which proof is not impacted by our specific variance profile. The following notions will be needed below. Define the set

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

For each $i \in [n]$, we define the section

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

The next proposition shows that in the large dimensional regime, the joint moments of the elements of a vector $\boldsymbol{z}_i^t$ issued from the non-backtracking iterations depend for large n only on the first two moments of the elements of W.

Proposition 10 (adaptation of Proposition 1 of [BLM15])**.** Let $\{\widetilde{X}_{ij}\}_{1\leqslant i<j\leqslant n}$ be a family of independent random variables satisfying Assumption 1, with distributions not necessarily identical to their analogues X_{ij}. Let $\widetilde{W}$ be the symmetric 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 $\{\tilde{\boldsymbol{z}}^0_{i\to j},\ i,j \in [n],\ i \neq j\}$ given as $\tilde{\boldsymbol{z}}^0_{i\to j} = \boldsymbol{x}^0_i$, define the vectors $\tilde{\boldsymbol{z}}^t_i \in \mathbb{R}^q$ by the recursion (22) and the equation (23), where W is replaced with $\widetilde{W}$. Then, for each $t \geqslant 1$ and each $\boldsymbol{m} \in \mathbb{N}^q$, there exists C such that for each $i \in [n]$,

$$\left|\mathbb{E}(\boldsymbol{z}_i^t)^{\boldsymbol{m}} - \mathbb{E}(\tilde{\boldsymbol{z}}_i^t)^{\boldsymbol{m}}\right| \leqslant CK^{-1/2}.$$

Proof. For simplicity, we restrict the proof to the case where the multi-index $\boldsymbol{m}$ satisfies $m(r) = m$ for some integer $m > 0$ and $m(s) = 0$ for $s \in [q]\backslash\{r\}$. By Lemma 9, we have

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

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

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

There is an integer constant $C_E = C_E(d, t, m)$ that bounds the total number of edges in the trees $T_1, \ldots, T_m \in \mathcal{T}_i^t(r)$, leading to

$$\sum_{k\in[m]} \sum_{j<l} \phi(T_k)_{jl} \leqslant C_E.$$

Given an integer $\mathfrak{m} \in [C_E]$, let

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

Then, since the elements of W beneath the diagonal are centered and independent,

$$\mathbb{E}z_i^t(r)^m = \sum_{\mathfrak{m}=1}^{C_E} \sum_{(T_1,\ldots,T_m)\in\mathcal{A}_i(\mathfrak{m})} \left(\prod_{k=1}^m \Gamma(T_k, \boldsymbol{\alpha}, t)\right) \left(\prod_{k=1}^m x(T_k)\right) \mathbb{E}\left[\prod_{k=1}^m W(T_k)\right]. \tag{24}$$

Notice that the contributions of the m–uples of trees in $\mathcal{A}_i(\mathfrak{m})$ for which

$$\forall j < l,\ \sum_{k\in[m]} \phi(T_k)_{jl} \in \{0, 2\}$$

are the same for $\mathbb{E}z_i^t(r)^m$ and $\mathbb{E}\tilde{z}_i^t(r)^m$ by the assumptions on the matrices W and $\widetilde{W}$. Thus, defining the set

$$\check{\mathcal{A}}_i(\mathfrak{m}) = \Big\{(T_1, \ldots, T_m) \in \mathcal{A}_i(\mathfrak{m})\ :\ \exists j < l,\ \sum_{k\in[m]} \phi(T_k)_{jl} \geqslant 3\Big\},$$

the proposition will be proven if we prove that for all $\mathfrak{m} \in [C_E]$, the real number

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

satisfies $|\xi_{\mathfrak{m}}| \leqslant CK^{-1/2}$.

Using the bounds (17) and (18) provided in the statement of Proposition 8, it is obvious that $\prod_{k=1}^m \Gamma(T_k, \boldsymbol{\alpha}, t)$ and $\prod_{k=1}^m x(T_k)$ are bounded.

Since $|\mathbb{E}W_{jl}^s| \leqslant CK^{-s/2}$ for each integer $s > 0$ by Assumptions 1 and 2, for each $(T_1, \ldots, T_m) \in \check{\mathcal{A}}_i(\mathfrak{m})$, we have

$$\left|\mathbb{E}\prod_{k=1}^m W(T_k)\right| = \prod_{j<l}\left|\mathbb{E}W_{jl}^{\sum_k \phi(T_k)_{jl}}\right| \leqslant CK^{-\mathfrak{m}/2}.$$

To complete the proof, we shall show that

$$\left|\check{\mathcal{A}}_i(\mathfrak{m})\right| \leqslant CK^{(\mathfrak{m}-1)/2}. \tag{25}$$

From a m–uple $(T_1, \ldots, T_m) \in \check{\mathcal{A}}_i(\mathfrak{m})$, let us construct a graph $G = \boldsymbol{G}(T_1, \ldots, T_m)$ as follows. The graph G is the rooted, undirected, labelled, and unmarked graph obtained by merging the m trees $T_1, \ldots, T_m$ and by identifying the nodes that have the same type; This common type will be the label of the obtained node in G. The root node of G will be the node obtained by merging the roots of the trees $T_1, \ldots, T_m$ (remember that they all have the same type i). The other nodes are numbered, say, in the increasing order of their labels. The edges of G are furthermore unweighted.

The number of edges of G is

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

Remember that when $\sum_k \phi(T_k)_{jl} > 0$, this sum is $\geqslant 2$, and for some $j < l$, it is $\geqslant 3$. Consequently, $2|E(G)| + 1 \leqslant \sum_{j<l}\sum_k \phi(T_k)_{jl}$, in other words,

$$|E(G)| \leqslant \frac{\mathfrak{m}-1}{2}.$$

Note that since G is connected as being obtained through the merger of trees, $|V(G)| \leqslant |E(G)|+1$. Thus, $|\{v \in G, v \neq \circ\}| \leqslant (\mathfrak{m}-1)/2$. Also, by construction, G satisfies the following property:

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

We shall denote as $\mathcal{G}_i^{\mathfrak{m}}$ the set of rooted, undirected, labelled and connected graphs such that $\ell(\circ) = i$, $|E(G)| \leqslant \frac{\mathfrak{m}-1}{2}$, and the last property is satisfied. We denote as $\mathcal{R}^{\mathfrak{m}}$ the set of all the elements of $\mathcal{G}_i^{\mathfrak{m}}$ but without the labels. Given a graph $G \in \mathcal{G}_i^{\mathfrak{m}}$, let us denote as $\bar{G} = \boldsymbol{U}(G) \in \mathcal{R}^{\mathfrak{m}}$ the unlabelled version of G. With these notations, we have

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

For each graph G, we have

$$\left|\left\{(T_1, \ldots, T_m) \in \check{\mathcal{A}}_i(\mathfrak{m}) \ :\ \boldsymbol{G}(T_1, \ldots, T_m) = G\right\}\right| \leqslant C, \tag{27}$$

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

$$\left|\left\{G \in \mathcal{G}_i^{\mathfrak{m}} \,:\, \boldsymbol{U}(G) = \bar{G}\right\}\right| \leqslant CK^{(\mathfrak{m}-1)/2}. \tag{28}$$

Given $\bar{G} \in \mathcal{R}^{\mathfrak{m}}$, denote as $\circ$ the root node of $\bar{G}$, write $M = |V(\bar{G})| - 1 \leqslant (\mathfrak{m}-1)/2$, and write $V(\bar{G})\backslash\{\circ\} = [M]$. Recalling that $\bar{G}$ is connected, let us consider a spanning tree of this graph rooted in $\circ$. Denote as $\boldsymbol{\pi}(v)$ the parent of the node v in this tree. Writing $j_\circ = i$, we obtain that

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

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

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

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

$$\left|\left\{(j_1, \ldots, j_M) \in [n]^M \,:\, \forall k \in [M],\ j_k \in \mathcal{K}_{j_{\boldsymbol{\pi}(k)}}\right\}\right| \leqslant CK^M \leqslant CK^{(\mathfrak{m}-1)/2}.$$

Inequality (28) follows.

It is furthermore easy to check that

$$|\mathcal{R}^{\mathfrak{m}}| \leqslant C.$$

Getting back to Inequality (26), and using this last inequality along with Inequalities (28) and (27), we obtain Inequality (25), and the proposition is established. □

Let us keep the notations of the former proof. Consider the subset $\widetilde{\mathcal{A}}_i(\mathfrak{m})$ of $\mathcal{A}_i(\mathfrak{m})$ defined as

$$\widetilde{\mathcal{A}}_i(\mathfrak{m}) = \Big\{(T_1, \ldots, T_m) \in \mathcal{A}_i(\mathfrak{m}) \,:\, \forall j < l,\ \sum_{k \in [m]} \phi(T_k)_{jl} \in \{0, 2\}\Big\}.$$

For $(T_1, \ldots, T_m) \in \widetilde{\mathcal{A}}_i(\mathfrak{m})$, let us denote as $G = \boldsymbol{G}(T_1, \ldots, T_m) \in \mathcal{G}_i^{\mathfrak{m}}$ the graph obtained by merging these trees, and $\bar{G} = \boldsymbol{U}(G) \in \mathcal{R}^{\mathfrak{m}}$ the unlabelled version of G, as we did for $(T_1, \ldots, T_m) \in \check{\mathcal{A}}_i(\mathfrak{m})$. As we did in (24), $\mathbb{E} z_i^t(r)^m$ can be written as

$$\mathbb{E} z_i^t(r)^m = \sum_{\mathfrak{m}=1}^{C_E} \chi_{\mathfrak{m}} + \sum_{\mathfrak{m}=1}^{C_E} \xi_{\mathfrak{m}}, \tag{29}$$

where

$$\begin{aligned}
\chi_{\mathfrak{m}} &= \sum_{(T_1,\ldots,T_m) \in \widetilde{\mathcal{A}}_i(\mathfrak{m})} \left(\prod_{k=1}^m \Gamma(T_k, \boldsymbol{\alpha}, t)\right) \left(\prod_{k=1}^m x(T_k)\right) \mathbb{E}\left[\prod_{k=1}^m W(T_k)\right] \\
&= \sum_{\bar{G} \in \mathcal{R}^{\mathfrak{m}}} \sum_{\substack{G \in \mathcal{G}_i^{\mathfrak{m}}: \\ \boldsymbol{U}(G) = \bar{G}}} \sum_{\substack{(T_1,\ldots,T_m) \in \widetilde{\mathcal{A}}_i(\mathfrak{m}): \\ \boldsymbol{G}(T_1,\ldots,T_m) = G}} \left(\prod_{k=1}^m \Gamma(T_k, \boldsymbol{\alpha}, t)\right) \left(\prod_{k=1}^m x(T_k)\right) \mathbb{E}\left[\prod_{k=1}^m W(T_k)\right],
\end{aligned}$$

and where we recall from the former proof that $\xi_{\mathfrak{m}} = \sum_{(T_1,\ldots,T_m) \in \check{\mathcal{A}}_i(\mathfrak{m})} \cdots$ satisfies $|\xi_{\mathfrak{m}}| \leqslant CK^{-1/2}$.

Let us further decompose $\chi_{\mathfrak{m}}$ as

$$\chi_{\mathfrak{m}} = \sum_{\substack{\bar{G}\in\mathcal{R}^{\mathfrak{m}}:\\ \bar{G}\text{ is a tree}}} \sum_{\substack{G\in\mathcal{G}_i^{\mathfrak{m}}:\\ \boldsymbol{U}(G)=\bar{G}}} \sum_{\substack{(T_1,\ldots,T_m)\in\widetilde{\mathcal{A}}_i(\mathfrak{m}):\\ \boldsymbol{G}(T_1,\ldots,T_m)=G}} \cdots + \sum_{\substack{\bar{G}\in\mathcal{R}^{\mathfrak{m}}:\\ \bar{G}\text{ not a tree}}} \sum_{\substack{G\in\mathcal{G}_i^{\mathfrak{m}}:\\ \boldsymbol{U}(G)=\bar{G}}} \sum_{\substack{(T_1,\ldots,T_m)\in\widetilde{\mathcal{A}}_i(\mathfrak{m}):\\ \boldsymbol{G}(T_1,\ldots,T_m)=G}} \cdots$$
$$= \chi_{\mathfrak{m}}^{\mathrm{T}} + \chi_{\mathfrak{m}}^{\mathrm{NT}}.$$

Then, the contribution of the term $\chi_{\mathfrak{m}}^{\mathrm{NT}}$ is negligible:

Lemma 11 (see Lemma 2 of [BLM15])**.** $\left|\chi_{\mathfrak{m}}^{\mathrm{T}}\right| \leqslant C$, and $\left|\chi_{\mathfrak{m}}^{\mathrm{NT}}\right| \leqslant C/K$.

Proof. By repeating the argument of the former proof, the terms $\prod_{k=1}^m \Gamma(T_k, \boldsymbol{\alpha}, t)$ and $\prod_{k=1}^m x(T_k)$ are both bounded, and the term $\mathbb{E}\prod_{k=1}^m W(T_k)$ accounts for a factor of order $K^{-\mathfrak{m}/2}$ in both $\chi_{\mathfrak{m}}^{\mathrm{T}}$ and $\chi_{\mathfrak{m}}^{\mathrm{NT}}$.

Furthermore, when $(T_1,\ldots,T_m) \in \widetilde{\mathcal{A}}_i(\mathfrak{m})$, the graph $G = \boldsymbol{G}(T_1,\ldots,T_m)$ satisfies $|E(G)| = \mathfrak{m}/2$. We further know that $\bar{G} = \boldsymbol{U}(G)$ satisfies $|V(\bar{G})| \leqslant |E(\bar{G})| + 1 = |E(G)| + 1$, with equality if and only if G is a tree. Therefore, when $\bar{G} \in \mathcal{R}^{\mathfrak{m}}$ is a tree, $|V(\bar{G})| = \mathfrak{m}/2 + 1$. Once we set $M = |V(G)| - 1 = \mathfrak{m}/2$, the argument for establishing Inequality (28) in the former proof can be reproduced word for word here to show that $|\{G \in \mathcal{G}_i^{\mathfrak{m}} : \boldsymbol{U}(G) = \bar{G}\}| \leqslant CK^{\mathfrak{m}/2}$. When $\bar{G} \in \mathcal{R}^{\mathfrak{m}}$ is not a tree, $|V(\bar{G})| \leqslant \mathfrak{m}/2$, and we obtain that $|\{G \in \mathcal{G}_i^{\mathfrak{m}} : \boldsymbol{U}(G) = \bar{G}\}| \leqslant CK^{\mathfrak{m}/2-1}$. Thus,

$$\left|\chi_{\mathfrak{m}}^{\mathrm{T}}\right| \leqslant CK^{-\mathfrak{m}/2}K^{\mathfrak{m}/2} = C,$$
$$\left|\chi_{\mathfrak{m}}^{\mathrm{NT}}\right| \leqslant CK^{-\mathfrak{m}/2}K^{\mathfrak{m}/2-1} = CK^{-1},$$

and the lemma is proven. □

Getting back to the expression (29) of $\mathbb{E}z_i^t(r)^m$ and using this lemma along with the bound $|\xi_{\mathfrak{m}}| \leqslant CK^{-1/2}$, we obtain that for each $t > 0$ and each multi-index $\boldsymbol{m} \in \mathbb{N}^q$, there exists a constant $C > 0$ such that

$$\max_{i\in[n]} \mathbb{E}|(\boldsymbol{z}_i^t)^{\boldsymbol{m}}| \leqslant C. \tag{30}$$

This bound will be needed below.

Recall that the samples $\boldsymbol{y}_i^t$ are obtained by drawing an independent matrix W^t at each iteration in the parameter t. We have:

Proposition 12 (Proposition 2 of [BLM15])**.** For each $t \geqslant 1$ and each $\boldsymbol{m} \in \mathbb{N}^q$, there exists C such that for each $i \in [n]$,

$$\left|\mathbb{E}(\boldsymbol{z}_i^t)^{\boldsymbol{m}} - \mathbb{E}(\boldsymbol{y}_i^t)^{\boldsymbol{m}}\right| \leqslant CK^{-1/2}.$$

Proof. Parallelling the quantities $W(T)$ introduced above for a tree $T \in \overline{\mathcal{T}}^t$, [BLM15] introduced the quantities

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

By an easy adaptation of the proof of [BLM15, Lm. 1] (Lemma 9 above), we can show that

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

Similarly to the proof of Proposition 10, we assume that $m(r) = m$ and $m(s) = 0$ for $s \in [q]\backslash\{r\}$. Similarly to Equation (24), and with the same notations, it holds that

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

As we did for $\mathbb{E}z_i^t(r)^m$ above, we partition $\mathcal{A}_i(\mathfrak{m})$ as $\mathcal{A}_i(\mathfrak{m}) = \widetilde{\mathcal{A}}_i(\mathfrak{m}) \cup \check{\mathcal{A}}_i(\mathfrak{m})$. We prove with the same arguments that the contribution of $\check{\mathcal{A}}_i(\mathfrak{m})$ is of order $K^{-1/2}$. Furthermore, parallelling Lemma 11, we also obtain that within $\widetilde{\mathcal{A}}_i(\mathfrak{m})$ we can limit ourselves to the terms $\bar{\chi}_{\mathfrak{m}}^{\mathrm{T}}$ defined as

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

the terms for which $\bar{G}$ is not a tree being of order K^{-1}. With this at hand, the proposition will be established once we show that $\bar{\chi}_{\mathfrak{m}}^{\mathrm{T}} = \chi_{\mathfrak{m}}^{\mathrm{T}}$, where we recall that $\chi_{\mathfrak{m}}^{\mathrm{T}}$, introduced in the last proof, has the same expression as $\bar{\chi}_{\mathfrak{m}}^{\mathrm{T}}$ except that the terms $\overline{W}(T_k)$ in the latter are replaced with $W(T_k)$.

Consider an arbitrary m–uple $(T_1, \ldots, T_m)$ in the inner sum above. We first notice that if $\mathbb{E}\prod_{k=1}^m \overline{W}(T_k) \neq 0$, then $\mathbb{E}\prod_{k=1}^m \overline{W}(T_k) = \mathbb{E}\prod_{k=1}^m W(T_k)$. This is due to the fact that if $j < l$ is active in $\mathbb{E}\prod_{k=1}^m \overline{W}(T_k)$ (in the sense that $\sum_k \phi(T_k)_{jl} \neq 0$, and thus, is equal to 2), then the corresponding contribution of this $j < l$ to $\mathbb{E}\prod_{k=1}^m W(T_k)$ will be exactly the same.

The proof will then be terminated if we show that if $\mathbb{E}\prod_{k=1}^m W(T_k) \neq 0$ and $\mathbb{E}\prod_{k=1}^m \overline{W}(T_k) = 0$, then necessarily, the graph $G = \boldsymbol{U}(T_1, \ldots, T_m)$ will not be a tree, *i.e.*, it will contain a cycle. Assume that $\mathbb{E}\prod_{k=1}^m W(T_k) \neq 0$ and $\mathbb{E}\prod_{k=1}^m \overline{W}(T_k) = 0$. Then, there will be an edge $\{u, v\}$ in G, with $\{\ell(u), \ell(v)\} = \{j, l\}$, but that will appear at two different distances to $\circ$ in the trees T_k. Let us consider the three possible cases where this could happen:

1. This happens in the same tree, say T_1, and on the same path to $\circ$. Then, due to the non-backtracking nature of T_1, a cycle appears in G.

2. This happens in two different trees, say T_1 and T_2. Namely, there exists two edges $u \to v \in E(T_1)$ and $u' \to v' \in E(T_2)$ such that $\{\ell(u), \ell(v)\} = \{\ell(u'), \ell(v')\}$, and $|u| \neq |u'|$. Then, keeping in mind the backtracking property, it is easy to observe that a cycle is created in G.

3. A similar remark can be made when this happens in the same tree but on two different paths to the root.

Thus, we have a contradiction in the three cases, and we get that $\bar{\chi}_{\mathfrak{m}}^{\mathrm{T}} = \chi_{\mathfrak{m}}^{\mathrm{T}}$. The proposition is proven. □

The following proposition links the joint moments of the elements of the vectors $\boldsymbol{z}_i^t$ with those of the vectors $\boldsymbol{x}_i^t$ provided by the AMP algorithm (16).

Proposition 13 (proposition 3 of [BLM15])**.** For each $t \geqslant 1$ and each $\boldsymbol{m} \in \mathbb{N}^q$, there exists C such that for each $i \in [n]$,

$$\left|\mathbb{E}(\boldsymbol{z}_i^t)^{\boldsymbol{m}} - \mathbb{E}(\boldsymbol{x}_i^t)^{\boldsymbol{m}}\right| \leqslant CK^{-1/2}.$$

Recalling the bound (30), the bound (20) can be deduced from this proposition.

To prove this proposition, new objects need to be introduced. In a directed and labelled graph,

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

Let $\overline{\mathcal{U}}^t$ be the set of equivalence classes of trees in $\overline{\mathcal{T}}^t$ from which the marks have been removed. Denote as $\mathcal{B}^t$ the set of trees T in $\overline{\mathcal{U}}^t$ that satisfy the following additional conditions:

- If $u \to v \in E(T)$, then $\ell(u) \neq \ell(v)$.
- There exists in T at least one backtracking path of length 3 or one backtracking star.

Finally, $\mathcal{B}_i^t$ is the subset of trees in $\mathcal{B}^t$ with the root having the type i, and such that the type of the child v of the root satisfies $\ell(v) \neq i$.

The proof of Proposition 13 relies on the following structural lemma:

Lemma 14 (Lemma 3 of [BLM15])**.**

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

where $|\widetilde{\Gamma}(T,t,r)| \leqslant C(d,t)$.

Proof of Proposition 13. Once again, we assume that $m(r) = m$ and $m(s) = 0$ for $s \in [q]\backslash\{r\}$. We write

$$\mathbb{E}(x_i^t(r))^m = \mathbb{E}\Big(z_i^t(r) + \sum_{T \in \mathcal{B}_i^t} W(T)\widetilde{\Gamma}(T,t,r)x(T)\Big)^m$$

where $z_i^t(r)$ is given by Lemma 9. With this at hand, the proposition will be established once we bound the terms of the type

$$\sum_{T_1 \in \mathcal{B}_i^t} \sum_{T_2,\ldots,T_m \in \mathcal{B}_i^t \cup \mathcal{T}_i^t(r)} \Big|\mathbb{E}\prod_{k=1}^m W(T_k)\Big|.$$

The argument is nearly the same as in the proof of Proposition 10. Defining the set

$$\begin{aligned}\mathcal{D}_i(\mathfrak{m}) = \Big\{(T_1,\ldots,T_k) \ :\ & T_1 \in \mathcal{B}_i^t,\ T_2,\ldots,T_m \in \mathcal{B}_i^t \cup \mathcal{T}_i^t(r),\\ & \forall j < l,\ \sum_{k\in[m]} \phi(T_k)_{jl} \neq 1,\\ & \forall j < l,\ \sum_{k\in[m]} \phi(T_k)_{jl} > 0\ \Rightarrow \{j,l\} \in \mathcal{K},\\ & \sum_{k\in[m]}\sum_{j<l} \phi(T_k)_{jl} = \mathfrak{m}\Big\}.\end{aligned}$$

Recalling the notations of the proof of Proposition 10, we need to show that

$$\delta_{\mathfrak{m}} = \sum_{\bar{G}\in\mathcal{R}^{\mathfrak{m}}} \sum_{\substack{G\in\mathcal{G}_i^{\mathfrak{m}}:\\ \boldsymbol{U}(G)=\bar{G}}} \sum_{\substack{(T_1,\ldots,T_m)\in\mathcal{D}_i(\mathfrak{m}):\\ \boldsymbol{G}(T_1,\ldots,T_m)=G}} \Big|\mathbb{E}\prod_{k=1}^m W(T_k)\Big|$$

satisfies $\delta_{\mathfrak{m}} \leqslant CK^{-1/2}$.

As usual, $|\mathbb{E}\prod_{k=1}^m W(T_k)| \leqslant CK^{-\mathfrak{m}/2}$. We need to bound $|\mathcal{D}_i(\mathfrak{m})|$. To this end, we observe that since $T_1 \in \mathcal{B}_i^t$, resulting in this tree having a backtracking path or a backtracking star, it is easy to see that the graph $\bar{G}$ has an edge that results from the fusion of three edges at least. This implies that $|E(\bar{G})| \leqslant (\mathfrak{m}-1)/2$. Reusing the argument of the proof of Proposition 10, we obtain that $|\mathcal{D}_i(\mathfrak{m})| \leqslant CK^{(\mathfrak{m}-1)/2}$, which shows that $\delta_{\mathfrak{m}} \leqslant CK^{-1/2}$ as required. Proposition 13 is proven. □

Making use of the independence of the matrices W^t, we now show that the joint moments of the elements of a sample $\boldsymbol{y}_i^t$ are close to their analogues for U_i^t, which distribution is provided before Proposition 8. It will be enough to consider that the matrices W^t are Gaussian.

Proposition 15. Assume that the matrix W is Gaussian. Then for each multi-index $\boldsymbol{m} \in \mathbb{N}^q$, each integer $t > 0$,

$$\max_{i\in[n]} \big|\mathbb{E}(\boldsymbol{y}_i^t)^{\boldsymbol{m}} - \mathbb{E}(U_i^t)^{\boldsymbol{m}}\big| \xrightarrow[n\to\infty]{} 0.$$

Proof. The uniform convergence we need to show can be equivalently stated as follows: for each sequence (i_n) valued in $[n]$, it holds that

$$\mathbb{E}(\boldsymbol{y}_{i_n}^t)^{\boldsymbol{m}} - \mathbb{E}(U_{i_n}^t)^{\boldsymbol{m}} \xrightarrow[n\to\infty]{} 0. \tag{31}$$

Remember that $\boldsymbol{y}^0_{i\to j} = \boldsymbol{x}^0_i$ and $\boldsymbol{y}^{t+1}_{i\to j} = \sum_{l\in[n]\setminus\{j\}} W^t_{il} f(\boldsymbol{y}^t_{l\to i}, l, t)$ for each $i \neq j$. First, using this equation, it is easy to establish by recurrence on t that

$$\forall t \geqslant 0,\ \forall \boldsymbol{m} \in \mathbb{N}^q,\ \sup_n \max_{i\neq j} \mathbb{E}|(\boldsymbol{y}^t_{i\to j})^{\boldsymbol{m}}| < \infty. \tag{32}$$

Fixing t, we shall show by recurrence on $u = 1, \ldots, t-1$ the following assertion that we denote as $\mathcal{A}(u)$: For each multi-index $\boldsymbol{m}$, each sequence (j_n) valued in $[n]$, and each $(n-1)$–uple $(b_\ell)_{\ell\in[n]\setminus\{j_n\}}$ with bounded elements, it holds that

$$\sum_{\ell\in[n]\setminus\{j\}} s_{i\ell} b_\ell (\boldsymbol{y}^u_{\ell\to j})^{\boldsymbol{m}} - s_{i\ell} b_\ell \mathbb{E}(U^u_\ell)^{\boldsymbol{m}} \xrightarrow[n\to\infty]{\mathcal{P}} 0,$$

where $j = j_n$. In all the proof, we shall need the covariance matrices H^u_{ij} defined for $u \geqslant 1$ and $i \neq j$ as

$$H^u_{ij} = \sum_{l\in[n]\setminus\{j\}} s_{il} f(\boldsymbol{y}^{u-1}_{l\to i}, l, u-1) f(\boldsymbol{y}^{u-1}_{l\to i}, l, u-1)^\top.$$

Starting with $\mathcal{A}(1)$, let us assume for notational simplicity and without generality loss that $j = n$. We have

$$\begin{aligned}
\sum_{l\in[n-1]} s_{il} b_l \left((\boldsymbol{y}^1_{l\to n})^{\boldsymbol{m}} - \mathbb{E}(U^1_l)^{\boldsymbol{m}} \right) &= \sum_{l\in[n-1]} s_{il} b_l \left((\boldsymbol{y}^1_{l\to n})^{\boldsymbol{m}} - \mathbb{E}(\boldsymbol{y}^1_{l\to n})^{\boldsymbol{m}} \right) \\
&\quad + \sum_{l\in[n-1]} s_{il} b_l \left(\mathbb{E}(\boldsymbol{y}^1_{l\to n})^{\boldsymbol{m}} - \mathbb{E}(U^1_l)^{\boldsymbol{m}} \right) \\
&= \chi_1 + \chi_2.
\end{aligned}$$

It is obvious that $\boldsymbol{y}^1_{l\to n} \sim \mathcal{N}(0, H^1_{ln})$, where $H^1_{ln} = Q^1_l - E_{l,n}$, and the rank-one matrix

$$E_{l,n} = s_{l,n} f(\boldsymbol{x}^0_n, n, 0) f(\boldsymbol{x}^0_n, n, 0)^\top$$

has a spectral norm that converges to zero by Assumptions 2 and 3. It is then easy to deduce that $\chi_2 \to_n 0$.

To deal with χ_1, we make use of Poincaré's inequality [PS11, Ch. 2]. For $u \geqslant 1$, let Σ^u_n be the the $q(n-1) \times q(n-1)$ covariance matrix defined as $\Sigma^u_n = \left[\Sigma^u_n(k,l)\right]^{n-1}_{k,\ell=1}$ where the $q \times q$ block $\Sigma^u_n(k,l)$ is given as

$$\Sigma^u_n(k,l) = \begin{cases} H^u_{k,n} \text{ if } k = l \\ s_{kl} f(\boldsymbol{y}^{u-1}_{l\to k}, l, u-1) f(\boldsymbol{y}^{u-1}_{k\to l}, k, u-1)^\top \text{ if not.} \end{cases}$$

Note that Σ^1_n is deterministic, and the $\mathbb{R}^{q(n-1)}$–valued vector $\boldsymbol{y}^1_{\cdot\to n} = \left[\boldsymbol{y}^1_{i\to n}\right]^{n-1}_{i=1}$ has the distribution $\mathcal{N}(0, \Sigma^1_n)$. Define the function $\Gamma(\boldsymbol{y}^1_{\cdot\to n}) = \sum_{l\in[n-1]} s_{il} b_l (\boldsymbol{y}^1_{l\to n})^{\boldsymbol{m}}$, and write $\nabla_{\boldsymbol{y}} \boldsymbol{y}^{\boldsymbol{m}} = p_{\boldsymbol{m}}(\boldsymbol{y})$, a $\mathbb{R}^q$–valued polynomial. Then, we obtain by Poincaré's inequality

$$\mathbb{E}\chi_1^2 = \mathbb{V}\mathrm{ar}\left(\Gamma(\boldsymbol{y}^1_{\cdot\to n})\right) \leqslant \sum_{k,l=1}^{n-1} s_{ik} s_{il} b_k b_l \mathbb{E} p_{\boldsymbol{m}}(\boldsymbol{y}^1_{k\to n})^\top \Sigma^1_n(k,l) p_{\boldsymbol{m}}(\boldsymbol{y}^1_{l\to n-1}).$$

Considering the expression of $\Sigma^1_n(k,l)$, and using Assumption 3 and the bound (32), the right hand side of the previous display satisfies the bounds

$$\left|\sum_{k=l} \cdots \right| \leqslant C \sum_{k=1}^{n-1} s^2_{ik} \leqslant \frac{C}{K}, \quad \text{and} \quad \left|\sum_{k\neq l} \cdots \right| \leqslant C \sum_{k,\ell=1}^{n-1} s_{ik} s_{il} s_{kl} \leqslant \frac{C}{K}.$$

It results that $\chi_1 \xrightarrow{\mathcal{P}}_n 0$, and $\mathcal{A}(1)$ is established.

Assuming that $\mathcal{A}(u)$ is true, let us establish $\mathcal{A}(u+1)$. Define the σ–field $\mathscr{F}^u = \sigma(W^0, \ldots, W^u)$. Then, still setting $j = n$, the conditional distribution $\mathscr{L}(\boldsymbol{y}^{u+1}_{\cdot\to n} \,|\, \mathscr{F}^{u-1})$ of the vector $\boldsymbol{y}^{u+1}_{\cdot\to n} = [\boldsymbol{y}^{u+1}_{i\to n}]_{i=1}^{n-1}$ given $\mathscr{F}^{u-1}$ is $\mathcal{N}(0, \Sigma^{u+1}_n)$. We also have from $\mathcal{A}(u)$ that

$$\forall (i_n) \text{ valued in } [n],\ H^{u+1}_{i_n n} - Q^{u+1}_{i_n} \xrightarrow[n\to\infty]{\mathcal{P}} 0. \tag{33}$$

With this at hand, we write

$$\begin{aligned}\sum_{l\in[n-1]} s_{il} b_l \left((\boldsymbol{y}^{u+1}_{l\to n})^{\boldsymbol{m}} - \mathbb{E}(U^{u+1}_l)^{\boldsymbol{m}}\right) &= \sum_{l\in[n-1]} s_{il} b_l \left((\boldsymbol{y}^{u+1}_{l\to n})^{\boldsymbol{m}} - \mathbb{E}\left[(\boldsymbol{y}^{u+1}_{l\to n})^{\boldsymbol{m}} \,|\, \mathscr{F}^{u-1}\right]\right) \\ &\quad + \sum_{l\in[n-1]} s_{il} b_l \left(\mathbb{E}\left[(\boldsymbol{y}^{u+1}_{l\to n})^{\boldsymbol{m}} \,|\, \mathscr{F}^{u-1}\right] - \mathbb{E}(U^{u+1}_l)^{\boldsymbol{m}}\right) \\ &= \chi_1 + \chi_2.\end{aligned}$$

Given a small $\delta > 0$, we write

$$\begin{aligned}|\chi_2| &\leqslant \sum_{l\in[n-1]} s_{il}|b_l| \sup_{H\geqslant 0\,:\,\|H-Q^{u+1}_l\|\leqslant\delta} \left|\varphi(H) - \varphi(Q^{u+1}_l)\right| \\ &\quad + \sum_{l\in[n-1]} s_{il}|b_l|\, |\varphi(H^{u+1}_{ln})| \mathbb{1}_{\|H^{u+1}_{ln} - Q^{u+1}_l\|>\delta} + \sum_{l\in[n-1]} s_{il}|b_l|\, |\varphi(Q^{u+1}_l)| \mathbb{1}_{\|H^{u+1}_{ln} - Q^{u+1}_l\|>\delta} \\ &= \chi_{2,1} + \chi_{2,2} + \chi_{2,3},\end{aligned}$$

where $\varphi(H) = \mathbb{E}Y^{\boldsymbol{m}}$ when $Y \sim \mathcal{N}(0, H)$. Observe that $\sum_\ell s_{il}|b_l|$ is bounded. Furthermore, assuming that H and Q belong to a compact, it holds that $\varphi(H) - \varphi(Q) \to 0$ when $\|H - Q\| \to 0$ by the continuity of φ. Thus, using the bound (19), we obtain that $\chi_{2,1} \to 0$ as $\delta \to 0$. We also have by the Jensen and the Cauchy-Schwarz inequalities that

$$\mathbb{E}\chi_{2,2} \leqslant \sum_{l\in[n-1]} s_{il}|b_l|\, (\mathbb{E}|(\boldsymbol{y}^{u+1}_{l\to n})^{\boldsymbol{m}}|^2)^{1/2} \mathbb{P}[\|H^{u+1}_{ln} - Q^{u+1}_l\| > \delta]^{1/2}.$$

The convergence (33) can be rewritten as

$$\forall\delta > 0, \max_{l\in[n]} \mathbb{P}[\|H^{u+1}_{ln} - Q^{u+1}_l\| > \delta] \xrightarrow[n\to\infty]{} 0.$$

Using the bound (32), we obtain that $\mathbb{E}\chi_{2,2} \to_n 0$, thus, $\chi_{2,2} \xrightarrow{\mathcal{P}}_n 0$. It is easy to show that $\chi_{2,3} \xrightarrow{\mathcal{P}}_n 0$. In conclusion, $\chi_2 \xrightarrow{\mathcal{P}}_n 0$.

To deal with χ_1, we use Poincaré's inequality involving this time the conditional distribution $\mathscr{L}(\boldsymbol{y}^{u+1}_{\cdot\to n} \,|\, \mathscr{F}^{u-1})$. By an argument similar to above, this leads to

$$\mathbb{E}\chi_1^2 = \mathbb{E}\,\mathbb{V}\mathrm{ar}\left(\Gamma(\boldsymbol{y}^{u+1}_{\cdot\to n}) \,|\, \mathscr{F}^{u-1}\right) \leqslant \sum_{k,l=1}^{n-1} s_{ik} s_{il} b_k b_l \mathbb{E}\left[p_{\boldsymbol{m}}(\boldsymbol{y}^{u+1}_{k\to n})^\top \Sigma^{u+1}_n(k,l) p_{\boldsymbol{m}}(\boldsymbol{y}^{u+1}_{l\to n})\right] \leqslant \frac{C}{K}$$

with the help of Inequality (32). It results that $\chi_1 \xrightarrow{\mathcal{P}}_n 0$, and $\mathcal{A}(u+1)$ is established.

We now use $\mathcal{A}(t-1)$ to prove the convergence stated by our proposition. Recall that $\boldsymbol{y}^t_i = \sum_{l\in[n]} W^{t-1}_{il} f(\boldsymbol{y}^{t-1}_{l\to i}, \ell, t-1)$. Set $i = i_n$ as in (31). Given $A > 0$, define the real function $\eta_A : \mathbb{R} \to \mathbb{R}$ as the function that coincides with the identity on $[-A, A]$ and is equal to A on (A, ∞) and to $-A$ on $(-\infty, -A)$. To study $\mathbb{E}(\boldsymbol{y}^t_i)^{\boldsymbol{m}} - \mathbb{E}(U^t_i)^{\boldsymbol{m}}$, we can assume that $\boldsymbol{y}^t_i = (H^t_i)^{1/2}\xi$ and $U^t_i = (Q^t_i)^{1/2}\xi$, where

$$H^t_i = \sum_{l\in[n]} s_{il} f(\boldsymbol{y}^{t-1}_{l\to i}, l, t-1) f(\boldsymbol{y}^{t-1}_{l\to i}, l, t-1)^\top,$$

$\xi \sim \mathcal{N}(0, I_q)$ is independent of H^t_i, and $(\cdot)^{1/2}$ is the semidefinite positive square root. Write $Y = (\boldsymbol{y}^t_i)^{\boldsymbol{m}}$ and $U = (U^t_i)^{\boldsymbol{m}}$. With this, we have

$$\begin{aligned}\mathbb{E}Y - \mathbb{E}U &= (\mathbb{E}\eta_A(Y) - \mathbb{E}\eta_A(U)) + (\mathbb{E}Y - \mathbb{E}\eta_A(Y)) + (\mathbb{E}U - \mathbb{E}\eta_A(U)) \\ &= \chi_1 + \chi_2 + \chi_3.\end{aligned}$$

For $\delta > 0$, we have

$$\begin{aligned}|\chi_1| &\leqslant \sup_{H \geqslant 0 \,:\, \|H - Q_i^t\| \leqslant \delta} \left| \mathbb{E}_\xi \eta_A(H^{1/2}\xi)^{\boldsymbol{m}}) - \mathbb{E}_\xi \eta_A(((Q_i^t)^{1/2}\xi)^{\boldsymbol{m}}) \right| + 2A\mathbb{P}\left[\|H_i^t - Q_i^t\| > \delta\right] \\ &\leqslant C_1(A, \delta) + C_2(A, \delta, n),\end{aligned}$$

where $C_1(A, \delta) \to 0$ as $\delta \to 0$, and $C_2(A, \delta, n) \to 0$ as $n \to \infty$ as in (33). Regarding χ_2, we have

$$\chi_2 = \mathbb{E}(Y - \eta_A(Y))\mathbb{1}_{|Y|>A} \leqslant \mathbb{E}(|Y| + A)\mathbb{1}_{|Y|>A} \leqslant \sup_n \frac{\mathbb{E}Y^2 + \mathbb{E}|Y|^3}{A^2} \leqslant \frac{C}{A^2}$$

for some $C > 0$. We have a similar bound for χ_3. By taking A large enough then δ small enough, we easily obtain that $\mathbb{E}(\boldsymbol{y}_i^t)^{\boldsymbol{m}} - \mathbb{E}(U_i^t)^{\boldsymbol{m}} \to_n 0$. □

The results of Propositions 10, 12, 13, and 15 will lead to the convergences (21), which will be the consequences of the two following propositions.

Proposition 16. Let $\psi^{(n)} : \mathbb{R}^q \times [n] \to \mathbb{R}$ be such that $\psi^{(n)}(\cdot, l)$ is a multivariate polynomial with a bounded degree and bounded coefficients as functions of (l, n). Then, for each set $\mathcal{S}^{(n)} \subset [n]$ with $|\mathcal{S}^{(n)}| \to_n \infty$, it holds that

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

Proof. By Propositions 10, 12 and 13, we obtain that for each $t \geqslant 1$ and each $\boldsymbol{m} \in \mathbb{N}^q$, there exists $C > 0$ such that $|\mathbb{E}(\boldsymbol{x}_i^t)^{\boldsymbol{m}} - \mathbb{E}(\boldsymbol{y}_i^t)^{\boldsymbol{m}}| \leqslant CK^{-1/2}$ for each $i \in [n]$, and furthermore, W can be assumed Gaussian in the construction of the $\boldsymbol{y}_i^t$. Using Proposition 15, we obtain that $\max_{i\in[n]} |\mathbb{E}(\boldsymbol{x}_i^t)^{\boldsymbol{m}} - \mathbb{E}(U_i^t)^{\boldsymbol{m}}| \to_n 0$.

Furthermore, using the moment bound (20) and observing that the mixed moments of the U_i^t are bounded by (19), we obtain the result. □

Proposition 17 (adaptation of Proposition 5 of [BLM15]). Let $\psi^{(n)} : \mathbb{R}^q \times [n] \to \mathbb{R}$ be as in the previous proposition. Let $\mathcal{S}^{(n)} \subset [n]$ be a non empty set such that $|\mathcal{S}^{(n)}| \leqslant CK^{(n)}$. Then, for each $t > 0$,

$$\begin{aligned}&\mathbb{V}\mathrm{ar}\Big(\frac{1}{K} \sum_{i\in\mathcal{S}} \psi(\boldsymbol{x}_i^t, i)\Big) \xrightarrow[n\to\infty]{} 0, \quad \text{and} \\ &\mathbb{V}\mathrm{ar}\Big(\frac{1}{n} \sum_{i\in[n]} \psi(\boldsymbol{x}_i^t, i)\Big) \xrightarrow[n\to\infty]{} 0.\end{aligned} \tag{34}$$

Proof. We adapt the proof of [BLM15, Prop. 5] to our situation. Let $\{\check{X}_{ij}\}_{1\leqslant i<j\leqslant n}$ be a set of real independent random variables that satisfy the same assumptions as $\{X_{ij}\}_{1\leqslant i<j\leqslant n}$. Assume furthermore that these two sets are independent. Write $\check{X}_{ji} = \check{X}_{ij}$ for $1 \leqslant i < j \leqslant n$, and let $\check{X}_{ii} = 0$ for $i \in [n]$. Define the $n \times n$ matrix $\check{X} = \left[\check{X}_{ij}\right]_{i,j=1}^n$.

Let $B = [b_{ij}]$ be the $n \times n$ symmetric matrix which first row is defined as $b_{1j} = \mathbb{1}_{j\in\mathcal{S}}$ for $j \in [n]$, and which have zeros outside its first row and first column. To establish the first convergence, we build the matrix $\breve{W} = \left[\breve{W}_{ij}\right]_{i,j=1}^n$ defined as

$$\breve{W}_{ij} = \frac{1}{\sqrt{K}} B \odot \check{X}$$

(here we assume without affecting the conclusion of the proposition that $1 \notin \mathcal{S}$). We construct a new AMP sequence around the $2n \times 2n$ matrix $\begin{bmatrix} W & 0 \\ 0 & \breve{W} \end{bmatrix}$, which obviously satisfies Assumptions 1 and 2. The new algorithm, which delivers the $2n$-uple $(\breve{\boldsymbol{y}}_1^k, \ldots, \breve{\boldsymbol{y}}_{2n}^k)$ at Iteration k, is written as follows: $\breve{\boldsymbol{y}}_i^0 = \boldsymbol{x}_i^0$ for $i \in [n]$ and $\breve{\boldsymbol{y}}_i^0 = 0$ otherwise,

$$\begin{bmatrix} (\breve{\boldsymbol{y}}_1^{k+1})^\top \\ \vdots \\ (\breve{\boldsymbol{y}}_n^{k+1})^\top \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} W & 0 \\ 0 & \breve{W} \end{bmatrix} \begin{bmatrix} f(\breve{\boldsymbol{y}}_1^k, 1, k)^\top \\ \vdots \\ f(\breve{\boldsymbol{y}}_n^k, n, k)^\top \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} f(\breve{\boldsymbol{y}}_1^{k-1}, 1, k-1)^\top \\ \vdots \\ f(\breve{\boldsymbol{y}}_n^{k-1}, n, k-1)^\top \\ 0 \\ \vdots \\ 0 \end{bmatrix} \left[\sum_l W_{i,l}^2 \frac{\partial f_r}{\partial y(s)}(\breve{\boldsymbol{y}}_l^k, l, k) \right]_{s,r=1}^q,$$

for $k = 0, \ldots, t-1$, and

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ (\breve{\boldsymbol{y}}_{n+1}^{t+1})^\top \\ \vdots \\ (\breve{\boldsymbol{y}}_{2n}^{t+1})^\top \end{bmatrix} = \begin{bmatrix} W & 0 \\ 0 & \breve{W} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ [\psi(\breve{\boldsymbol{y}}_1^t, 1), 0, \ldots, 0] \\ \vdots \\ [\psi(\breve{\boldsymbol{y}}_n^t, n), 0, \ldots, 0] \end{bmatrix}$$

(next iterations are irrelevant). This sequence enters the framework of Proposition 8.

It is clear that $\breve{\boldsymbol{y}}_i^k = \boldsymbol{x}_i^k$ for $k \in [t]$ and $i \in [n]$, and therefore,

$$\breve{\boldsymbol{y}}_{n+i}^{t+1}(1) = \sum_{l \in [n]} \breve{W}_{il} \psi(\boldsymbol{x}_l^t, l) \quad \text{for } i \in [n].$$

Set $i = 1$. On the one hand, we have

$$\begin{aligned} \mathbb{E} \breve{\boldsymbol{y}}_{n+1}^{t+1}(1)^4 &= \sum_{l_1 l_2 l_3 l_4 \in \mathcal{S}} \mathbb{E} \breve{W}_{1l_1} \breve{W}_{1l_2} \breve{W}_{1l_3} \breve{W}_{1l_4} \mathbb{E} \psi(\boldsymbol{x}_{l_1}^t, l_1) \psi(\boldsymbol{x}_{l_2}^t, l_2) \psi(\boldsymbol{x}_{l_3}^t, l_3) \psi(\boldsymbol{x}_{l_4}^t, l_4) \\ &= \frac{3}{K^2} \sum_{l_1, l_2 \in \mathcal{S}} \mathbb{E} \psi(\boldsymbol{x}_{l_1}^t, l_1)^2 \psi(\boldsymbol{x}_{l_2}^t, l_2)^2 + \varepsilon \end{aligned} \tag{35}$$

where $|\varepsilon| \leqslant C/K$. On the other hand, Propositions 10, 12, 13 and 15 applied to our new AMP sequence show that

$$\mathbb{E} \breve{\boldsymbol{y}}_{n+1}^{t+1}(1)^4 - \mathbb{E}(\widetilde{U}^{t+1})^4 \xrightarrow[n \to \infty]{} 0,$$

where $\widetilde{U}^{t+1} \sim \mathcal{N}(0, (\sigma^{t+1})^2)$, with

$$(\sigma^{t+1})^2 = \frac{1}{K} \sum_{l \in \mathcal{S}} \mathbb{E} \psi(U_l^t, l)^2,$$

By the Gaussianity and centeredness of $\widetilde{U}^{t+1}$, we thus have

$$\mathbb{E} \breve{\boldsymbol{y}}_{n+1}^{t+1}(1)^4 - \frac{3}{K^2} \Big(\mathbb{E} \sum_{l \in \mathcal{S}} \psi(U_l^t, l)^2 \Big)^2 \xrightarrow[n \to \infty]{} 0,$$

and since $\max_{i \in [n]} |\mathbb{E}(\boldsymbol{x}_i^t)^{\boldsymbol{m}} - \mathbb{E}(U_i^t)^{\boldsymbol{m}}| \to_n 0$, we get that

$$\mathbb{E} \breve{\boldsymbol{y}}_{n+1}^{t+1}(1)^4 - \frac{3}{K^2} \Big(\mathbb{E} \sum_{l \in \mathcal{S}} \psi(\boldsymbol{x}_l^t, l)^2 \Big)^2 \xrightarrow[n \to \infty]{} 0,$$

Combining this convergence with (35), we obtain the first convergence stated by our proposition for polynomials of the type $\psi(\cdot, l)^2$. To obtain this convergence for arbitrary polynomials, write $\Psi(\boldsymbol{x}, l) = (1 + \varepsilon \psi(\boldsymbol{x}, l))$ for $\varepsilon > 0$. Since

$$\frac{1}{\varepsilon^2} \mathbb{V}\mathrm{ar} \Big(\frac{1}{K} \sum_{i \in \mathcal{S}} \Psi(\boldsymbol{x}_i^t, i)^2 \Big) = \mathbb{V}\mathrm{ar} \Big(\frac{\varepsilon}{K} \sum_{i \in \mathcal{S}} \psi(\boldsymbol{x}_i^t, i)^2 + \frac{2}{K} \sum_{i \in \mathcal{S}} \psi(\boldsymbol{x}_i^t, i) \Big)$$

must vanish for all $\varepsilon > 0$, we get the convergence (34).

To establish the other convergence in the statement, let $\mathcal{S}_i^{(n)} = \{i+1, ((i+1) \bmod n) + 1, \ldots, ((i+K_n-1) \bmod n)+1\}$ for $i \in \{0, \ldots, n-1\}$. Then, we have that

$$\forall l \in [n], \quad \psi(\boldsymbol{x}_l^t, l) = \frac{1}{K} \sum_{i=0}^{n-1} \psi(\boldsymbol{x}_l^t, l) \mathbb{1}_{l \in S_i}.$$

Therefore, writing

$$\frac{1}{n} \sum_{l \in [n]} \psi(\boldsymbol{x}_l^t, l) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{K} \sum_{l \in S_i} \psi(\boldsymbol{x}_l^t, l),$$

we can use Minkowski's inequality along with the convergence (34) to show that the variance of the left hand side converges to zero. □

The convergences (21) follow at once from Propositions 16 and 17. Proposition 8 is proven.

4.1.3 Proof of Proposition 7

To establish Proposition 7 for a given fixed $t > 0$, we apply Proposition 8 with $q = t$ and a properly designed sequence of functions $f(\cdot, \cdot, 0), \ldots, f(\cdot, \cdot, t-1)$, along the idea used in [BLM15, Proof of Th. 5].

For $k = 0, \ldots, t-1$ define the function $f(\boldsymbol{x}, i, k)$ of the statement of Proposition 8 as follows. Consider the initial vector x^0 in Algorithm (12) as a constant parameter vector. For $\boldsymbol{x} = [x(1), \ldots, x(t)]$, set

$$\begin{aligned}
f(\boldsymbol{x}, i, 0)^\top &= [\; p(x_i^0, i, 0) \quad 0 \quad \ldots \quad 0 \;] \\
f(\boldsymbol{x}, i, 1)^\top &= [\; p(x_i^0, i, 0) \quad p(x(1), i, 1) \quad \ldots \quad 0 \;] \\
&\vdots \\
f(\boldsymbol{x}, i, t-1)^\top &= [\; p(x_i^0, i, 0) \quad p(x(1), i, 1) \quad \ldots \quad p(x(t-1), i, t-1) \;].
\end{aligned}$$

(note that $f(\cdot, i, 0)$ is a polynomial with degree zero). With this construction, if we start Algorithm (12) with the initial value x^0 and Algorithm (16) with an arbitrary initial value, then we can easily show by recurrence on the first t iterations running in parallel for both algorithms that

$$\begin{aligned}
(\boldsymbol{x}_i^1)^\top &= [\; \check{x}_i^1 \quad 0 \quad \ldots \quad 0 \;] \\
(\boldsymbol{x}_i^2)^\top &= [\; \check{x}_i^1 \quad \check{x}_i^2 \quad \ldots \quad 0 \;] \\
&\vdots \\
(\boldsymbol{x}_i^t)^\top &= [\; \check{x}_i^1 \quad \check{x}_i^2 \quad \ldots \quad \check{x}_i^t \;].
\end{aligned}$$

We also notice that

$$\forall i \in [n], \quad Q_i^t = \check{R}_i^t.$$

With this at hand, Proposition 7 follows from Proposition 8.

In order to deduce Theorems 2 and 4 from Proposition 7, we now need to approximate the activation functions present in the statements of these theorems with polynomials. The next subsection is devoted to this purpose.

4.2 From polynomial to general activation functions

In all this subsection, $h : \mathbb{R} \times \mathcal{Q}_\eta \times \mathbb{N} \to \mathbb{R}$ is a function that complies with either Assumption 5 or Assumption 6.

The proof of the following lemma makes use of the density of the polynomials in the Hilbert space $L^2(\mathbb{R}, \nu)$ when ν is a Gaussian measure. Density arguments of this kind have been used in the AMP literature in, *e.g.*, [DLS22, WZF22].

Lemma 18. Fix t, and fix two positive numbers $0 < \sigma^2_{\min} \leqslant \sigma^2_{\max}$. Let $\varepsilon > 0$ be an arbitrarily small number. Then, there exists a function $g_\varepsilon(\cdot,\cdot,t) : \mathbb{R} \times \mathcal{Q}_\eta \to \mathbb{R}$ that satisfies the following properties: For each $\eta \in \mathcal{Q}_\eta$, the function $g_\varepsilon(\cdot,\eta,t)$ is a polynomial. Denoting as ∂g_ε the derivative of g_ε with respect to the first parameter, the inequalities

$$\mathbb{E}(h(\underline{\xi},\eta,t) - g_\varepsilon(\underline{\xi},\eta,t))^2 \leqslant \varepsilon, \quad \text{and} \quad |\mathbb{E}[\partial h(\underline{\xi},\eta,t) - \partial g_\varepsilon(\underline{\xi},\eta,t)]| \leqslant \varepsilon$$

hold true for each random variable $\underline{\xi} \sim \mathcal{N}(0,\sigma^2)$ with $\sigma^2 \in [\sigma^2_{\min}, \sigma^2_{\max}]$ and each $\eta \in \mathcal{Q}_\eta$. Finally, the function $p^{(n)}(x,i,t) = g_\varepsilon(x,\eta_i^{(n)},t)$ satisfies the assumptions of Proposition 7.

Proof. Let $\varepsilon' > 0$ be a small number to be set at the end of the proof. Let κ be the function from Assumption 5 or 6. If Assumption 5 is chosen, set $m = 1$, otherwise, let m be the integer specified in Assumption 6. Define $\delta > 0$ as

$$\delta = \max\Big\{ e \in (0, D_{\mathcal{Q}_\eta}], : \kappa(e)^2 \leqslant \frac{\varepsilon'}{\mathbb{E}(1+|\xi|^m)^2} \Big\},$$

where $\xi \sim \mathcal{N}(0,\sigma^2_{\max})$, and where $D_{\mathcal{Q}_\eta}$ is the diameter of $\mathcal{Q}_\eta$. Since $\mathcal{Q}_\eta$ is compact, it contains a δ–net with finite cardinality. Let $\mathcal{S} \subset \mathcal{Q}_\eta$ be a δ–net of $\mathcal{Q}_\eta$ with the smallest cardinality, and write $M = |\mathcal{S}|$. Let $\phi : \mathcal{Q}_\eta \to \mathcal{S}$ be such that $\phi(\eta)$ is the closest element in $\mathcal{S}$ to η if this closest element is unique, and $\phi(\eta)$ is the closest element smaller than η if not. Denoting as $\boldsymbol{\eta}_1 < \boldsymbol{\eta}_2 < \cdots < \boldsymbol{\eta}_M$ the elements of $\mathcal{S}$, define the function $\psi : \mathcal{Q}_\eta \to [M]$ as $\phi(\eta) = \boldsymbol{\eta}_{\psi(\eta)}$. With these definitions, we have $\mathbb{E}(h(\xi,\eta,t) - h(\xi,\phi(\eta),t))^2 \leqslant \varepsilon'$.

It is now well-known that the polynomials are dense in the space $L^2(\mathbb{R},\nu)$ when ν is a Gaussian distribution. Therefore, there are M polynomials $P(\cdot,l,t)$ such that $\mathbb{E}(h(\xi,\boldsymbol{\eta}_l,t) - P(\xi,l,t))^2 \leqslant \varepsilon'$ for each $l \in [M]$. Fixing l, let $\varphi(x) = (h(x,\boldsymbol{\eta}_l,t) - P(x,l,t))^2$. For $\underline{\xi} \sim \mathcal{N}(0,\sigma^2)$ with $\sigma^2 \in [\sigma^2_{\min}, \sigma^2_{\max}]$, we have

$$\mathbb{E}(h(\underline{\xi},\boldsymbol{\eta}_l,t) - P(\underline{\xi},l,t))^2 = \frac{1}{\sqrt{2\pi\sigma^2}} \int \varphi(x) e^{-\frac{x^2}{2\sigma^2}} dx \leqslant \frac{\sigma_{\max}}{\sigma_{\min}} \frac{1}{\sqrt{2\pi\sigma^2_{\max}}} \int \varphi(x) e^{-\frac{x^2}{2\sigma^2_{\max}}} dx = \frac{\sigma_{\max}}{\sigma_{\min}} \varepsilon'.$$

Putting things together, we obtain that

$$\begin{aligned}\mathbb{E}(h(\underline{\xi},\eta,t) - P(\underline{\xi},\psi(\eta),t))^2 &\leqslant 2\mathbb{E}(h(\underline{\xi},\eta,t) - h(\underline{\xi},\phi(\eta),t))^2 + 2\mathbb{E}(h(\underline{\xi},\phi(\eta),t) - P(\underline{\xi},\psi(\eta),t))^2 \\ &\leqslant 2(1 + \sigma_{\max}/\sigma_{\min})\varepsilon'.\end{aligned}$$

We also have by Stein's lemma that

$$\mathbb{E}[\partial h(\underline{\xi},\eta,t) - \partial_{\underline{\xi}} P(\underline{\xi},\psi(\eta),t)] = \frac{1}{\sigma^2} \mathbb{E}[\underline{\xi}(h(\underline{\xi},\eta,t) - P(\underline{\xi},\psi(\eta),t))].$$

Thus, we obtain by Cauchy-Schwarz and the previous result that

$$|\mathbb{E}[\partial h(\underline{\xi},\eta,t) - \partial_{\underline{\xi}} P(\underline{\xi},\psi(\eta),t)]| \leqslant \frac{\sigma_{\max}}{\sigma^2_{\min}} \sqrt{2(1+\sigma_{\max}/\sigma_{\min})}\sqrt{\varepsilon'},$$

By adjusting ε', we thus obtain that

$$\mathbb{E}(h(\underline{\xi},\eta,t) - P(\underline{\xi},\psi(\eta),t))^2 \leqslant \varepsilon, \quad \text{and} \quad |\mathbb{E}[\partial h(\underline{\xi},\eta,t) - \partial_{\underline{\xi}} P(\underline{\xi},\psi(\eta),t)]| \leqslant \varepsilon,$$

and it remains to set $g_\varepsilon(x,\eta,t) = P(x,\psi(\eta),t)$. We also notice that the polynomials $p^{(n)}(x,i,t) = g_\varepsilon(x,\eta_i^{(n)},t)$ have bounded degrees and bounded coefficients, and thus comply with the statement of Proposition 7. □

Lemma 19. There exists a constant $c > 0$ such that $R_i(t,t) \geqslant c$ for all $i \in [n]$ and $t \in [t_{\max}]$. Let $e > 0$ be a small number. Then, there exists a set of $\mathbb{R} \times \mathcal{Q}_\eta \to \mathbb{R}$ functions $\{g_e(\cdot,\cdot,t)\}_{t=0}^{t_{\max}-1}$ such that for each $\eta \in \mathcal{Q}_\eta$, the function $g_e(\cdot,\eta,t)$ is a polynomial, and furthermore,

$$\forall i \in [n], \quad |h(x_i^0,\eta_i,0) - g_e(x_i^0,\eta_i,0)| \leqslant e,$$

and $\forall t \in [t_{\max} - 1]$, $\forall i \in [n]$,

$$\mathbb{E}(h(Z_i^t, \eta_i, t) - g_e(Z_i^t, \eta_i, t))^2 \leqslant e, \quad |\mathbb{E}[\partial h(Z_i^t, \eta_i, t) - \partial g_e(Z_i^t, \eta_i, t)]| \leqslant e,$$

and the functions $p(x, i, t) = g_e(x, \eta_i, t)$ defined for $0 \leqslant t \leqslant t_{\max} - 1$ satisfy the assumptions of Proposition 7.

Furthermore, there exists a non-negative function $\delta(e)$ on a small interval $(0, \varepsilon)$ that converges to zero as $e \to 0$, and that satisfies the following property. Given $e \in (0, \varepsilon)$, construct the function g_e as above, and let $(\breve{\boldsymbol{Z}}^t = [\breve{Z}_i^t]_{i=1}^n)_{t \in [t_{\max}]}$ be the finite sequence of centered Gaussian vectors which distribution is determined by the (S, g_e, η, x^0)–state evolution equations stopped at $t_{\max}$, and leading to the covariance matrices $\{\breve{R}_i^{t_{\max}}\}_{i \in [n]}$. Then, $\|R_i^{t_{\max}} - \breve{R}_i^{t_{\max}}\| \leqslant \delta(e)$ for all $i \in [n]$. In particular,

$$\sup_n \max_{i \in [n]} \|\breve{R}_i^{t_{\max}}\| < \infty. \tag{36}$$

Proof. Recall the notations introduced in Assumptions 2 and 7, and recall that $\mathcal{K}_i = \{j \in [n],\ s_{ij} > 0\}$. We first show that for each $i \in [n]$, the set $\mathcal{I}_i \subset \mathcal{K}_i$ defined as $\mathcal{I}_i = \{j \in [n],\ s_{ij} > c_S/(2C_{\text{card}}K)\}$ satisfies $|\mathcal{I}_i| > \alpha_S K$. Using Assumptions 2 and 7, we indeed have

$$c_S \leqslant \sum_{j \in \mathcal{K}_i} s_{ij} = \sum_{j \in \mathcal{I}_i} s_{ij} + \sum_{j \in \mathcal{K}_i \setminus \mathcal{I}_i} s_{ij} < \frac{C_S}{K} |\mathcal{I}_i| + \frac{c_S}{2C_{\text{card}}K} (C_{\text{card}}K - |\mathcal{I}_i|),$$

and the result is obtained by rearranging.

Using Assumption 7 again, we then have $R_i(1,1) = \sum_l s_{il} h(x_l^0, \eta_l, 0)^2 \geqslant c > 0$ for each $i \in [n]$, by lower bounding the sum to $l \in \mathcal{I}_i$. Assuming that $R_i(t,t) \geqslant c > 0$ for each $i \in [n]$, we also have that $R_i(t+1, t+1) = \sum_l s_{il} \mathbb{E} h(Z_l^t, \eta_l, t)^2 \geqslant c > 0$: This can be checked by noticing from Assumption 7 and the properties of h that there exists $c > 0$ such that $\sum_l s_{il} \mathbb{E} h(\underline{\xi}, \eta_l, t)^2 \geqslant c_{\text{h}}(t)$ when $\underline{\xi} \sim \mathcal{N}(0,1)$, and by making standard Gaussian calculations to modify the variance of $\underline{\xi}$ and the bound $c_{\text{h}}(t)$. This establishes the first result of the lemma.

With this result, the existence of the functions $\{g_e(x, \eta, t)\}_{t=1}^{t_{\max}-1}$ follows at once from Lemma 18, with the numbers $\sigma^2_{\min}$ and $\sigma^2_{\max}$ in the statement of that lemma chosen as $\sigma^2_{\min} = \min_{i,t} R_i(t,t)/2$ and $\sigma^2_{\max} = 2 \max_{i,t} R_i(t,t)$. As regards $g(x, \eta, 0)$, we can use Assumption 3 and either Assumption 5 or Assumption 6, and invoke the Stone-Weierstrass theorem to choose $g(\cdot, \cdot, 0)$ as a multivariate polynomial that satisfies $\max_{(x,\eta) \in \mathcal{Q}_x \times \mathcal{Q}_\eta} |g(x, \eta, 0) - h(x, \eta, 0)| \leqslant e$. It is readily checked that $p(x, l, 0) = g_e(x, \eta_l, 0)$ complies with the assumptions of Proposition 7.

We show the last result by recurrence on t. We first have $|R_i^1 - \breve{R}_i^1| \leqslant \sum_i s_{il}(h(x_l^0, \eta_l, 0)^2 - g_e(x_l^0, \eta_l, 0)^2) \leqslant Ce^2$. Assume that there exists a non-negative function $\delta_t(e)$ that converges to zero as $e \to 0$, and such that $\|R_i^t - \breve{R}_i^t\| \leqslant \delta_t(e)$. We have

$$\begin{aligned}
|R_i^{t+1}(t+1, 1) - \breve{R}_i^{t+1}(t+1, 1)| &\leqslant \sum_l s_{il} |\mathbb{E} h(Z_l^t, \eta_l, t) h(x_l^0, \eta_l, 0) - \mathbb{E} g_e(\breve{Z}_l^t, \eta_l, t) g_e(x_l^0, \eta_l, 0)| \\
&\leqslant \sum_l s_{il} |\mathbb{E}[h(Z_l^t, \eta_l, t) - h(\breve{Z}_l^t, \eta_l, t)]|\, |h(x_l^0, \eta_l, 0)| \\
&\quad + \sum_l s_{il} |\mathbb{E} h(\breve{Z}_l^t, \eta_l, t)|\, |h(x_l^0, \eta_l, 0) - g_e(x_l^0, \eta_l, 0)| \\
&\quad + \sum_l s_{il} |\mathbb{E}[h(\breve{Z}_l^t, \eta_l, t) - g_e(\breve{Z}_l^t, \eta_l, t)]|\, |g_e(x_l^0, \eta_l, 0)|.
\end{aligned}$$

Using Assumptions 3, 4, any of the two Assumptions 5 or 6, the bound (4), and the recurrence assumption, we obtain that the term at the second line can be bounded by some non-negative function $\delta_{t+1}(e)$ converging to zero with e. Indeed, when the triple $(\eta, \sigma^2, \breve{\sigma}^2)$ belongs to a compact set, the function $\varphi(\eta, \sigma^2, \breve{\sigma}^2) = \mathbb{E} h(\underline{\xi}, \eta, t) - \mathbb{E} h(\underline{\breve{\xi}}, \eta, t)$ with $\underline{\xi} \sim \mathcal{N}(0, \sigma^2)$ and $\underline{\breve{\xi}} \sim \mathcal{N}(0, \breve{\sigma}^2)$ is (uniformly) continuous on this compact, and vanishes on the set $\{(\eta, \sigma^2, \breve{\sigma}^2) : \sigma^2 = \breve{\sigma}^2\}$. One can readily check that the terms at the third and fourth line of the previous display can be bounded by

Ce and $C\sqrt{e}$ respectively. We thus obtain that $|R_i^{t+1}(t+1,1)-\breve{R}_i^{t+1}(t+1,1)| \leqslant \delta_{t+1}(e)$, by possibly modifying the function δ_{t+1}. A similar treatment can be applied to $R_i^{t+1}(t+1,k) - \breve{R}_i^{t+1}(t+1,k)$ for $k = 2, \dots, t+1$, leading to $\|R_i^{t+1} - \breve{R}_i^{t+1}\| \leqslant \delta_{t+1}(e)$ by possibly modifying $\delta_{t+1}(e)$ once again. This leads to the required result. □

We now use the previous lemma for $e > 0$ small to construct a polynomial AMP sequence. Starting with $\breve{x}^0 = x^0$, this sequence denoted as $(\breve{x}^t)_{t\in[t_{\max}]}$ is given as

$$\breve{x}^{t+1} = W g_e(\breve{x}^t, \eta, t) - \mathrm{diag}\Big(W^{\odot 2} \partial g_e(\breve{x}^t, \eta, t)\Big) g_e(\breve{x}^{t-1}, \eta, t-1). \tag{37}$$

Proposition 7 can be applied to this sequence, which implies in particular that the convergences (15) hold true for polynomial test functions. We now show that the polynomial test function can be replaced with a continuous function that increases at most polynomially near the infinity.

Lemma 20. Consider the polynomial AMP sequence $(\breve{x}^t)_{t\in[t_{\max}]}$ provided by Algorithm (37). Let the n–uple $(\beta_1, \dots, \beta_n)$ be as in the statements of Theorems 2 or 4. Let $\varphi : \mathcal{Q}_\eta \times \mathbb{R}^{t_{\max}} \to \mathbb{R}$ be a continuous function such that $|\varphi(\alpha, u_1, \dots, u_{t_{\max}})| \leqslant C(1 + |u_1|^m + \cdots + |u_{t_{\max}}|^m)$ for a given arbitrarily integer $m > 0$. Then,

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

Proof. Define the $\mathcal{P}(\mathbb{R}^{t_{\max}+2})$–valued random probability measure $\hat{\varrho}^{(n)}$ as

$$\hat{\varrho}^{(n)} = \frac{1}{n} \sum_{i\in[n]} \delta_{(\beta_i, \eta_i, \breve{x}_i^1, \dots, \breve{x}_i^{t_{\max}})},$$

and the deterministic measure $\boldsymbol{\varrho}^{(n)}$ valued on the same space as

$$\boldsymbol{\varrho}^{(n)} = \mathscr{L}((\beta_\theta, \eta_\theta, \breve{Z}_\theta^1, \dots, \breve{Z}_\theta^{t_{\max}})),$$

where the random variable θ is uniformly distributed on $[n]$ and is independent of $\{\breve{Z}_i^t\}_{t\in[t_{\max}], i\in[n]}$. Let $P(\eta, x^1, \dots, x^{t_{\max}})$ be a multivariate polynomial. Since the elements η_i of the vector η are in $\mathcal{Q}_\eta$, and since the β_i are bounded by assumption, the function $\psi : \mathbb{R}^{t_{\max}} \times [n] \to \mathbb{R}$, $(x^1, \dots, x^{t_{\max}}, l) \mapsto \beta_l P(\eta_l, x^1, \dots, x^{t_{\max}})$ complies with the assumptions of the statement of Proposition 7. Observing that $\int \underline{\beta} P(\eta, x^1, \dots, x^{t_{\max}})\, \boldsymbol{\varrho}^{(n)}(d\underline{\beta}, d\eta, dx^1, \dots, dx^{t_{\max}}) = n^{-1} \sum_i \beta_i \mathbb{E} P(\eta_i, \breve{Z}_i^1, \dots, \breve{Z}_i^{t_{\max}})$, we obtain from Proposition 7 that

$$\int \underline{\beta} P(\eta, x^1, \dots, x^{t_{\max}})\, d\hat{\varrho}^{(n)} - \int \underline{\beta} P(\eta, x^1, \dots, x^{t_{\max}})\, d\boldsymbol{\varrho}^{(n)} \xrightarrow[n\to\infty]{\mathcal{P}} 0. \tag{38}$$

We now show that this convergence remains true when P is replaced with the function φ of the statement. Recalling the bound (36) and the bounds on the β_i and the η_i, we know that for each sequence (n) of integers, there is a subsequence (that will still denote as (n)) and a deterministic measure $\boldsymbol{\varrho}^\infty \in \mathcal{P}(\mathbb{R}^{t_{\max}+2})$ such that $\boldsymbol{\varrho}^{(n)}$ converges narrowly to $\boldsymbol{\varrho}^\infty$. We shall show that $\hat{\varrho}^{(n)}$ converges narrowly in probability towards $\boldsymbol{\varrho}^\infty$, or, equivalently, that

$$\forall \boldsymbol{\omega} \in \mathbb{R}^{t_{\max}+2}, \quad \Phi_{\hat{\varrho}^{(n)}}(\boldsymbol{\omega}) \xrightarrow[n\to\infty]{\mathcal{P}} \Phi_{\boldsymbol{\varrho}^\infty}(\boldsymbol{\omega}), \tag{39}$$

where $\Phi_{\hat{\varrho}^{(n)}}$ (respectively $\Phi_{\boldsymbol{\varrho}^\infty}$) is the characteristic function of $\hat{\varrho}^{(n)}$ (respectively $\boldsymbol{\varrho}^\infty$). This narrow convergence, coupled with the moment convergence implied by (38) along our subsequence, leads to the result of the lemma.

Define respectively as $\hat{\varrho}_{\boldsymbol{\omega}}^{(n)}, \boldsymbol{\varrho}_{\boldsymbol{\omega}}^{(n)}, \boldsymbol{\varrho}_{\boldsymbol{\omega}}^{\infty} \in \mathcal{P}(\mathbb{R})$ the push-forward of $\hat{\varrho}^{(n)}$, $\boldsymbol{\varrho}^{(n)}$, and $\boldsymbol{\varrho}^{\infty}$ by the linear function $\langle \boldsymbol{\omega}, \cdot \rangle$. Writing $\boldsymbol{\omega} = [\omega_\beta, \omega_\eta, \boldsymbol{\omega}_x^\top]^\top$ where ω_β and ω_η are scalars, the moment generating function $\psi_{\boldsymbol{\omega}}^{(n)}(z)$ of $\boldsymbol{\varrho}_{\boldsymbol{\omega}}^{(n)}$ is given as

$$\psi_{\boldsymbol{\omega}}^{(n)}(z) = \frac{1}{n} \sum_{i=1}^{n} \exp((\omega_\beta \beta_i + \omega_\eta \eta_i) z - \boldsymbol{\omega}_x^\top \breve{R}_i^{t_{\max}} \boldsymbol{\omega}_x \, z^2/2).$$

Using the normal family theorem in conjunction with the boundedness of the β_i, the η_i and the norms $\|\breve{R}_i^{t_{\max}}\|$, we obtain that the moment generating function $\psi_{\boldsymbol{\omega}}^{\infty}(z)$ of $\boldsymbol{\varrho}_{\boldsymbol{\omega}}^{\infty}$ is holomorphic in a neighborhood of zero (even entire), thus, $\boldsymbol{\varrho}_{\boldsymbol{\omega}}^{\infty}$ is determined by its moments. Moreover, we obtain from the convergence (38) that each moment of $\hat{\varrho}_{\boldsymbol{\omega}}^{(n)}$ converges in probability to its analogue for $\boldsymbol{\varrho}_{\boldsymbol{\omega}}^{\infty}$. This implies that $\hat{\varrho}_{\boldsymbol{\omega}}^{(n)}$ converges narrowly in probability towards $\boldsymbol{\varrho}_{\boldsymbol{\omega}}^{\infty}$ for each $\boldsymbol{\omega}$, which is equivalent to the convergence (39). ∎

In Algorithm (37), the term $\mathrm{diag}(W^{\odot 2} \partial g_e(\breve{x}^t, \eta, t))$ can be approximated with the more manipulable term $\mathrm{diag}(S \partial g_e(\breve{x}^t, \eta, t))$. This will be a consequence of the next lemma.

Lemma 21. For each $t \in [t_{\max} - 1]$, there is a constant $C \geqslant 0$ such that for each $i \in [n]$,

$$\mathbb{E}\Big[\Big(\sum_{l \in [n]} (W_{il}^2 - s_{il}) \partial g_e(\breve{x}_l^t, \eta_l, t) \Big)^4 \Big] \leqslant C/K^2.$$

The proof of this lemma follows closely the development of [BLM15, Sec. A.2], with an adaptation to our variance profile model very similar to what we did in the proof of Proposition 10 above. We omit the details.

We are now in position to prove Theorems 2 and 4. In the next two subsections, the following notational conventions will be adopted. We shall most often write $g_e(x) = g_e(x, \eta, t)$ and $h(x) = h(x, \eta, t)$ for conciseness; The values of η and t will be clear from the context. Denote as $\mathcal{E}_W$ the event $\mathcal{E}_W = [\|W\| \leqslant C_W]$. Given a sequence $(\xi^{(n)})$ of non-negative random variables and a number $e > 0$, the notations $\xi^{(n)} \overset{\mathcal{P}}{\leqslant} e$ and $\xi^{(n)} \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} e$ will stand respectively for $\mathbb{P}[\xi^{(n)} \geqslant e] \to_n 0$ and $\mathbb{P}[[\xi^{(n)} \geqslant e] \cap \mathcal{E}_W] \to_n 0$. The relations $\overset{\mathcal{P}}{\leqslant}$ and $\overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W}$ satisfy some obvious calculation rules, such as $\xi_1^{(n)} + \xi_2^{(n)} \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} e_1 + e_2$ when $\xi_1^{(n)} \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} e_1$ and $\xi_2^{(n)} \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} e_2$. In both proofs, the function $\delta : (0, \epsilon) \to \mathbb{R}_+$ defined for some $\epsilon > 0$ is a generic function, independent of n, such that $\delta(e) \to 0$ as $e \to 0$. This function can change from a display to another.

4.3 Proof of Theorem 2

Given an arbitrarily small $e > 0$, let us construct the functions $g_e(\cdot, \cdot, t)$ for $t = 0, \ldots, t_{\max} - 1$ as specified by Lemma 19. With these functions at hand, let $(\breve{x}^1, \ldots, \breve{x}^{t_{\max}})$ be the iterates obtained by Algorithm (37). We shall compare the iterates of Algorithm (5) with those of the former, and show by recurrence on $t = 1, \ldots, t_{\max}$ that

$$\|\breve{x}^t - x^t\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e), \quad \text{and} \quad \|h(x^t) - g_e(\breve{x}^t)\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e).$$

Starting with $t = 1$, since $x^1 - \breve{x}^1 = W(h(x^0) - g_e(x^0))$, it holds that

$$\|x^1 - \breve{x}^1\|_n \leqslant C_W \|h(x^0) - g_e(x^0)\|_n \leqslant C_W e$$

on the event $\mathcal{E}_W$, by the construction of g_e specified by Lemma 19. Using this bound along with the Lipschitz property of $h(\cdot, \eta_i, 1)$ provided by Assumption 5, we obtain

$$\|h(x^1) - g_e(\breve{x}^1)\|_n \leqslant \|h(x^1) - h(\breve{x}^1)\|_n + \|h(\breve{x}^1) - g_e(\breve{x}^1)\|_n \leqslant C\|x^1 - \breve{x}^1\|_n + \|h(\breve{x}^1) - g_e(\breve{x}^1)\|_n.$$

Writing

$$\frac{1}{n}\sum_i (h(\breve{x}_i^1) - g_e(\breve{x}_i^1))^2 = \frac{1}{n}\sum_i \Big((h(\breve{x}_i^1) - g_e(\breve{x}_i^1))^2 - \mathbb{E}(h(\breve{Z}_i^1) - g_e(\breve{Z}_i^1))^2\Big) + \frac{1}{n}\sum_i \mathbb{E}(h(\breve{Z}_i^1) - g_e(\breve{Z}_i^1))^2,$$

the first term at the right hand side converges to zero in probability by Lemma 20, and the second term is bounded by e by construction of the function $g_e(\cdot,\cdot,1)$. This establishes the recurrence property for $t = 1$.

Let $t \in [t_{\max} - 1]$, and assume that $\|x^s - \breve{x}^s\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e)$ and $\|h(x^s) - g_e(\breve{x}^s)\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e)$ for $s = 1, \ldots, t$. To establish these bounds for $s = t + 1$, we first show that

$$\big\|\mathrm{diag}(W^{\odot 2}\partial g_e(\breve{x}^t)) g_e(\breve{x}^{t-1}) - \mathrm{diag}(S\mathbb{E}\partial h(\boldsymbol{Z}^t)) h(x^{t-1})\big\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e). \tag{40}$$

To this end, we write

$$\begin{aligned}
&\mathrm{diag}(W^{\odot 2}\partial g_e(\breve{x}^t)) g_e(\breve{x}^{t-1}) - \mathrm{diag}(S\mathbb{E}\partial h(\boldsymbol{Z}^t)) h(x^{t-1}) \\
&= \mathrm{diag}\big((W^{\odot 2} - S)\partial g_e(\breve{x}^t)\big) g_e(\breve{x}^{t-1}) + \mathrm{diag}\big(S\big(\partial g_e(\breve{x}^t) - \mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)\big)\big) g_e(\breve{x}^{t-1}) \\
&\quad + \mathrm{diag}\big(S\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)\big)\big(g_e(\breve{x}^{t-1}) - h(x^{t-1})\big) + \mathrm{diag}\big(S\big(\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t) - \mathbb{E}\partial h(\boldsymbol{Z}^t)\big)\big) h(x^{t-1}).
\end{aligned}$$

We limit ourselves to $t \geqslant 2$, and omit the easy adaptations of the proof to manage the terms $g_e(x^0)$ and $h(x^0)$ above when $t = 1$. Using Lemma (21) and the bound (14), the term χ_1 defined as

$$\chi_1 = \|\,\mathrm{diag}((W^{\odot 2} - S)\partial g_e(\breve{x}^t)) g_e(\breve{x}^{t-1})\|_n^2 = \frac{1}{n}\sum_{i \in [n]} [(W^{\odot 2} - S)\partial g_e(\breve{x}^t)]_i^2 [g_e(\breve{x}^{t-1})]_i^2$$

satisfies $\mathbb{E}\chi_1 \to 0$ by the Cauchy-Schwarz inequality, thus, $\chi_1 \xrightarrow{\mathcal{P}} 0$.

We now deal with the next term $\chi_2 = \|\,\mathrm{diag}(S(\partial g_e(\breve{x}^t) - \mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t))) g_e(\breve{x}^{t-1})\|_n^2$. For $i = i_n \in [n]$, write $\xi_i = [S(\partial g_e(\breve{x}^t) - \mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)]_i$. Applying the convergence (15a) in the statement of Proposition 7 with $\mathcal{S} = \mathcal{K}_i = \{j \in [n],\ s_{ij} > 0\}$ and $\psi(\breve{x}_j^1, \ldots, \breve{x}_j^t, j) = K s_{ij}\partial g_e(\breve{x}_j^t)$, we obtain that $\xi_{i_n} \xrightarrow{\mathcal{P}}_n 0$. Furthermore, for each integer $m > 0$, we obtain from the bound (14) and Minkowski's inequality that $\max_{j \in [n]} \mathbb{E}|\xi_j|^m$ is bounded. Let $\varepsilon > 0$. Writing $\xi_i = \xi_i \mathbb{1}_{|\xi_i| \geqslant \epsilon} + \xi_i \mathbb{1}_{|\xi_i| < \epsilon}$ and using Cauchy-Schwarz inequality along with the bound (14), we obtain that

$$\mathbb{E}\chi_2 \leqslant \frac{C}{n}\sum_{i \in [n]} \mathbb{P}[\xi_i \geqslant \varepsilon]^{1/2} + C\varepsilon^2,$$

thus, $\limsup_n \mathbb{E}\chi_2 \leqslant C\varepsilon^2$, and since ε is arbitrary, $\mathbb{E}\chi_2 \to 0$.

Considering the term $\chi_4 = \|\,\mathrm{diag}\big(S\big(\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t) - \mathbb{E}\partial h(\boldsymbol{Z}^t)\big)\big) h(x^{t-1})\|_n^2$, we have

$$\begin{aligned}
|[S\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t) - S\mathbb{E}\partial h(\boldsymbol{Z}^t)]_i| &\leqslant |[S\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t) - S\mathbb{E}\partial h(\breve{\boldsymbol{Z}}^t)]_i| + |[S\mathbb{E}\partial h(\breve{\boldsymbol{Z}}^t) - S\mathbb{E}\partial h(\boldsymbol{Z}^t)]_i| \\
&\leqslant Ce + \delta(e),
\end{aligned} \tag{41}$$

where the bound Ce on the first term comes from the construction of g_e in Lemma 19, and the bound $\delta(e)$ on the second term is due to the inequality $\max_i \|\breve{R}_i^{t_{\max}} - R_i^{t_{\max}}\| \leqslant \delta(e)$ stated by the same lemma. Writing $\|h(x^{t-1})\|_n \leqslant \|h(x^{t-1}) - h(\breve{x}^{t-1})\|_n + \|h(\breve{x}^{t-1})\|_n \leqslant C\|x^{t-1} - \breve{x}^{t-1}\| + \|h(\breve{x}^{t-1})\|_n$ and using Proposition 20 to control $\|h(\breve{x}^{t-1})\|_n$, we obtain that there exists a constant $C > 0$ such that $\mathbb{P}[[\|h(x^{t-1})\|_n \geqslant C] \cap \mathcal{E}_W] \to 0$. By consequence, $\chi_4 \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e)$.

We finally consider the term $\chi_3 = \|\,\mathrm{diag}\big(S\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)\big)\big(g_e(\breve{x}^{t-1}) - h(x^{t-1})\big)\|_n^2$. By Inequality (41) and the bound (4), the real numbers $b_i = [S\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)]_i^2$ are bounded. Using the recurrence assumption $\|g_e(\breve{x}^{t-1}) - h(x^{t-1})\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e)$, we thus obtain that $\chi_3 \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e)$.

Gathering these results on χ_1 to χ_4, we obtain the convergence expressed by (40).

With this at hand, we obtain from Equations (37) and (5) that

$$\|x^{t+1} - \check{x}^{t+1}\|_n \leqslant \|W\| \|h(x^t) - g_e(\check{x}^t)\|_n + \left\| \mathrm{diag}(W^{\odot 2} \partial g_e(\check{x}^t)) g_e(\check{x}^{t-1}) - \mathrm{diag}(S \mathbb{E} \partial h(\boldsymbol{Z}^t)) h(x^{t-1}) \right\|_n$$

which shows with the help of the recurrence assumption again that $\|x^{t+1} - \check{x}^{t+1}\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e)$. Similarly to what we did for $\|h(x^1) - g_e(\check{x}^1)\|_n$, we also obtain that $\|h(x^{t+1}) - g_e(\check{x}^{t+1})\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e)$.

We now have all the elements to show the convergence (6) provided in the statement of the theorem.

Let $\varphi \in \mathrm{PL}_2(\mathbb{R}^{t_{\max}+1})$. Write $\boldsymbol{x}_i = \left[\eta_i, x_i^1, \ldots, x_i^{t_{\max}}\right]^\top$ and $\check{\boldsymbol{x}}_i = \left[\eta_i, \check{x}_i^1, \ldots, \check{x}_i^{t_{\max}}\right]^\top$, and let $\boldsymbol{u} = \left[\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_n^\top\right]^\top$, and $\check{\boldsymbol{u}} = \left[\check{\boldsymbol{x}}_1^\top, \ldots, \check{\boldsymbol{x}}_n^\top\right]^\top$. Then, by the Cauchy-Schwarz inequality, the inequality $(a+b+c)^2 \leqslant 3(a^2+b^2+c^2)$, and Assumption 4, we have

$$\begin{aligned}
\frac{1}{n}\Big| \sum_{i \in [n]} \beta_i \varphi(\eta_i, x_i^1, \ldots, x_i^{t_{\max}}) - \beta_i \varphi(\eta_i, \check{x}_i^1, \ldots, \check{x}_i^{t_{\max}}) \Big| &\leqslant \frac{C}{n} \sum_{i \in [n]} \|\boldsymbol{x}_i - \check{\boldsymbol{x}}_i\| (1 + \|\boldsymbol{x}_i\| + \|\check{\boldsymbol{x}}_i\|) \\
&\leqslant \frac{C}{n} \|\boldsymbol{u} - \check{\boldsymbol{u}}\| \Big(\sum_{i \in [n]} (1 + \|\boldsymbol{x}_i\| + \|\check{\boldsymbol{x}}_i\|)^2 \Big)^{1/2} \\
&\leqslant C \|\boldsymbol{u} - \check{\boldsymbol{u}}\|_n (1 + \|\boldsymbol{u}\|_n + \|\check{\boldsymbol{u}}\|_n) \\
&\leqslant C \Big(\sum_{t=1}^{t_{\max}} \|x^t - \check{x}^t\|_n \Big) \Big(1 + \sum_{t=1}^{t_{\max}} \|x^t\|_n + \|\check{x}^t\|_n \Big).
\end{aligned}$$

We just showed that $\sum_{t=1}^{t_{\max}} \|x^t - \check{x}^t\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e)$. We furthermore have that

$$\forall t \in [t_{\max}], \quad \frac{1}{n} \sum_{i \in [n]} (\check{x}_i^t)^2 - \mathbb{E}(\check{Z}_i^t)^2 \xrightarrow[n \to \infty]{\mathcal{P}} 0$$

by Proposition 7, thus, $\|\check{x}^t\|_n \overset{\mathcal{P}}{\leqslant} C$ for each $t \in [t_{\max}]$. Also, $\|x^t\|_n \leqslant \|x^t - \check{x}^t\|_n + \|\check{x}^t\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} C$. It follows that

$$\frac{1}{n}\Big| \sum_{i \in [n]} \beta_i \varphi(\eta_i, x_i^1, \ldots, x_i^{t_{\max}}) - \beta_i \varphi(\eta_i, \check{x}_i^1, \ldots, \check{x}_i^{t_{\max}}) \Big| \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e).$$

Using Lemma 20 in conjunction with the bound

$$\Big| \frac{1}{n} \sum_{i \in [n]} \beta_i \mathbb{E} \varphi(\eta_i, \check{Z}_i^1, \ldots, \check{Z}_i^t) - \beta_i \mathbb{E} \varphi(\eta_i, Z_i^1, \ldots, Z_i^t) \Big| \leqslant \delta(e),$$

we obtain that

$$\frac{1}{n}\Big| \sum_{i \in [n]} \beta_i \varphi(\eta_i, x_i^1, \ldots, x_i^{t_{\max}}) - \beta_i \mathbb{E} \varphi(\eta_i, Z_i^1, \ldots, Z_i^t) \Big| \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta(e),$$

and since e is arbitrarily small, we obtain the convergence (6). Theorem 2 is proven.

4.4 Proof of Theorem 4

Here also, our starting point will be a polynomial AMP sequence. Given a small $e > 0$, construct the functions $g_e(\cdot, \cdot, t)$ for $t = 0, \ldots, t_{\max} - 1$ as specified by Lemma 19. With these functions at hand, let $(\check{x}^1, \ldots, \check{x}^{t_{\max}})$ be the iterates obtained by Algorithm (37), starting with $\check{x}^0 = x^0$.

We shall make use of this sequence in a different way than in the previous subsection. As a matter of fact, we shall not be able to construct a "true" AMP sequence (x^t) as in (5) and compare it with the sequence $(\breve{x}^t)$ recursively on t as we did in the previous section, because we have lost the Lipschitz character of $h(\cdot, \eta_i, t)$. However, by writing for each $t = 0, 1, \ldots, t_{\max} - 1$,

$$\bar{x}^{(n),t+1} = W^{(n)} h(\breve{x}^{(n),t}, \eta^{(n)}, t) - \operatorname{diag}\left(S^{(n)} \mathbb{E}\partial h(\boldsymbol{Z}^{(n),t}, \eta^{(n)}, t)\right) h(\breve{x}^{(n),t-1}, \eta^{(n)}, t-1),$$

with the term $\operatorname{diag}(\cdots)h(\cdots)$ being zero for $t = 0$, we shall be able to show that there exists a function $\delta_x(e) \geqslant 0$, defined for e small enough, such that $\delta_x(e) \to 0$ as $e \to 0$, and such that

$$\forall t \in [t_{\max}], \quad \|\breve{x}^t - \bar{x}^t\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} \delta_x(e). \tag{42}$$

This will be the first step to prove Theorem 4.

We first establish

$$\left\|\operatorname{diag}(W^{\odot 2} \partial g_e(\breve{x}^t)) g_e(\breve{x}^{t-1}) - \operatorname{diag}(S\mathbb{E}\partial h(\boldsymbol{Z}^t)) h(\breve{x}^{t-1})\right\|_n \overset{\mathcal{P}}{\leqslant} \delta(e) \tag{43}$$

for $t \in [t_{\max} - 1]$. To this end, we write

$$\begin{aligned}
&\operatorname{diag}(W^{\odot 2} \partial g_e(\breve{x}^t)) g_e(\breve{x}^{t-1}) - \operatorname{diag}(S\mathbb{E}\partial h(\boldsymbol{Z}^t)) h(\breve{x}^{t-1}) \\
&= \operatorname{diag}\big((W^{\odot 2} - S)\partial g_e(\breve{x}^t)\big) g_e(\breve{x}^{t-1}) + \operatorname{diag}\big(S\big(\partial g_e(\breve{x}^t) - \mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)\big)\big) g_e(\breve{x}^{t-1}) \\
&\quad + \operatorname{diag}\big(S\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)\big)\big(g_e(\breve{x}^{t-1}) - h(\breve{x}^{t-1})\big) + \operatorname{diag}\big(S\big(\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t) - \mathbb{E}\partial h(\boldsymbol{Z}^t)\big)\big) h(\breve{x}^{t-1}),
\end{aligned}$$

Assume for simplicity that $t \geqslant 2$ as in the previous proof. The terms $\chi_1 = \|\operatorname{diag}((W^{\odot 2} - S)\partial g_e(\breve{x}^t)) g_e(\breve{x}^{t-1})\|_n^2$ and $\chi_2 = \|\operatorname{diag}(S(\partial g_e(\breve{x}^t) - \mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t))) g_e(\breve{x}^{t-1})\|_n^2$ are identical to their analogues in the previous proof. To manage the term $\chi_3 = \|\operatorname{diag}\big(S\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)\big)\big(g_e(\breve{x}^{t-1}) - h(\breve{x}^{t-1})\big)\|_n^2$, we use Lemma 20 to obtain that

$$\left\|g_e(\breve{x}^{t-1}) - h(\breve{x}^{t-1})\right\|_n^2 - \frac{1}{n}\sum_{i \in [n]} \mathbb{E}(g_e(\breve{Z}_i^{t-1}) - h(\breve{Z}_i^{t-1}))^2 \xrightarrow[n \to \infty]{\mathcal{P}} 0.$$

Since $\mathbb{E}(g_e(\breve{Z}_i^{t-1}) - h(\breve{Z}_i^{t-1}))^2 \leqslant e$ by the construction of g_e and the $[S\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t)]_i$ are bounded, we obtain that $\chi_3 \overset{\mathcal{P}}{\leqslant} \delta(e)$. The term $\chi_4 = \|\operatorname{diag}\big(S\big(\mathbb{E}\partial g_e(\breve{\boldsymbol{Z}}^t) - \mathbb{E}\partial h(\boldsymbol{Z}^t)\big)\big) h(\breve{x}^{t-1})\|_n^2$ can be managed by a similar method. Details are omitted. This leads to (43).

On the event $\mathcal{E}_W$, we furthermore have $\|W(h(\breve{x}^t) - g_e(\breve{x}^t))\| \leqslant C_W \|h(\breve{x}^t) - g_e(\breve{x}^t)\|$ for $t = 0, \ldots, t_{\max} - 1$. By working similarly as for χ_3 above, we obtain that

$$\forall t \in [t_{\max}], \quad \left\|W(h(\breve{x}^t) - g_e(\breve{x}^t))\right\|_n \overset{\mathcal{P}}{\leqslant}_{\mathcal{E}_W} C\sqrt{e}$$

(specificities for $t = 0$ obvious). Combining this with the convergence (43), we obtain the convergences (42).

To pursue, let $\varphi : \mathcal{Q}_\eta \times \mathbb{R}^{t_{\max}} \to \mathbb{R}$ be as in the statement of Theorem 4. From Lemma 20 and from the bound

$$\left|\frac{1}{n}\sum_{i \in [n]} \beta_i \mathbb{E}\varphi(\eta_i, \breve{Z}_i^1, \ldots, \breve{Z}_i^t) - \beta_i \mathbb{E}\varphi(\eta_i, Z_i^1, \ldots, Z_i^t)\right| \leqslant \delta(e),$$

we obtain that there exists a function $\delta_\varphi(e) \geqslant 0$ converging to zero as $e \to 0$, and such that

$$\left|\frac{1}{n}\sum_{i \in [n]} \beta_i \varphi(\eta_i, \breve{x}_i^1, \ldots, \breve{x}_i^{t_{\max}}) - \beta_i \mathbb{E}\varphi(\eta_i, Z_i^1, \ldots, Z_i^{t_{\max}})\right| \overset{\mathcal{P}}{\leqslant} \delta_\varphi(e). \tag{44}$$

With the help of these results, we can now construct the sequence of matrices $(\widetilde{\boldsymbol{X}}^{(n)})$ provided in the statement of the theorem. Given a large enough integer k_0, let $k \geqslant k_0$, and choose $e > 0$

small enough so that $\delta_x(e) \vee \delta_\varphi(e) = 1/k$. By (42), there exists a random matrix ${}^k\boldsymbol{X}^{(n)} = \begin{bmatrix} {}^k\check{x}^{(n),1} & \dots & {}^k\check{x}^{(n),t_{\max}} \end{bmatrix} \in \mathbb{R}^{n \times t_{\max}}$ such that

$$\mathbb{P}\left[\left[\left\| {}^k\check{x}^{t+1} - \left(W h({}^k\check{x}^t) - \operatorname{diag}\left(S \mathbb{E}\partial h(\boldsymbol{Z}^t) \right) h({}^k\check{x}^{t-1}) \right) \right\|_n \geqslant 1/k \right] \cap \mathcal{E}_W \right] \xrightarrow[n\to\infty]{} 0$$

for each $t = 0, \dots, t_{\max} - 1$, and by (44),

$$\mathbb{P}\left[\left| \frac{1}{n} \sum_{i \in [n]} \beta_i \varphi(\eta_i, [{}^k\check{x}^1]_i, \dots, [{}^k\check{x}^{t_{\max}}]_i) - \beta_i \mathbb{E}\varphi(\eta_i, Z_i^1, \dots, Z_i^{t_{\max}}) \right| \geqslant 1/k \right] \xrightarrow[n\to\infty]{} 0,$$

where the $[{}^k\check{x}^t]_i$ are the elements of the vector ${}^k\check{x}^t$. If $k = k_0$, let $N_k \in \mathbb{N}_*$ be the smallest integer such that for all $n \geqslant N_k$, the two probabilities above are upper bounded by $1/k$. If $k > k_0$, do the same thing for N_k with the supplementary condition that $N_k > N_{k-1}$. Construct the sequence of matrices $(\widetilde{\boldsymbol{X}}^{(n)})_{n \geqslant 2}$, with $\widetilde{\boldsymbol{X}}^{(n)} = \begin{bmatrix} \tilde{x}^{(n),1} & \cdots & \tilde{x}^{(n),t_{\max}} \end{bmatrix} \in \mathbb{R}^{n \times t_{\max}}$ in such a way that $\widetilde{\boldsymbol{X}}^{(n)} = {}^k\boldsymbol{X}^{(n)}$ for $n = N_k, N_k + 1, \dots, N_{k+1} - 1$. Then, this sequence satisfies the properties provided in the statement of Theorem 4.

5 Proof of Theorem 5

As said in the introduction, $u_\star^{(n)}$ can be identified as the solution of a LCP problem [Tak96]. We recall herein the elements of the LCP theory needed in this paper. As above, we often drop the superscript ${}^{(n)}$. We shall also use the well-known notations $\succcurlyeq$, $\preccurlyeq$, and $\prec$ to refer to element-wise inequalities between vectors. Recall that the notation xy refers to the elementwise product of the $\mathbb{R}^n$–valued vectors x and y.

5.1 The equilibrium $u_\star$ as the solution of a LCP problem

Given a matrix $B \in \mathbb{R}^{n \times n}$ and a vector $b \in \mathbb{R}^n$, the LCP problem, denoted as $\mathrm{LCP}(B, b)$, consists in finding a couple of vectors $(z, y) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

$$\begin{aligned} &y = Bz + b \succcurlyeq 0, \\ &z \succcurlyeq 0, \\ &\langle z, y \rangle = 0. \end{aligned}$$

When a solution (z, y) exists and is unique, we write $z = \mathrm{LCP}(B, b)$.

Obviously, an equilibrium $\bar{u}$ of our LV dynamical system (9) is defined by the system

$$\begin{aligned} &\bar{u} \succcurlyeq 0, \\ &\bar{u}\,(r + (\Sigma - I)\,\bar{u}) = 0. \end{aligned}$$

Furthermore, the supplementary condition

$$r + (\Sigma - I)\,\bar{u} \preccurlyeq 0$$

turns out to be a necessary condition for the equilibrium $\bar{u}$ to be stable in the classical sense of Lyapounov theory (see [Tak96, Chapter 3] to recall the different notions of stability, and [Tak96, Theorem 3.2.5] for this result). These three conditions can be rewritten as

$$\begin{aligned} &\bar{w} = (I - \Sigma)\,\bar{u} - r \succcurlyeq 0, \\ &\bar{u} \succcurlyeq 0, \\ &\langle \bar{w}, \bar{u} \rangle = 0, \end{aligned}$$

in other words, the couple $(\bar{u}, \bar{w})$ solves the problem $\mathrm{LCP}(I - \Sigma, -r)$.

Theorem 5 is more specific, since it asserts that the ODE (9) has a (unique) *globally stable* equilibrium when $\|\Sigma\| < 1$. Recalling that Σ is a symmetric matrix, this can be obtained from the following result:

Proposition 22 (Lemma 3.2.2 and Theorem 3.2.1 of [Tak96])**.** Given a symmetric matrix $B \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$, consider the LV ODE

$$\dot{u}(t) = u(t)\,(b + Bu(t))\,, \quad t \geqslant 0,\ u(0) \in \mathbb{R}^n_{*+}.$$

Then, the LCP problem $\mathrm{LCP}(-B, -b)$ has an unique solution for each $b \in \mathbb{R}^N$ if and only if B is negative definite (notation $B < 0$). On the domain where $B < 0, b \in \mathbb{R}^n$, the function $x = \mathrm{LCP}(-B, -b)$ is measurable. Moreover, if $B < 0$, then for each $b \in \mathbb{R}^N$, the ODE above has a globally stable equilibrium u given as $u = \mathrm{LCP}(-B, -b)$.

Considering our LV ODE (9), the first part of Theorem 5 is true by setting

$$u_\star^{(n)} = \begin{cases} \mathrm{LCP}(I_n - \Sigma^{(n)}, -r^{(n)}) & \text{if } \|\Sigma^{(n)}\| < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

5.2 Behavior of $\mu^{u_\star}$ by an AMP approach: proof outline

In the remainder, given $x, y, z, v \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$, expressions such as xAy or $xAyzv$ always denote $\mathbb{R}^n$–valued vectors with the bracketing always starting from the right. Thus, $xAy = x(Ay)$ and $xAyzv = x(A(yzv))$.

Let us outline our approach for studying the asymptotic behavior of the measure $\mu^{u_\star^{(n)}}$.

Let $S^{(n)} \in \mathbb{R}_+^{n\times n}$ and $\eta^{(n)} \in \mathbb{R}_+^n$ be respectively a deterministic symmetric matrix and a deterministic vector that comply with Assumptions 2 and 4 respectively. These objects will be specified below. For the moment, we just assume that

$$\limsup_n \|S^{(n)}\| < 1. \tag{46}$$

Given the matrix S, let W be random symmetric $n \times n$ matrix given by Equation (1). In this equation, the matrix X is precisely the one used to construct the interaction matrix Σ of our LV model. Regarding the function $h : \mathbb{R} \times \mathcal{Q}_\eta \times \mathbb{N} \to \mathbb{R}$ of Section 2, set $h(x, \eta, t) = (x + \eta)_+$. Observe that this function is uniformly Lipschitz in x and satisfies Assumption 5. With a small notational modification, we shall consider that h is a $\mathbb{R} \times \mathcal{Q}_\eta \to \mathbb{R}$ function, and write

$$h(x, \eta) = (x + \eta)_+. \tag{47}$$

Let us consider a Gaussian sequence $(\boldsymbol{Z}^t = [Z_i^t]_{i=1}^n)_t$ which distribution is determined by the $(S, h, \eta, 1_n)$–state evolution equations. In particular, writing $a_i^t = \mathbb{E}(Z_i^t)^2$, the vector of variances $a^t = [a_i^t]_{i=1}^n$ satisfies the recursion

$$a^1 = S\,(1 + \eta)^2\,, \quad \text{and} \quad a^{t+1} = S\mathbb{E}\left(\sqrt{a^t}\xi + \eta\right)_+^2, \tag{48}$$

with $\xi \sim \mathcal{N}(0, I_n)$.

Notice that $\mathbb{E}\partial h(Z^t, \eta) = \mathbb{P}[Z^t + \eta \geqslant 0] = \mathbb{P}[\sqrt{a^t}\xi + \eta \geqslant 0]$ with $\partial h(x, \eta) = \partial h(\cdot, \eta)|_x$. Writing $\zeta^t = S\mathbb{P}[\sqrt{a^t}\xi + \eta \geqslant 0]$, we can apply Theorem 2 to the AMP sequence in t

$$x^0 = 1, \quad x^{t+1} = W(x^t + \eta)_+ - \zeta^t(x^{t-1} + \eta)_+. \tag{49}$$

Thus, given a random variable θ uniformly distributed on $[n]$ and independent of $\boldsymbol{Z}^t$, the distribution μ^{x^t} can be approximated by $\mathscr{L}(Z_\theta^t)$ in the sense that $\boldsymbol{d}_2(\mu^{x^t}, \mathscr{L}(Z_\theta^t)) \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$.

In all what follows, when we say that a sequence $(y^t)_{t=1,2,\ldots}$ of $\mathbb{R}^n$–valued vectors converge to the vector y as $t \to \infty$, we always consider that this convergence occurs in the max-norm $\|\cdot\|_\infty$ and is uniform in n.

Using the condition $\limsup_n |||S||| < 1$, it is not difficult to show that the iterates a^t converge as $t \to \infty$ to the vector a defined as the unique solution of the identity

$$a = S\mathbb{E}\left(\sqrt{a}\xi + \eta\right)_+^2 . \tag{50}$$

On the one hand, this implies that for large values of t, we can replace the vector $\boldsymbol{Z}^t$ with a vector $\boldsymbol{Z} = [Z_i]_{i=1}^n \stackrel{\mathcal{L}}{=} \sqrt{a}\xi$, which implies that μ^{x^t} can be approximated by $\mathscr{L}(Z_\theta)$, where the uniformly distributed random variable θ on $[n]$ is independent of $\boldsymbol{Z}$. On the other hand, ζ^t converges in t to the vector ζ given as

$$\zeta = S\mathbb{P}[\sqrt{a}\xi + \eta \geqslant 0]. \tag{51}$$

We can thus write

$$x^{t+1} = W(x^t + \eta)_+ - \zeta(x^{t-1} + \eta)_+ + \chi_1,$$

where $\chi_1 = (\zeta - \zeta^t)(x^{t-1} + \eta)_+$ is such that $\|\chi_1\|_n$ is small for large t.

Next, we need to show that the vectors x^t and x^{t+1} tend to become aligned as t grows. To that end, building on a result of [MR16, DM16], we show that the correlation vector q^t, defined as

$$q^t = \frac{\mathbb{E}\boldsymbol{Z}^t\boldsymbol{Z}^{t-1}}{\sqrt{a^t a^{t-1}}},$$

converges to the vector 1_n as $t \to \infty$. With these elements, the AMP iteration can be written as

$$x^t = W(x^t + \eta)_+ - \zeta(x^t + \eta)_+ + \chi_1 + \chi_2,$$

where

$$\chi_2 = \zeta\left[(x^t + \eta)_+ - (x^{t-1} + \eta)_+\right] + x^t - x^{t+1}$$

has the property that $\|\chi_2\|_n$ is small for large values of t. The next to last equation can be rewritten

$$-(x^t + \eta)_- = W(x^t + \eta)_+ - (1 + \zeta)(x^t + \eta)_+ + \eta + \chi_1 + \chi_2,$$

which leads to the equivalent equation

$$\begin{aligned} -(1+\zeta)^{-1/2}(x^t + \eta)_- &= (1+\zeta)^{-1/2}W(1+\zeta)^{-1/2}(1+\zeta)^{1/2}(x^t+\eta)_+ - (1+\zeta)^{1/2}(x^t+\eta)_+ \\ &\quad + (1+\zeta)^{-1/2}\eta + \boldsymbol{\varepsilon}^t, \end{aligned} \tag{52}$$

with $\boldsymbol{\varepsilon}^t = (1+\zeta)^{-1/2}(\chi_1 + \chi_2)$.

We know that when $\|\Sigma\| < 1$, the equilibrium vector of the ODE (9) is $u_\star = \mathrm{LCP}(I - \Sigma, -r)$. On the other hand, the previous equation shows that the vector $u^t = (1+\zeta)^{1/2}(x^t + \eta)_+$ is a solution to the LCP problem $\mathrm{LCP}(I - \mathrm{diag}(1+\zeta)^{-1/2}W\,\mathrm{diag}(1+\zeta)^{-1/2}, -(1+\zeta)^{-1/2}\eta - \boldsymbol{\varepsilon}^t)$. Thus, if we choose the matrix W as $W = \mathrm{diag}(1+\zeta)^{1/2}\Sigma\,\mathrm{diag}(1+\zeta)^{1/2}$, or, equivalently, if we put

$$S = \mathrm{diag}(1+\zeta)V\,\mathrm{diag}(1+\zeta), \tag{53}$$

and furthermore, if we set

$$\eta = (1+\zeta)^{1/2}r, \tag{54}$$

then we obtain that

$$u^t = \mathrm{LCP}(I - \Sigma, -r - \boldsymbol{\varepsilon}^t).$$

Thus, we constructed with the help of an AMP approach a "perturbed" version of $u_\star$ which empirical distribution can be evaluated with arbitrary precision for large n, and the perturbation $\boldsymbol{\varepsilon}^t$ can be made "arbitrarily small" in the norm $\|\cdot\|_n$ for large t.

Recalling Equations (50) and (51), the choices made in (53) and (54) result in the following system of $2n$ equations in the unknown vectors a and ζ:

$$a = (1+\zeta)V(1+\zeta)\mathbb{E}\left(\sqrt{a}\xi + (1+\zeta)^{1/2}r\right)_+^2$$
$$\zeta = (1+\zeta)V(1+\zeta)\mathbb{P}\left[\sqrt{a}\xi + (1+\zeta)^{1/2}r \geqslant 0\right].$$

It is a bit more convenient to write

$$p = \frac{a}{1+\zeta}, \tag{55}$$

resulting in the equivalent system (11). Remembering that $\mu^{x^t} \simeq \mathscr{L}(Z_\theta)$ and observing that $u^t = (1+\zeta)^{1/2}(x^t + (1+\zeta)^{1/2}r)_+$, we obtain from what precedes that $\mu^{u_\star} \simeq \mathscr{L}((Y_\theta)_+)$ where Y_θ is the random variable specified in the statement of Theorem 5.

To prove this theorem rigorously, we thus need to perform the following steps:

1. Prove that under Hypothesis 3, the system (11) admits an unique solution $(p, \zeta) \in \mathbb{R}_+^n \times [0,1]^n$. This is the content of Lemma 23 below. With this solution, construct the matrix S and the vector η as in (53) and (54) respectively. Note that the bound (46) is satisfied thanks to Hypothesis 3.

2. With the help of Theorem 2, establish Equation (52) with a control on the error term ε^t.

3. We just saw that u^t is a perturbed version of $u_\star$. To be able to approximate $\mu^{u_\star}$ with μ^{u^t}, we need a LCP perturbation result. This will be provided by the work of Chen and Xiang in [CX07], which will let us control the norm $\|u^t - u_\star\|$.

5.3 Behavior of $\mu^{u_\star}$: Detailed proof

Lemma 23. The system (11) admits an unique solution $(p^{(n)}, \zeta^{(n)}) \in \mathbb{R}_+^n \times [0,1]$, and this solution satisfies $\sup_n \|p^{(n)}\|_\infty < \infty$.

This lemma will be proven in Section 5.4 below.

We now consider the second step of the proof. Given the solution (p, ζ) of System (11), let $S \in \mathbb{R}_+^{n\times n}$ and $\eta \in \mathbb{R}_{*+}^n$ be given by equations (53) and (54) respectively. Note that the bound (46) is satisfied thanks to Hypothesis 3. Define the random matrix $W = S^{\odot 1/2} \odot X$. Define the function h as in (47), and let $x^0 = 1_n$.

We these definitions, let us check that the assumptions leading to Theorem 2 are satisfied. Using Hypothesis 1, it is clear that our matrix W satisfies Assumptions 1 and 2. Trivially, $x^0 = 1_n$ satisfies Assumption 3. Letting $r_{\min}, r_{\max} > 0$ be respectively the minimum and maximum values of the elements of the compact $\mathcal{Q}_r$ of Hypothesis 4, we obtain from Equation (54) that the elements of η belong to the compact interval $\mathcal{Q}_\eta = [\eta_{\min}, \eta_{\max}] \subset \mathbb{R}_{*+}$, where $\eta_{\min} = r_{\min}$ and $\eta_{\max} = \sqrt{2}r_{\max}$, and Assumption 4 is satisfied. Finally, one can readily check that the function h and the matrix S satisfy Assumptions 5 and 7.

Lemma 24. As $t \to \infty$, the sequence (a^t) converges to the vector $a \in \mathbb{R}_{*+}^n$ given as the unique solution of Equation (50), the sequence (ζ^t) converges to ζ, and the sequence (q^t) of correlation vectors converges to 1_n.

Proof. Let $\underline{\xi} \sim \mathcal{N}(0,1)$, and define the function $f : \mathbb{R}_+ \times \mathcal{Q}_\eta \to \mathbb{R}_+, (\underline{a}, \eta) \mapsto \mathbb{E}(\sqrt{\underline{a}}\underline{\xi} + \eta)_+^2$. A small calculation that we omit reveals that the derivative $\partial f(\underline{a}, \eta) = \partial f(\cdot, \eta)|_{\underline{a}}$ on $\mathbb{R}_{*+}$ is given as $\partial f(\underline{a}, \eta) = \mathbb{P}[\sqrt{\underline{a}}\underline{\xi} + \eta \geqslant 0]$. In particular, $f(\cdot, \eta)$ is increasing. Therefore, $f(\underline{a}, \eta) \geqslant f(0, \eta) = \eta^2 \geqslant \eta_{\min}$. We thus have from Assumption 7 and Equations (48) that

$$a^1 \geqslant c_S 1 \quad \text{and} \quad \forall t \geqslant 1,\ a^{t+1} \geqslant S\eta^2 \geqslant c_S \eta_{\min}^2 1.$$

Moreover, since $\|S\| < 1$, and since $f(\underline{a}, \cdot)$ is also increasing, we have

$$a^{t+1} = Sf(a^t, \eta) \leqslant Sf(a^t, \eta_{\max}1) < f(\|a^t\|_\infty, \eta_{\max})1,$$

thus, $\|a^{t+1}\|_\infty < f(\|a^t\|_\infty, \eta_{\max})$. For $\underline{a}$ large, we have that $f(\underline{a}, \eta_{\max}) = \underline{a}\mathbb{E}(\underline{\xi} + \eta_{\max}/\sqrt{\underline{a}})_+^2 \sim \underline{a}/2$. Noting from (48) that $\|a^1\|_\infty < (1+\eta_{\max})^2$, we readily obtain that for each t, the elements of the vector a^t stay in a compact interval $\mathcal{Q}_a \subset \mathbb{R}_{*+}$ which is independent of n.

Let us study the iterations $a^{t+1} = Sf(a^t, \eta)$ given by (48). The vector-valued matrix function $F(a, \eta) = Sf(a, \eta)$ satisfies $|[F(a,\eta) - F(a',\eta)]_i| \leqslant \|S\| \max_i |f(a_i, \eta_i) - f(a'_i, \eta_i)| \leqslant \|S\| \, \|a - a'\|_\infty$, where $a = [a_i]$, $a' = [a'_i]$, and $\eta = [\eta_i]$. Thus $\|F(a,\eta) - F(a',\eta)\|_\infty \leqslant \|S\| \|a - a'\|_\infty$, and the convergence $a^t \to_t a$, where a is defined as the unique solution of Equation (50), is obtained by Banach's fixed point theorem. That this convergence is uniform in n in the norm $\|\cdot\|_\infty$ results from the inequality $\|a^t - a\|_\infty \leqslant \|S\|^{t-1} \|a^1 - a\|_\infty \leqslant \|S\|^{t-1} D_{\mathcal{Q}_a}$, where $D_{\mathcal{Q}_a}$ is the diameter of the compact $\mathcal{Q}_a$.

Recall that $\zeta^t = S\mathbb{P}[\sqrt{a^t}\xi + \eta \geqslant 0]$. The convergence $\zeta^t \to_t \zeta$ uniformly in n in the norm $\|\cdot\|_\infty$ is due to the boundedness of $(\|S^{(n)}\|)$, the continuity of the function $(\underline{a}, \eta) \mapsto \mathbb{P}[\sqrt{\underline{a}}\underline{\xi} + \eta \geqslant 0]$ on the compact $\mathcal{Q}_a \times \mathcal{Q}_\eta$, and to the convergence of a^t to a as above.

We now establish the convergence $q^t \to_t 1$. Let $[G_1, G_2]$ be a centered Gaussian vector such that $\mathbb{E}G_1^2 = \mathbb{E}G_2^2 = 1$, and such that $\mathbb{E}G_1 G_2 = \underline{q} \in [0,1]$. Writing $\mathcal{Q}_a = [a_{\min}, a_{\max}]$, define the compact interval $\mathcal{Q}_{\mathsf{H}} = [\eta_{\min}/\sqrt{a_{\max}}, \eta_{\max}/\sqrt{a_{\min}}]$, and define the continuous function $\mathsf{H} : [0,1] \times \mathcal{Q}_{\mathsf{H}} \to [0,1]$ as

$$(\underline{q}, \underline{b}) \mapsto \mathsf{H}(\underline{q}, \underline{b}) = \frac{\mathbb{E}(G_1 + \underline{b})_+ (G_2 + \underline{b})_+}{\mathbb{E}(G_1 + \underline{b})_+^2}.$$

A consequence of [MR16, Lemma 38 and proof of Lemma 37] is that for each $\underline{b} \in \mathcal{Q}_{\mathsf{H}}$, the function $\mathsf{H}(\cdot, \underline{b})$ satisfies $\mathsf{H}(q, \underline{b}) > q$ for all $q \in [0,1)$, and $\mathsf{H}(1, \underline{b}) = 1$.

We also define the continuous function $H : [0,1] \times \mathcal{Q}_{\mathsf{H}} \times \mathcal{Q}_{\mathsf{H}} \to [0,1]$ as

$$(\underline{q}, \underline{b}, \underline{d}) \mapsto H(\underline{q}, \underline{b}, \underline{d}) = \frac{\mathbb{E}(G_1 + \underline{b})_+ (G_2 + \underline{d})_+}{\left(\mathbb{E}(G_1 + \underline{b})_+^2 \mathbb{E}(G_2 + \underline{d})_+^2\right)^{1/2}}.$$

Given $t > 0$ and $k = t-1, t$, write the Gaussian vectors $\boldsymbol{Z}^k$ from the $(S, h, \eta, 1)$–state evolution equations as $\boldsymbol{Z}^k = \sqrt{a^k}\xi^k$, with $\xi^k \sim \mathcal{N}(0, I_n)$, and with $q^t = \mathbb{E}\xi^t \xi^{t-1}$. By the $(S, h, \eta, 1)$–state evolution equations, we have

$$\begin{aligned}
\mathbb{E}\boldsymbol{Z}^{t+1}\boldsymbol{Z}^t &= S\mathbb{E}(\boldsymbol{Z}^t + \eta)_+ (\boldsymbol{Z}^{t-1} + \eta)_+ \\
&= S\left(\mathbb{E}(\boldsymbol{Z}^t + \eta)_+^2 \mathbb{E}(\boldsymbol{Z}^{t-1} + \eta)_+^2\right)^{1/2} H(q^t, \eta/\sqrt{a^t}, \eta/\sqrt{a^{t-1}}) \\
&= S\left(f(a^t, \eta) f(a^{t-1}, \eta)\right)^{1/2} H(q^t, \eta/\sqrt{a^t}, \eta/\sqrt{a^{t-1}}),
\end{aligned}$$

and using the state evolution equations again, we obtain that

$$q^{t+1} = \left(a^{t+1} a^t\right)^{-1/2} S \left(f(a^t, \eta) f(a^{t-1}, \eta)\right)^{1/2} H(q^t, \eta/\sqrt{a^t}, \eta/\sqrt{a^{t-1}}).$$

Notice that $H(\underline{q}, \underline{b}, \underline{b}) = \mathsf{H}(\underline{q}, \underline{b})$. By the continuity of H on $[0,1] \times \mathcal{Q}_{\mathsf{H}} \times \mathcal{Q}_{\mathsf{H}}$ and the uniform convergence $\|a^t - a\|_\infty \to_t 0$, we obtain that $\|H(q^t, \eta/\sqrt{a^t}, \eta/\sqrt{a^{t-1}}) - \mathsf{H}(q^t, \eta/\sqrt{a})\|_\infty \to_t 0$ uniformly in n. We also have that $\|(a^{t+1}a^t)^{-1/2} - a^{-1}\|_\infty \to_t 0$ uniformly, and by using the continuity of f on $\mathcal{Q}_a \times \mathcal{Q}_\eta$, that $\|(f(a^t, \eta) f(a^{t-1}, \eta))^{1/2} - f(a, \eta)\|_\infty \to_t 0$ uniformly. Using the boundedness of $\|S\|$ and observing that $a^{-1} S f(a, \eta) = 1$, we obtain that $\|q^{t+1} - \mathsf{H}(q^t, \eta/\sqrt{a})\|_\infty \to_t 0$ uniformly in n.

Since the set of zeros of the continuous function $(\underline{q}, \underline{b}) \mapsto \underline{q} - \mathsf{H}(\underline{q}, \underline{b})$ on the compact $[0,1] \times \mathcal{Q}_{\mathsf{H}}$ is reduced to $\{1\} \times \mathcal{Q}_{\mathsf{H}}$, we deduce from the last convergence that $\|q^{(n),t} - 1_n\|_\infty \to_t 0$ uniformly in n. □

In the remainder of the proof, we reuse the notation $\overset{\mathcal{P}}{\leqslant}$ that was introduced before the subsection 4.3. Similarly, $\delta : (0, \epsilon) \to \mathbb{R}_+$ defined for some $\epsilon > 0$ is a generic function, independent of n, such that $\delta(e) \to 0$ as $e \to 0$. This function can change from a display to another.

Recalling that $W = \operatorname{diag}(1+\zeta)^{1/2} \Sigma \operatorname{diag}(1+\zeta)^{1/2}$, we have by Hypothesis 2 that $\limsup_n \|W\| < 2$ with probability one. Let us set $C_W = 2$ in the statement of Theorem 2.

Consider the AMP sequence (49). Fix a small number $e > 0$, and choose the index t in this sequence to be large enough (independently of n) so that

$$\begin{aligned}
&\mathbb{E}\|\boldsymbol{Z}^{t+1} - \boldsymbol{Z}^t\|_n^2 + \mathbb{E}\|\boldsymbol{Z}^t - \boldsymbol{Z}^{t-1}\|_n^2 \leqslant e, \\
&\|\zeta^t - \zeta\|_\infty \leqslant e, \quad \text{and} \\
&\|a^t - a\|_\infty \leqslant e,
\end{aligned} \tag{56}$$

which is possible by Lemma 24, after noting that $\mathbb{E}\|\boldsymbol{Z}^t - \boldsymbol{Z}^{t-1}\|_n^2 = n^{-1}\sum_i a_i^t + a_i^{t-1} - 2(a_i^t a_i^{t-1})^{1/2} q_i^t$.

Repeating the derivations that follow Equation (49), we reach the equation (52), and we obtain the bounds

$$\begin{aligned}
\|(1+\zeta)^{-1/2}\chi_1\|_n^2 &\leqslant \|(\zeta - \zeta^t)(x^{t-1} + \eta)_+\|_n^2 \\
&\leqslant \|(\zeta - \zeta^t)\|_\infty^2 \big(\|(x^{t-1}+\eta)_+\|_n^2 - \mathbb{E}\|(\boldsymbol{Z}^{t-1}+\eta)_+\|_n^2\big) + \|(\zeta-\zeta^t)\|_\infty^2 \mathbb{E}\|(\boldsymbol{Z}^{t-1}+\eta)_+\|_n^2 \\
&\overset{\mathcal{P}}{\leqslant} Ce^2
\end{aligned}$$

by Theorem 2 and Lemma 24, and

$$\begin{aligned}
\|(1+\zeta)^{-1/2}\chi_2\|_n^2 &\leqslant 2\|(x^t+\eta)_+ - (x^{t-1}+\eta)_+\|_n^2 + 2\|x^{t+1} - x^t\|_n^2 \\
&\leqslant 2\|(x^t+\eta)_+ - (x^{t-1}+\eta)_+\|_n^2 - 2\mathbb{E}\|(\boldsymbol{Z}^t+\eta)_+ - (\boldsymbol{Z}^{t-1}+\eta)_+\|_n^2 \\
&\quad + 2\mathbb{E}\|(\boldsymbol{Z}^t+\eta)_+ - (\boldsymbol{Z}^{t-1}+\eta)_+\|_n^2 \\
&\quad + 2\|x^{t+1} - x^t\|_n^2 - 2\mathbb{E}\|\boldsymbol{Z}^{t+1} - \boldsymbol{Z}^t\|_n^2 \\
&\quad + 2\mathbb{E}\|\boldsymbol{Z}^{t+1} - \boldsymbol{Z}^t\|_n^2 \\
&\overset{\mathcal{P}}{\leqslant} \delta(e),
\end{aligned}$$

by Theorem 2 and Lemma 24 again, and by noticing that $\mathbb{E}\|(\boldsymbol{Z}^t+\eta)_+ - (\boldsymbol{Z}^{t-1}+\eta)_+\|_n^2 + \mathbb{E}\|\boldsymbol{Z}^{t+1} - \boldsymbol{Z}^t\|_n^2 \leqslant e$ thanks to the Lipshitz property of the function $x \mapsto (x+\eta)_+$.

All in all, we have that the vector $u^t = (1+\zeta)^{1/2}(x^t+\eta)_+$ satisfies $u^t = \mathrm{LCP}(I - \Sigma, -r - \boldsymbol{\varepsilon}^t)$ on the event $[\|\Sigma\| < 1]$, with

$$\|\boldsymbol{\varepsilon}^t\|_n \overset{\mathcal{P}}{\leqslant} \delta(e). \tag{57}$$

We now tackle the third step of the proof, applying the LCP perturbation result of Chen and Xiang in [CX07]. By [CX07, Th. 2.7 and Th. 2.8], we have

$$\|u^t - u_\star\| \leqslant \big\|(I-\Sigma)^{-1}\big\| \, \|\boldsymbol{\varepsilon}^t\|,$$

thus, $\|u^t - u_\star\| \leqslant C\|\boldsymbol{\varepsilon}^t\|$ with probability one for all n large by Hypothesis 2.

Recall the definition of the Gaussian vector $Y = (1+\zeta)(\sqrt{p}\xi + r)$ in the statement of Theorem 5. To finish the proof of this theorem, it remains to prove that

$$\forall \varphi \in \mathrm{PL}_2(\mathbb{R}), \quad \frac{1}{n}\sum_{i\in[n]} \varphi(u_{\star,i}) - \mathbb{E}\varphi((Y_i)_+) \xrightarrow[n\to\infty]{\mathcal{P}} 0. \tag{58}$$

Notice that $Y = (1+\zeta)^{1/2}(\sqrt{a}\xi + \eta) \overset{\mathcal{L}}{=} (1+\zeta)^{1/2}(\boldsymbol{Z}+\eta)$ by Equations (54) and (55). For $\varphi \in \mathrm{PL}_2(\mathbb{R})$, we write

$$\begin{aligned}
&\frac{1}{n}\sum_{i\in[n]} \varphi(u_{\star,i}) - \mathbb{E}\varphi((1+\zeta_i)^{1/2}(Z_i+\eta_i)_+) \\
&= \frac{1}{n}\sum_{i\in[n]} \big(\varphi(u_{\star,i}) - \varphi(u_i^t)\big) + \frac{1}{n}\sum_{i\in[n]} \Big(\varphi(u_i^t) - \mathbb{E}\varphi((1+\zeta_i)^{1/2}(Z_i^t+\eta_i)_+)\Big) \\
&\quad + \frac{1}{n}\sum_{i\in[n]} \mathbb{E}\varphi((1+\zeta_i)^{1/2}(Z_i^t+\eta_i)_+) - \mathbb{E}\varphi((1+\zeta_i)^{1/2}(Z_i+\eta_i)_+) \\
&= \epsilon_1 + \epsilon_2 + \epsilon_3.
\end{aligned}$$

We have

$$|\epsilon_1| \leqslant \frac{C}{n} \sum_{i\in[n]} |u_{\star,i} - u_i^t|(1 + 2|u_i^t| + |u_{\star,i} - u_i^t|) \leqslant C\|u_\star - u^t\|_n(1 + 2\|u^t\|_n + \|u_\star - u^t\|_n)$$

by Cauchy-Schwarz. Thus, with probability one, $|\epsilon_1| \leqslant C\|\boldsymbol{\varepsilon}^t\|_n(1 + \|u^t\|_n + \|\boldsymbol{\varepsilon}^t\|_n)$ for all n large. Applying Theorem 2 to $\|u^t\|_n$ and using the bound $\|\boldsymbol{\varepsilon}^t\|_n \overset{\mathcal{P}}{\leqslant} \delta(e)$, we obtain that $|\epsilon_1| \overset{\mathcal{P}}{\leqslant} \delta(e)$.

We also have that

$$\epsilon_2 \xrightarrow[n\to\infty]{\mathcal{P}} 0 \tag{59}$$

by Theorem 2.

To deal with ϵ_3, we can write $\boldsymbol{Z}^t = \sqrt{a^t}\xi$ and $\boldsymbol{Z} = \sqrt{a}\xi$, with $\xi \sim \mathcal{N}(0, I_n)$, and use the pseudo-Lipschitz property of φ along with the bound (56) to show after a small derivation that $|\epsilon_3| \leqslant Ce$.

Putting these bounds together, we obtain that

$$\Big|\frac{1}{n}\sum_{i\in[n]} \varphi(u_{\star,i}) - \mathbb{E}\varphi((Y_i)_+)\Big| \overset{\mathcal{P}}{\leqslant} \delta(e).$$

Since e is arbitrary, the convergence (58) follows, and Theorem 5 is proven.

5.4 Proof of Lemma 23

Given $\zeta \in [0,1]^n$, Equation (11a) is $p = V(1+\zeta)^2 f(p,r)$, where f is the function introduced in the proof of Lemma 24. We have that $2f(\underline{p},\underline{r})/\underline{p} \to 1$ as $\underline{p} \to \infty$, uniformly in $\underline{r} \in \mathcal{Q}_r$ from Hypothesis 4. Thus, there exists $p_{\max} > 0$ such that $f(\underline{p},\underline{r}) \leqslant 3\underline{p}/4$ for each $\underline{p} \geqslant p_{\max}$ and each $\underline{r} \in \mathcal{Q}_r$. By consequence, if $\|p\|_\infty > p_{\max}$, then $\|V(1+\zeta)^2 f(p,r)\|_\infty \leqslant (3\|p\|_\infty/4)\|V(1+\zeta)^2 1_n\|_\infty < (3\|p\|_\infty/4)$ by Hypothesis 3, thus, $V(1+\zeta)^2 f(p,r)$ cannot be equal to p when $p \notin [0, p_{\max}]^n$. On the other hand, if $p \in [0,p_{\max}]^n$, it holds that $\|V(1+\zeta)^2 f(p,r)\|_\infty \leqslant f(p_{\max}, r_{\max})\|V(1+\zeta)^2 1_n\|_\infty \leqslant 3p_{\max}/4$, thus, $V(1+\zeta)^2 f(p,r) \in [0,p_{\max}]^n$.

Turning to Equation (11b), for each $\zeta \in [0,1]^n$, we see that $(1+\zeta)V(1+\zeta)\mathbb{P}\left[\sqrt{p}\xi + r \geqslant 0\right] \subset [0,1]^n$ for each p by Hypothesis 3 again.

Thus, writing the system (11) as $(p,\zeta) = G(p,\zeta)$, we obtain that $G([0,p_{\max}]^n \times [0,1]^n) \subset [0,p_{\max}]^n \times [0,1]^n$, and furthermore, G does not have a fixed point outside $(\mathbb{R}_+^n \backslash [0,p_{\max}]^n) \times [0,1]^n$. Since $[0,p_{\max}]^n \times [0,1]^n$ is a compact convex set of $\mathbb{R}^{2n}$, we obtain by Brouwer's fixed point theorem that G has a fixed point in this set. In particular, when (p,ζ) is a fixed point, $\|p\|_\infty < p_{\max}$.

To complete the proof, we need to show that this fixed point is unique. To that end, we rely on the construction of the previous paragraph, where we note that this uniqueness is never used.

In all the remainder of this proof, the integer n is fixed. Choose a solution (p,ζ) of the system (11). From this solution, construct the matrix $S^{(n)}$ and the vector $\eta^{(n)}$ according to (53) and (54) respectively. Let (M) be a sequence of integers converging to infinity. For each M, construct the symmetric matrix $\boldsymbol{S}^{(M)} \in \mathbb{R}_+^{nM\times nM}$ and the vector $\boldsymbol{\eta}^{(M)} \in \mathbb{R}_{*+}^{nM}$ as

$$\boldsymbol{S}^{(nM)} = S^{(n)} \otimes \left(M^{-1}1_M 1_M^\top\right), \quad \text{and} \quad \boldsymbol{\eta}^{(nM)} = \eta^{(n)} \otimes 1_M,$$

where $\otimes$ is the Kronecker product. Let $\boldsymbol{X}^{(nM)} = [\boldsymbol{X}_{ij}^{(nM)}]_{1\leqslant i,j\leqslant nM}$ be a real random symmetric $nM \times nM$ matrix such that the random variables $\{\boldsymbol{X}_{ij}^{(nM)}\}_{1\leqslant i<j\leqslant nM}$ are independent $\mathcal{N}(0,1)$ random variables, and such that $\boldsymbol{X}_{ii}^{(nM)} = 0$ for $i \in [nM]$. Define the random matrix $\boldsymbol{W}^{(nM)}$ as

$$\boldsymbol{W}^{(nM)} = (\boldsymbol{S}^{(nM)})^{\odot 1/2} \odot \boldsymbol{X}^{(nM)}.$$

We shall consider herein the $\mathbb{R}^{nM}$–valued AMP iterates based on the matrix $\boldsymbol{W}^{(nM)}$ and on the $(\boldsymbol{S}^{(nM)}, h, \boldsymbol{\eta}^{(nM)}, 1_{nM})$–state evolution equations, which take the form

$$\boldsymbol{u}^{(nM),t+1} = \boldsymbol{W}^{(nM)}\left(\boldsymbol{u}^{(nM),t} + \boldsymbol{\eta}^{(nM)}\right)_+ - \boldsymbol{\zeta}^{(nM),t}\left(\boldsymbol{u}^{(nM),t-1} + \boldsymbol{\eta}^{(nM)}\right)_+$$

(expression of $\boldsymbol{\zeta}^{(nM),t}$ omitted). In this context, one can check that Assumptions 1–5 and 7 applied to this model are satisfied with n and K_n there replaced with nM and K_nM respectively. Write

$$\boldsymbol{V}^{(nM)} = V^{(n)} \otimes (M^{-1} 1_M 1_M^\top) \quad \text{and} \quad \boldsymbol{r}^{(nM)} = r^{(n)} \otimes 1_M$$

(so that $\boldsymbol{S} = \operatorname{diag}(1_{nM} + \boldsymbol{\zeta})\boldsymbol{V} \operatorname{diag}(1_{nM} + \boldsymbol{\zeta})$ with $\boldsymbol{\zeta} = \zeta \otimes 1_M$ and $\boldsymbol{\eta} = (1_{nM} + \boldsymbol{\zeta})^{1/2}\boldsymbol{r}$), and let

$$\boldsymbol{\Sigma}^{(nM)} = (\boldsymbol{V}^{(nM)})^{\odot 1/2} \odot \boldsymbol{X}^{(nM)}.$$

Observe that since n is now fixed, our matrices $\boldsymbol{V}^{(nM)}$ are no more sparse, and $\|\boldsymbol{V}^{(nM)}\|_\infty \lesssim 1/M$. Thus, recalling the spectral norm controls made above in the Gaussian case, the positive number

$$\boldsymbol{T}^{(nM)}_{\text{Gauss}} = (1+\varepsilon)\left(2 \|\boldsymbol{V}^{(nM)}\|^{1/2} + \frac{6}{\sqrt{\log(1+\varepsilon)}} (\|\boldsymbol{V}^{(nM)}\|_\infty \log(nM))^{1/2} \right)$$

satisfies $\limsup_M \boldsymbol{T}^{(nM)}_{\text{Gauss}} < 1$ by choosing ε small enough. Thus, by the Gaussian concentration such as in (10), we obtain that $\limsup_M \|\boldsymbol{\Sigma}^{(nM)}\| < 1$ with the probability one. By consequence, on this event, the equation $\boldsymbol{u}_\star^{(nM)} = \operatorname{LCP}(I - \boldsymbol{\Sigma}^{(nM)}, -\boldsymbol{r}^{(nM)})$ is well-defined for all M large. It is important to note that this vector does not depend on the chosen solution (p, ζ) of the system (11).

Write $\boldsymbol{u}_\star^{(nM)} = \left[\boldsymbol{u}_\star^{(nM)}(1)^\top, \dots, \boldsymbol{u}_\star^{(nM)}(n)^\top\right]^\top$ where $\boldsymbol{u}_\star^{(nM)}(i) = \left[\boldsymbol{u}_{\star,1}^{(nM)}(i), \dots, \boldsymbol{u}_{\star,M}^{(nM)}(i)\right]^\top$. Recall the definition of the $\mathbb{R}^n$–valued Gaussian vector $Y^{(n)}$ as provided in the statement of Theorem 5. By repeating the argument of the previous paragraph, by relying this time on the AMP sequence $(\boldsymbol{u}^{(nM),t})_t$, we are able to show that

$$\forall \varphi \in \operatorname{PL}_2(\mathbb{R}), \quad \frac{1}{nM} \sum_{i\in[n]} \sum_{j \in [M]} \varphi(\boldsymbol{u}_{\star,j}^{(nM)}(i)) \xrightarrow[M\to\infty]{\mathcal{P}} \frac{1}{n} \sum_{i \in [n]} \mathbb{E}\varphi((Y_i)_+).$$

However, we need here a bit more than this convergence, which requires a slight modification of the approach of the previous paragraph. By relying on our new AMP construction, we have that for each $e > 0$, there exists an integer $t > 0$ and a random $\mathbb{R}^{nM}$–valued error vector $\boldsymbol{\varepsilon}^{(nM),t}$ such that

$$\left\|\boldsymbol{u}_\star^{(nM)} - \boldsymbol{u}^{(nM),t}\right\| \leqslant \left\|\left(I - \boldsymbol{\Sigma}^{(nM)}\right)^{-1}\right\| \left\|\boldsymbol{\varepsilon}^{(nM),t}\right\|$$

with

$$\mathbb{P}\left[\left\|\boldsymbol{\varepsilon}^{(nM),t}\right\|^2_{nM} \geqslant e\right] \xrightarrow[M\to\infty]{} 0$$

Writing $\boldsymbol{u}^{(nM),t} = \left[\boldsymbol{u}^{(nM),t}(1)^\top, \dots, \boldsymbol{u}^{(nM),t}(n)^\top\right]^\top$ with $\boldsymbol{u}^{(nM),t}(i) = \left[\boldsymbol{u}_1^{(nM),t}(i), \dots, \boldsymbol{u}_M^{(nM),t}(i)\right]^\top$ and remembering that n is fixed, this implies that there is $C > 0$ such that

$$\forall i \in [n], \ \mathbb{P}\left[\left\|\boldsymbol{u}_\star^{(nM)}(i) - \boldsymbol{u}^{(nM),t}(i)\right\|^2_M \geqslant Cne\right] \xrightarrow[M\to\infty]{} 0.$$

This is the analogue of the convergence (57).

Furthermore, let $\varphi \in \operatorname{PL}_2(\mathbb{R})$, let $i \in [n]$, and define the nM–uple

$$(\beta_1^{(nM)}, \dots \beta_{nM}^{(nM)}) = (0, \dots, 0, \underbrace{1, 1, \dots 1}_{\text{length } M}, 0, \dots, 0),$$

where the first element of the M–uple $(1, \dots, 1)$ is at the $((i-1)M + 1)^{\text{th}}$ place. By applying Theorem 2 with these weights, we obtain

$$\forall i \in [n], \quad \frac{1}{M} \sum_{l=1}^M \varphi(\boldsymbol{u}_l^{(nM),t}(i)) - \mathbb{E}\varphi((1+\zeta_i)^{1/2}(\sqrt{a_i^t}\xi + \eta_i)_+) \xrightarrow[M\to\infty]{\mathcal{P}} 0,$$

where $\xi \sim \mathcal{N}(0,1)$, and where the vector $a^t = [a_i^t]_{i=1}^n$ is precisely the one given by the recursion (48). This is the analogue of the convergence (59).

Completing the argument of the previous paragraph with these new convergences, we obtain that

$$\forall i \in [n],\ \mu^{\boldsymbol{u}_\star^{(nM)}(i)} \xrightarrow[M\to\infty]{\mathcal{P}} \mathscr{L}\left((1+\zeta_i)(\sqrt{p_i}\underline{\xi} + r_i)_+\right) \quad \text{in } \mathcal{P}_2(\mathbb{R}).$$

From the uniqueness of these limits, we deduce that the solution $(p = [p_i]_{i=1}^n, \zeta = [\zeta_i]_{i=1}^n)$ of the system (11) is unique.

Acknowledgements

I would like to thank Cédric Gerbelot, Marc Lelarge, and the authors of [ABC+22] who were all members of the French CNRS 80 prime project KARATE, for the inspiring and stimulating discussions.

A Proof of Proposition 3

Assume that $\rho > 0$. Following the notations of [BvH16], write $\sigma = \max_{i\in[n]}(\sum_{j\in[n]} s_{ij})^{1/2}$ and $\sigma_* = \max_{ij}\sqrt{s_{ij}}$. Then, the proof of [BvH16, Cor. 3.5] shows that

$$\left(\mathbb{E}\|W\|^{2\lceil \log n\rceil}\right)^{1/(2\lceil \log n\rceil)} \leqslant C\left(\sigma + \sigma_*(\log n)^{(\rho\vee 1)/2}\right) \leqslant C\left(1 + \sqrt{\frac{(\log n)^{\rho\vee 1}}{K_n}}\right),$$

where the second inequality is due to our Assumption 2. Using Markov's inequality and the hypothesis $K_n \gtrsim (\log n)^{\rho\vee 1}$, we obtain that

$$\mathbb{P}[\|W\| \geqslant \eta]^{1/(2\lceil \log n\rceil)} \leqslant C/\eta$$

for any $\eta > 0$. Choosing η large enough, the result follows from the Borel-Cantelli lemma.

If $\rho = 0$, we can just apply the concentration results provided by [BvH16, Cor 3.12 and Rem. 3.13].

References

[ABC+22] I. Akjouj, M. Barbier, M. Clenet, W. Hachem, M. Maïda, F. Massol, J. Najim, and V. C. Tran. Complex systems in ecology: A guided tour with large Lotka-Volterra models and random matrices. *arXiv preprint arXiv:2212.06136*, 2022.

[AHMN23] I. Akjouj, W. Hachem, M. Maïda, and J. Najim. Equilibria of large random Lotka-Volterra systems with vanishing species: a mathematical approach. *arXiv preprint arXiv:2302.07820*, 2023.

[AT12] S. Allesina and S. Tang. Stability criteria for complex ecosystems. *Nature*, 483(7388):205–208, 2012.

[AT15] S. Allesina and S. Tang. The stability–complexity relationship at age 40: a random matrix perspective. *Population Ecology*, 57(1):63–75, 2015.

[BBC18] G. Biroli, G. Bunin, and C. Cammarota. Marginally stable equilibria in critical ecosystems. *New Journal of Physics*, 20(8):083051, 2018.

[BGP14] F. Benaych-Georges and S. Péché. Largest eigenvalues and eigenvectors of band or sparse random matrices. *Electron. Commun. Probab.*, 19:no. 4, 9, 2014.

[BLM15] M. Bayati, M. Lelarge, and A. Montanari. Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.*, 25(2):753–822, 2015.

[BM11] M. Bayati and A. Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[BMN19] R. Berthier, A. Montanari, and P.-M. Nguyen. State evolution for approximate message passing with non-separable functions. *Information and Inference: A Journal of the IMA*, 9(1):33–79, 01 2019.

[Bol14] E. Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.

[BR22] J. K. Behne and G. Reeves. Fundamental limits for rank-one matrix estimation with groupwise heteroskedasticity. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 8650–8672. PMLR, 28–30 Mar 2022.

[BSHM17] D. M. Busiello, S. Suweis, J. Hidalgo, and A. Maritan. Explorability and the origin of network sparsity in living systems. *Scientific reports*, 7(1):12323, 2017.

[Bun17] G. Bunin. Ecological communities with Lotka-Volterra dynamics. *Phys. Rev. E*, 95:042414, Apr 2017.

[BvH16] A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *The Annals of Probability*, 44(4):2479 – 2506, 2016.

[CEN22] M. Clenet, H. El Ferchichi, and J. Najim. Equilibrium in a large Lotka–Volterra system with pairwise correlated interactions. *Stochastic Processes and their Applications*, 153:423–444, 2022.

[CL21] W.-K. Chen and W.-K. Lam. Universality of approximate message passing algorithms. *Electronic Journal of Probability*, 26(none):1 – 44, 2021.

[CNM22] M. Clenet, J. Najim, and F. Massol. Equilibrium and surviving species in a large Lotka-Volterra system of differential equations. *ArXiv preprint*, arXiv:2205.00735, 2022.

[CPS09] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The linear complementarity problem*, volume 60 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009. Corrected reprint of the 1992 original [MR1150683].

[CX07] X. Chen and S. Xiang. Perturbation bounds of P-matrix linear complementarity problems. *SIAM J. Optim.*, 18(4):1250–1265, 2007.

[DLS22] R. Dudeja, Y. M. Lu, and S. Sen. Universality of approximate message passing with semi-random matrices. *arXiv preprint arXiv:2204.04281*, 2022.

[DM16] D. Donoho and A. Montanari. High dimensional robust M-estimation: asymptotic variance via approximate message passing. *Probab. Theory Related Fields*, 166(3-4):935–969, 2016.

[Fan22] Z. Fan. Approximate message passing algorithms for rotationally invariant matrices. *Ann. Statist.*, 50(1):197–224, 2022.

[FSW18] D. Fourdrinier, W. E. Strawderman, and M. T. Wells. *Shrinkage estimation.* Springer Series in Statistics. Springer, Cham, 2018.

[FVRS22] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth. A unifying tutorial on approximate message passing. *Foundations and Trends® in Machine Learning*, 15(4):335–536, 2022.

[Gal18] T. Galla. Dynamically evolved community size and stability of random Lotka-Volterra ecosystems. *Europhysics Letters*, 123(4):48004, sep 2018.

[GB21] C. Gerbelot and R. Berthier. Graph-based Approximate Message Passing iterations. *ArXiv preprint*, arXiv:2109.11905, 2021.

[GKKZ22] A. Guionnet, J. Ko, F. Krzakala, and L. Zdeborová. Low-rank matrix estimation with inhomogeneous noise. *arXiv preprint arXiv:2208.05918*, 2022.

[JM13] A. Javanmard and A. Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013.

[Kho08] O. Khorunzhiy. Estimates for moments of random matrices with Gaussian elements. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 51–92. Springer, Berlin, 2008.

[LJM09] X. Li, D. Jiang, and X. Mao. Population dynamical behavior of Lotka-Volterra system under regime switching. *J. Comput. Appl. Math.*, 232(2):427–448, 2009.

[May72] R. M. May. Will a large complex system be stable? *Nature*, 238(5364):413–414, 1972.

[MR16] A. Montanari and E. Richard. Non-negative principal component analysis: message passing algorithms and sharp asymptotics. *IEEE Trans. Inform. Theory*, 62(3):1458–1484, 2016.

[OD92] M. Opper and S. Diederich. Phase transition and $1/f$ noise in a game dynamical model. *Phys. Rev. Lett.*, 69:1616–1619, Sep 1992.

[PKK23] A. Pak, J. Ko, and F. Krzakala. Optimal algorithms for the inhomogeneous spiked Wigner model. *arXiv preprint arXiv:2302.06665*, 2023.

[PS11] L. A. Pastur and M. Shcherbina. *Eigenvalue Distribution of Large Random Matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

[Ran11] S. Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, 2011.

[Sod10] S. Sodin. The spectral edge of some random band matrices. *Ann. of Math. (2)*, 172(3):2223–2251, 2010.

[Tak96] Y. Takeuchi. *Global dynamical properties of Lotka-Volterra systems.* World Scientific Publishing Co., Inc., River Edge, NJ, 1996.

[Tok04] K. Tokita. Species abundance patterns in complex evolutionary dynamics. *Phys. Rev. Lett.*, 93:178102, Oct 2004.

[Vil09] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 2009. Old and new.

[WZF22] T. Wang, X. Zhong, and Z. Fan. Universality of approximate message passing algorithms and tensor networks. *arXiv preprint arXiv:2206.13037*, 2022.