Approximate Message Passing for orthogonally invariant ensembles:
Multivariate non-linearities and spectral initialization

Xinyi Zhong\* Tianhao Wang\* Zhou Fan

Abstract

We study a class of Approximate Message Passing (AMP) algorithms for symmetric and rectangular spiked random matrix models with orthogonally invariant noise. The AMP iterates have fixed dimension $K \geq 1$, a multivariate non-linearity is applied in each AMP iteration, and the algorithm is spectrally initialized with $K$ super-critical sample eigenvectors. We derive the forms of the Onsager debiasing coefficients and corresponding AMP state evolution, which depend on the free cumulants of the noise spectral distribution. This extends previous results for such models with $K = 1$ and an independent initialization.

Applying this approach to Bayesian principal components analysis, we introduce a Bayes-OAMP algorithm that uses as its non-linearity the posterior mean conditional on all preceding AMP iterates. We describe a practical implementation of this algorithm, where all debiasing and state evolution parameters are estimated from the observed data, and we illustrate the accuracy and stability of this approach in simulations.

1 Introduction

In recent years, Approximate Message Passing (AMP) algorithms have been used in an increasingly diverse range of applications. These algorithms were originally derived as approximations to message passing and belief propagation algorithms for densely connected graphical models [Kab03, DMM09, DMM10a, DMM10b]. They have since been successfully adapted to perform both optimization and Bayesian inference in many problems arising in high-dimensional statistics and machine learning, and we refer to [FVRS21] for a recent review.

By design, AMP algorithms are closely tailored to distributional assumptions for the data matrices to which they are applied. These algorithms exhibit fast rates of convergence for typical realizations of such random data [Mal10], and can achieve near-optimal estimation risk in many contexts of Bayesian inference [KMS+12, DM14, DMK+16, DAM17, BKM+19, BMDK20]. Furthermore, iterates of AMP admit an exact asymptotic distributional characterization, known as its “state evolution”, that is simpler than that of alternative first-order procedures. Thus AMP has also served as a broadly useful theoretical tool for analyzing the asymptotic behavior of statistical methods [BM11b, MAYB13, DJM13, DM16, SBC17, SCC19, BKRS20] as well as probabilistic models [Bol18, DS19, FW21, CFM21].

The most common examples of AMP algorithms are tailored to data matrices with i.i.d. entries, and a line of work [Bol14, BM11a, JM13, BLM15, BMN20, CL21] has rigorously established the validity of their state evolutions in this context. This state evolution may no longer correctly

\*These authors contributed equally.

Yale University, Department of Statistics and Data Science.
xinyi.zhong@yale.edu, tianhao.wang@yale.edu, zhou.fan@yale.edu
describe the iterates for non-i.i.d. data, where such AMP algorithms may also exhibit divergent behavior [VSR+15, RSFS19]. More recently, several AMP algorithms have been developed for a broader class of random matrices that are orthogonally or unitarily invariant in law, but that can have arbitrary spectral distribution. These include the Orthogonal AMP [Tak17, MP17], Vector AMP [RSF19, SRF16], Convolutional AMP [Tak19, Tak20b, Tak20a], and Memory AMP [LHK21] procedures for linear models and generalized linear models with orthogonally invariant designs, as well as a general class of AMP procedures for symmetric and rectangular orthogonally invariant ensembles studied in [OCW16, CO20, Fan20]. Broadly, these algorithms adapt to the more general spectral laws that may arise in such models, by either using divergence-free non-linearities or applying Onsager corrections that are tailored to these spectral laws.

1.1 Contributions

Motivated by applications to statistical principal components analysis (PCA), our work extends the class of AMP algorithms for orthogonally invariant random matrix ensembles studied in [OCW16, CO20, Fan20], in two directions:

1. We extend the AMP procedures from vector-valued iterates $u_t \in \mathbb{R}^n$ to matrix-valued iterates $U_t \in \mathbb{R}^{n \times K}$, for an arbitrary fixed dimension $K \geq 1$. We consider such AMP algorithms that apply multivariate non-linearities in every iteration, and derive the forms of the Onsager corrections and state evolutions for algorithms of this type.

   Importantly, the non-linearities need not be separable across the $K$ dimensions. This generalization is particularly useful for PCA, where (empirically, in many domains of application) there is often joint structure across coordinates of multiple PCs, and a multivariate non-linearity should be used to regularize estimates towards this structure [ZSF20].

2. In a model of data consisting of low-rank signal plus additive noise, we develop a method of spectral initialization for the general AMP algorithms of [Fan20], using the sample eigenvectors or singular vectors of the data. This is analogous to the work of [MV21c] that developed this extension for AMP algorithms when the noise has i.i.d. entries.

   Such an extension eliminates the need for an informative initialization that is independent of the data, which is typically unavailable in practice. Our analysis shows that the AMP Onsager correction and state evolution must treat the spectral initialization separately, and they take different forms from the descriptions of [Fan20].

The first generalization above is a more direct extension of the previous analyses in [Fan20], and we describe these results in Section 2. The second generalization to a spectral initialization constitutes the larger technical contribution of our work, and we describe it for symmetric and rectangular matrices in Sections 3 and 4 respectively. Our proof is different from that of [MV21c], and instead follows a strategy introduced in [MV21a] of approximating the spectral initialization by a sequence of $\tau$ linear AMP steps that converge to the sample eigenvectors, as $\tau \to \infty$.

Recent independent work of [MV21b] has used this approach to derive also the forms of spectrally-initialized AMP algorithms corresponding to the “single-iterate posterior mean” PCA procedure described in [Fan20]; this constitutes an important case of our current results. Our results expand upon [MV21b], analyzing instead the general AMP algorithms in [Fan20] whose non-linearities may be functions of all preceding AMP iterates, and in the above context of multivariate iterates with dimension $K \geq 1$. At a technical level, we avoid the restrictive assumption imposed in [MV21b] that all free cumulants of the noise spectral law be positive, and we use an
alternative strategy for analyzing the convergence of the linear AMP iterations that leads to a
different and explicit assumption of sufficiently large signal strength.

Finally, as an application of these results, we propose in Section 5 a Bayes-OAMP algorithm
for PCA with a Bayesian prior for the PCs. This algorithm differs from the single-iterate posterior
mean procedure that was analyzed in [Fan20], computing instead the Bayes posterior mean based on
the multivariate Gaussian joint law of all preceding AMP iterates, as described by the above state
evolution. We demonstrate in Section 5 that this can yield a sizeable improvement in estimation
accuracy for rotationally invariant noise ensembles in settings of weak signal strength. We remark
that the Bayes-optimality of this approach remains (to us) unclear, and we leave this as an open
question for future work.

Notational conventions. For random variables \( X \) and \( Y \), \( X \perp \perp Y \) denotes that they are independent. \( \| \cdot \| \) denotes the \( \ell_2 \)-norm for vectors and the \( \ell_2 \to \ell_2 \) operator norm for matrices. \( \| \cdot \|_F \) is the Frobenius norm for matrices. \( \text{diag}(v) \in \mathbb{R}^{K \times K} \) is the diagonal matrix with \( v \in \mathbb{R}^K \) on its diagonal, and we write \( \text{diag}(v) \in \mathbb{R}^{K \times K} \) to indicate this matrix right-padded by \( K' - K \) columns of 0. We adopt the convention
\( M^0 = \text{Id} \) for the 0th power of any square matrix \( M \). For a block
matrix \( M \in \mathbb{R}^{tK \times tK} \) and \( \kappa \in \mathbb{R}^{K \times K} \), \( M \odot \kappa \in \mathbb{R}^{tK \times tK} \) and \( \kappa \odot M \in \mathbb{R}^{tK \times tK} \) denote the block-wise
right- and left- multiplication by \( \kappa \).

For a matrix \( U \in \mathbb{R}^{n \times K} \) and random vector \( U \in \mathbb{R}^K \), we write \( U \overset{W_2}{\to} U \) for the Wasserstein-2
convergence of the empirical distribution of rows of \( U \), as \( n \to \infty \). Letting \( u_i \) be the \( i \)th row of \( U \),
this means
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n g(u_i) = \mathbb{E}[g(U)]
\]
for any continuous function \( g : \mathbb{R}^K \to \mathbb{R} \) such that \( |g(u)| \leq C(1 + \|u\|^2) \) for a constant \( C > 0 \). We
write \( \langle U \rangle = n^{-1} \sum_{i=1}^n u_i \in \mathbb{R}^K \) for the empirical average of rows of \( U \).

2 AMP with independent initialization

In this section, we first describe extensions of the AMP algorithms and state evolution characteriza-
tions of [Fan20] from vector-valued to matrix-valued iterates, for orthogonally invariant matrices
and an independent initialization. We will then discuss signal-plus-noise models and spectral ini-
tializations in Sections 3 and 4.

2.1 Symmetric matrices

Let \( W \in \mathbb{R}^{n \times n} \) be a symmetric matrix, with eigen-decomposition \( W = O^\top \Lambda O \) where \( \Lambda = \text{diag}(\lambda) \).
We will assume that \( O \) is a Haar-distributed orthogonal basis of eigenvectors, so \( W \) is orthogonally
invariant in law.

For fixed dimensions \( J \geq 0 \) and \( K \geq 1 \), consider a possible additional matrix \( E \in \mathbb{R}^{n \times J} \) of “side
information”, and a sequence of Lipschitz functions \( u_2, u_3, \ldots \) where each \( u_{t+1} : \mathbb{R}^{K+J} \to \mathbb{R}^K \).
(We may set \( J = 0 \) if there is no such side information.) We consider an AMP algorithm with
initialization \( U_1 \in \mathbb{R}^{n \times K} \) independent of \( W \), having the iterates
\[
Z_t = WU_t - U_1b_1^\top - U_2b_2^\top - \cdots - U_tb_t^\top
\]
\[
U_{t+1} = u_{t+1}(Z_1, \ldots, Z_t, E).
\]
Here $b_{ts} \in \mathbb{R}^{K \times K}$ is a matrix-valued Onsager debiasing coefficient for each $s = 1, \ldots, t$, and $u_{t+1}(\cdot)$ is applied row-wise to $(Z_1, \ldots, Z_t, E) \in \mathbb{R}^{n \times (tK + J)}$ to yield each next iterate $U_{t+1} \in \mathbb{R}^{n \times K}$.

**Debiasing coefficients.** Let $\partial_s u_{t+1}(Z_1, \ldots, Z_t, E) \in \mathbb{R}^{K \times K}$ denote the Jacobian of $u_{t+1}$ in its vector argument $Z_s$, which exists Lebesgue-a.e. since $u_{t+1}(\cdot)$ is Lipschitz [Zie12, Theorem 2.2.1]. Denote

$$\langle \partial_s U_{t+1} \rangle = \frac{1}{n} \sum_{i=1}^{n} \partial_s u_{t+1}(z_{1,i}, \ldots, z_{t,i}, e_i)$$

where $z_{t,i} \in \mathbb{R}^K$ and $e_i \in \mathbb{R}^J$ are the $i^{th}$ rows of $Z_t$ and $E$. For each $T \geq 1$, define the $TK \times TK$ block-lower-triangular matrix

$$\phi_T = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \langle \partial_1 U_2 \rangle & 0 & \cdots & 0 & 0 \\ \langle \partial_1 U_3 \rangle & \langle \partial_2 U_3 \rangle & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \langle \partial_1 U_T \rangle & \langle \partial_2 U_T \rangle & \cdots & \langle \partial_{T-1} U_T \rangle & 0 \end{pmatrix}.$$ (2.3)

Let $\Lambda$ be a random variable on $\mathbb{R}$ with compact support, which will be the limit eigenvalue distribution of $W$ as $n \to \infty$. Let $\{\kappa_j\}_{j=1}^{\infty}$ be the free cumulants of $\Lambda$—see e.g. [Fan20, Section 2.3] for definitions. Applying the convention $\phi_0^T = \text{Id}$, we take the debiasing coefficient matrices $\{b_{ts}\}$ in (2.1) up to iteration $T$ to be the blocks of

$$b_T = \sum_{j=0}^{\infty} \kappa_{j+1} \phi_j^T = \begin{pmatrix} b_{11} & b_{21} & b_{22} & \cdots & b_{TT} \\ b_{21} & b_{22} & \cdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{T1} & b_{T2} & \cdots & b_{TT} & \end{pmatrix} \in \mathbb{R}^{TK \times TK}.$$ This may be interpreted as the $R$-transform of $\Lambda$ applied to $\phi_T$ (cf. Section 3.1). Note that this is a finite sum, because $\phi_j^T = 0$ for all $j \geq T$. We have $b_{tt} = \kappa_1 \text{Id}$ for every $t \geq 1$, which vanishes if $\Lambda$ has mean $\kappa_1 = 0$.

**State Evolution.** This choice of debiasing coefficients leads to the empirical distribution of rows of $(Z_1, \ldots, Z_T)$ having an asymptotically mean-zero multivariate Gaussian limit, as $n \to \infty$ for any fixed iteration $T$. The limit Gaussian law is described by its covariance matrix $\Sigma_T \in \mathbb{R}^{TK \times TK}$, which may be defined recursively via the following state evolution:

Let $(U_1, E)$ be an initial random vector with $U_1 \in \mathbb{R}^K$ and $E \in \mathbb{R}^J$, representing the limit empirical distribution of rows of $(U_1, E)$. Inductively for $t = 1, 2, 3, \ldots$, having defined the joint law of $(U_1, \ldots, U_{t-1}, Z_1, \ldots, Z_{t-1}, E)$, define the $tK \times tK$ matrices

$$\Delta_t = \begin{pmatrix} \mathbb{E}[U_1 U_1^\top] & \mathbb{E}[U_1 U_2^\top] & \cdots & \mathbb{E}[U_1 U_T^\top] \\ \mathbb{E}[U_2 U_1^\top] & \mathbb{E}[U_2 U_2^\top] & \cdots & \mathbb{E}[U_2 U_T^\top] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[U_T U_1^\top] & \mathbb{E}[U_T U_2^\top] & \cdots & \mathbb{E}[U_T U_T^\top] \end{pmatrix},$$

$$\Phi_t = \begin{pmatrix} \mathbb{E}[\partial_1 u_2(Z_1, E)] & 0 & \cdots & 0 & 0 \\ \mathbb{E}[\partial_1 u_3(Z_1, Z_2, E)] & \mathbb{E}[\partial_2 u_2(Z_1, Z_2, E)] & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbb{E}[\partial_1 u_t(Z_1, \ldots, Z_{t-1}, E)] & \mathbb{E}[\partial_2 u_t(Z_1, \ldots, Z_{t-1}, E)] & \cdots & \mathbb{E}[\partial_{t-1} u_{t}(Z_1, \ldots, Z_{t-1}, E)] & 0 \end{pmatrix}.$$ (2.5)
Here, $\Phi_t$ corresponds to the large-$n$ limit of $\phi_t$ defined in (2.3). Then define the covariance $\Sigma_t$ by

$$
\Sigma_t = \sum_{j=0}^{\infty} \Theta^{(j)}[\Phi_t, \kappa_{j+2}\Delta_t] \quad \text{where} \quad \Theta^{(j)}[\Phi, \kappa\Delta] = \sum_{i=0}^{j} \Phi^i(\kappa\Delta)(\Phi^\top)^{j-i}. 
$$

(2.6)

Define the next joint law of $(U_1, \ldots, U_{t+1}, Z_1, \ldots, Z_t, E)$ by

$$(Z_1, \ldots, Z_t) \sim \mathcal{N}(0, \Sigma_t) \perp \perp (U_1, E), \quad U_{s+1} = u_{s+1}(Z_1, \ldots, Z_s, E) \text{ for each } s = 1, \ldots, t. \quad (2.7)$$

Under these inductive definitions, it may be checked that the upper-left $(t-1) \times (t-1)$ blocks of $\Sigma_t$ coincide with $\Sigma_{t-1}$.

This state evolution characterizes the iterates of the AMP algorithm (2.1–2.2), under the following assumptions.

**Assumption 2.1.** The matrix $W = O^\top \text{diag}(\lambda)O$ and random variable $\Lambda$ satisfy

(a) $O$ is random and Haar-distributed over the orthogonal group.

(b) $\lambda$ is independent of $O$, and its empirical distribution converges weakly a.s. to $\Lambda$ as $n \to \infty$.

(c) $\Lambda$ has compact support $\text{supp}(\Lambda)$. Denoting $(\lambda_-, \lambda_+) = (\min \text{supp}(\Lambda), \max \text{supp}(\Lambda))$, we have $\min(\lambda) \to \lambda_-$ and $\max(\lambda) \to \lambda_+$ a.s. as $n \to \infty$.

**Assumption 2.2.** The AMP initialization $U_1$, functions $u_2, u_3, \ldots$, and random vectors $(U_1, E)$ satisfy

(a) $(U_1, E) \in \mathbb{R}^{n \times (K+J)}$ is independent of $O$, and $(U_1, E) \xrightarrow{W_2} (U_1, E)$ a.s. as $n \to \infty$.

(b) Each $u_{t+1}(\cdot)$ is Lipschitz in all arguments. For each $s = 1, \ldots, t$, $\partial_s u_{t+1}(Z_1, \ldots, Z_t, E)$ exists and is continuous on a set of probability 1 under the law of $(Z_1, \ldots, Z_t, E)$ defined by (2.7).

**Theorem 2.3.** Suppose Assumptions 2.1 and 2.2 hold. For any $T \geq 1$, consider the AMP algorithm (2.1–2.2) up to iteration $T$, and define $(U_1, \ldots, U_{T+1}, Z_1, \ldots, Z_T, E)$ by the state evolution (2.7). Then almost surely as $n \to \infty$,

$$(U_1, \ldots, U_{T+1}, Z_1, \ldots, Z_T, E) \xrightarrow{W_2} (U_1, \ldots, U_{T+1}, Z_1, \ldots, Z_T, E).$$

The proof of Theorem 2.3 is an extension of that of [Fan20, Theorem 4.3 and Corollary 4.4]. Compared with [Fan20, Corollary 4.4], Theorem 2.3 considers matrix-valued iterates having dimension $n \times K$, relaxes the needed convergence $(U_1, E) \to (U_1, E)$ from Wasserstein-$p$ for all orders $p \geq 1$ to only Wasserstein-2, and relaxes the continuous-differentiability requirement for each function $u_{t+1}(\cdot)$ to the weaker condition of Assumption 2.2(b). We describe the modifications of the proofs of [Fan20] needed to establish Theorem 2.3 in Appendix C.

### 2.2 Rectangular matrices

Now let $W \in \mathbb{R}^{m \times n}$ be a rectangular matrix, with singular value decomposition $W = O^\top \Lambda Q$. To simplify the exposition, let us assume that

$$\gamma \overset{\text{def}}{=} m/n \leq 1.$$
For $m/n > 1$, our results may be applied to $W^T$. We denote $\Lambda = \text{diag}(\lambda)$ where $\lambda \in \mathbb{R}^m$ are the singular values of $W$. We assume that $O$ and $Q$ are Haar-distributed orthogonal bases of singular vectors, so $W$ is bi-rotationally invariant in law.

For fixed $J, L \geq 0$ and $K \geq 1$, consider matrices of side information $E \in \mathbb{R}^{n \times J}$ and $F \in \mathbb{R}^{n \times L}$, and sequences of Lipschitz functions $u_2, u_3, \ldots$ and $v_1, v_2, \ldots$. We study an AMP algorithm with initialization $U_1 \in \mathbb{R}^{m \times K}$ independent of $W$, having the iterates (of dimensions $m \times K$ and $n \times K$)

$$Z_t = W^T U_t - V_1 b_{t1}^T - \ldots - V_{t-1} b_{t,t-1}^T \quad (2.8)$$

$$V_t = v_t(Z_1, \ldots, Z_t, F) \quad (2.9)$$

$$Y_t = W V_t - U_1 a_{11}^T - \ldots - U_t a_{tt}^T \quad (2.10)$$

$$U_{t+1} = u_{t+1}(Y_1, \ldots, Y_t, E). \quad (2.11)$$

Here again, $b_{ts}, a_{ts} \in \mathbb{R}^{K \times K}$ are Onsager debiasing coefficient matrices, and $v_t(\cdot)$ and $u_{t+1}(\cdot)$ are applied row-wise.

**Debiasing coefficients.** Let $\Lambda \in \mathbb{R}$ be a non-negative random variable with compact support, which will be the limit singular value distribution of $W$. Let $\{\kappa_{2j}\}_{j \geq 1}$ be the rectangular free cumulants of $\Lambda$ with aspect ratio $\gamma = m/n$—see [Fan20, Section 2.4] for definitions. For $T \geq 1$, define analogously to (2.3) the block-lower-triangular $TK \times TK$ matrices

$$\phi_T = \left(\langle \partial_s U_r \rangle\right)_{r,s \in \{1, \ldots, T\}} \quad \psi_T = \left(\langle \partial_s V_r \rangle\right)_{r,s \in \{1, \ldots, T\}}$$

with rows blocks indexed by $r$ and column blocks by $s$. Here, denoting $y_{t,i}, z_{t,i}, e_i, f_i$ as the $i$th rows of $Y_t, Z_t, E, F$, we set

$$\langle \partial_s U_r \rangle = \frac{1}{m} \sum_{i=1}^m \partial_s u_r(y_{1,i}, \ldots, y_{r-1,i}, e_i), \quad \langle \partial_s V_r \rangle = \frac{1}{n} \sum_{i=1}^n \partial_s v_r(z_{1,i}, \ldots, z_{r,i}, f_i)$$

and use the conventions $\partial_s u_r = 0$ for $s \geq r$ and $\partial_s v_r = 0$ for $s > r$. Then the matrices $a_{ts}$ and $b_{ts}$ in (2.8) and (2.10) up to iteration $T$ are the blocks of

$$a_T = \sum_{j=0}^\infty \kappa_{2(j+1)} \phi_T (\phi_T \psi_T)^j$$

and

$$b_T = \gamma \sum_{j=0}^\infty \kappa_{2(j+1)} \phi_T (\psi_T \phi_T)^j$$

**State Evolution.** The state of this algorithm up to iteration $T$ is characterized by two covariance matrices $\Sigma_T, \Omega_T \in \mathbb{R}^{TK \times TK}$ describing the limit multivariate Gaussian laws of the rows of $(Y_1, \ldots, Y_T), (Z_1, \ldots, Z_T)$. These are defined recursively as follows.

Let $(U_1, E) \in \mathbb{R}^{K+J}$ and $F \in \mathbb{R}^L$ be random vectors, which will be the limit empirical distribution of rows of $(U_1, E)$ and $F$. Inductively for $t = 1, 2, 3, \ldots$ having defined joint laws of
(\(U_1, \ldots, U_t, Y_1, \ldots, Y_{t-1}, E\)) and (\(V_1, \ldots, V_{t-1}, Z_1, \ldots, Z_t, F\)), we define the \(tK \times tK\) block matrices analogous to (2.4) and (2.5),
\[
\Delta_t = \left( \mathbb{E}[U_r U_s^T] \right)_{r,s \in \{1, \ldots, t\}}, \quad \Phi_t = \left( \mathbb{E}[\partial_s u_r(Y_1, \ldots, Y_{t-1}, E)] \right)_{r,s \in \{1, \ldots, t\}},
\]
\[
\Gamma_t = \left( \mathbb{E}[V_r V_s^T] \right)_{r,s \in \{1, \ldots, t\}}, \quad \Psi_t = \left( \mathbb{E}[\partial_s v_r(Z_1, \ldots, Z_t, E)] \right)_{r,s \in \{1, \ldots, t\}},
\]
with row blocks indexed by \(r\) and column blocks by \(s\). We then define the covariance
\[
\Omega_t = \gamma \sum_{j=0}^{\infty} \Theta^{(j)}[\Phi_t, \Psi_t, \kappa_2(j+1)\Delta_t, \kappa_2(j+1)\Gamma_t]
\]
where we denote
\[
\Theta^{(j)}[\Phi, \Psi, \kappa, \kappa \Gamma] = \sum_{i=0}^{j} (\Phi \Psi)^i (\kappa \Delta) (\Psi^T \Phi^T)^{j-i} + \sum_{i=0}^{j-1} (\Phi \Psi)^i (\kappa \Gamma) (\Psi^T \Phi^T)^{j-1-i}.
\]
(2.12)

Note that the final \(t^{\text{th}}\) row block of \(\Psi_t\) and \(t^{\text{th}}\) row and column blocks of \(\Gamma_t\) are not yet well-defined, as we have not yet defined \((V_t, Z_t)\). We permit this ambiguity because the matrix products defining \(\Theta^{(j)}\) in (2.12) do not depend on these blocks, as the last column block of \(\Phi_t\) is 0. From \(\Omega_t\), we define the joint law of \((V_1, \ldots, V_t, Z_1, \ldots, Z_t, F)\) by
\[
(Z_1, \ldots, Z_t) \sim \mathcal{N}(0, \Omega_t) \perp F, \quad V_s = v_s(Z_1, \ldots, Z_s, F) \text{ for each } s = 1, \ldots, t.
\]
(2.13)

Now, having defined \((U_1, \ldots, U_t, Y_1, \ldots, Y_{t-1}, E)\) and \((V_1, \ldots, V_t, Z_1, \ldots, Z_t, F)\), all blocks of \(\Gamma_t\) and \(\Psi_t\) are well-defined, and we define the covariance
\[
\Sigma_t = \sum_{j=0}^{\infty} \Xi^{(j)}[\Phi_t, \Psi_t, \kappa_2(j+1)\Delta_t, \kappa_2(j+1)\Gamma_t]
\]
where
\[
\Xi^{(j)}[\Phi, \Psi, \kappa \Delta, \kappa \Gamma] = \sum_{i=0}^{j} (\Phi \Psi)^i (\kappa \Gamma) (\Psi^T \Phi^T)^{j-i} + \sum_{i=0}^{j-1} (\Phi \Psi)^i (\kappa \Delta) (\Psi^T \Phi^T)^{j-1-i}.
\]
(2.14)

From \(\Sigma_t\), we then define the joint law of \((U_1, \ldots, U_{t+1}, Y_1, \ldots, Y_t, E)\) by
\[
(Y_1, \ldots, Y_t) \sim \mathcal{N}(0, \Sigma_t) \perp (U_1, E), \quad U_{s+1} = u_{s+1}(Y_1, \ldots, Y_s, E) \text{ for each } s = 1, \ldots, t,
\]
(2.15)
completing these inductive definitions. Under these definitions, it may be checked that the upper-left \((t-1) \times (t-1)\) blocks of \(\Sigma_t\) and \(\Omega_t\) coincide with \(\Sigma_{t-1}\) and \(\Omega_{t-1}\).

This state evolution characterizes the iterates of the AMP algorithm (2.8–2.11) under the following assumptions.

**Assumption 2.4.** The ratio \(\gamma = m/n \leq 1\) is fixed as \(m, n \to \infty\). The matrix \(W = O^\top \text{diag}(\lambda) Q\) and random variable \(\Lambda\) satisfy

(a) \(O\) and \(Q\) are random, independent, and Haar-distributed over the orthogonal groups.

(b) \(\lambda\) is independent of \(O, Q\), with empirical distribution converging weakly a.s. to \(\Lambda\) as \(m, n \to \infty\).

7
(c) \( \Lambda \) has compact support. Denoting \( \lambda_+ = \max \text{supp}(\Lambda) \), \( \max(\lambda) \to \lambda_+ \) a.s. as \( m, n \to \infty \).

**Assumption 2.5.** The AMP initialization \( U_1, E, F \) are independent of \( O, Q \), with \( (U_1, E) \xrightarrow{W_2} (U_1, E) \) and \( F \xrightarrow{W_2} F \) as \( m, n \to \infty \).

(b) Each \( u_{t+1} \cdot \) and \( v_t \cdot \) is Lipschitz in all arguments. For each \( s = 1, \ldots, t \), \( \partial_s u_{t+1}(Y_1, \ldots, Y_t, E) \) and \( \partial_s v_t(Z_1, \ldots, Z_t, F) \) exist and are continuous on sets of probability 1 under the laws of \( (Y_1, \ldots, Y_t, E) \) and \( (Z_1, \ldots, Z_t, F) \) defined by (2.15) and (2.13).

**Theorem 2.6.** Suppose Assumptions 2.4 and 2.5 hold. Fix any \( T \geq 1 \), consider the AMP algorithm (2.8–2.11) up to iteration \( T \), and define \( (U_1, \ldots, U_{T+1}, Y_1, \ldots, Y_T, E) \) and \( (V_1, \ldots, V_T, Z_1, \ldots, Z_T, F) \) by (2.15) and (2.13). Then almost surely as \( m, n \to \infty \),

\[
(U_1, \ldots, U_{T+1}, Y_1, \ldots, Y_T, E) \xrightarrow{W_2} (U_1, \ldots, U_{T+1}, Y_1, \ldots, Y_T, E),
(V_1, \ldots, V_T, Z_1, \ldots, Z_T, F) \xrightarrow{W_2} (V_1, \ldots, V_T, Z_1, \ldots, Z_T, F).
\]

Theorem 2.6 extends [Fan20, Theorem 5.3 and Corollary 5.4] in ways that are analogous to the extensions provided by the preceding Theorem 2.3 in the square symmetric setting. The proof of Theorem 2.6 modifies the proofs of [Fan20, Theorem 5.3 and Corollary 5.4] using the same ideas as described in Appendix C to prove Theorem 2.3, and we omit this proof for brevity.

**Remark 2.7.** We have defined \( b_{ts}, a_{ts} \) in (2.1), (2.8), and (2.10) using the free cumulants or rectangular free cumulants \( \{ \kappa_j \} \) of the limit spectral distributions. Theorems 2.3 and 2.6 then also hold for any AMP algorithm where \( b_{ts}, a_{ts} \) are replaced by \( b'_{ts}, a'_{ts} \) such that \( \|b_{ts} - b'_{ts}\| \to 0 \) and \( \|a_{ts} - a'_{ts}\| \to 0 \) a.s. as \( m, n \to \infty \). In particular, they hold if \( b_{ts}, a_{ts} \) are instead defined with \( \{ \kappa_j \} \) being any consistent estimates of these limit free cumulants.

## 3 Spectral initialization for the symmetric spiked model

We now develop versions of the preceding AMP algorithms for “spiked” signal-plus-noise models, with spectral initialization. Consider a rank-\( K' \) symmetric spiked model

\[
X = \sum_{k=1}^{K'} \frac{\theta_k}{n} u_k^* u_k^{*\top} + W \in \mathbb{R}^{n \times n},
\]

where \( u_1^*, \ldots, u_{K'}^* \) are \( K' \) orthogonal “signal” eigenvectors, \( \theta_1, \ldots, \theta_{K'} \) are non-zero signal eigenvalues, and the noise matrix \( W = O^\top \Lambda O \) is symmetric and rotationally-invariant in law. We distinguish this rank \( K' \) from the dimension \( K \) of the AMP iterates, in anticipation of applications where \( K' \) may be modeled as large, and the practitioner may wish to apply AMP to estimate small subsets of \( K \) signal eigenvectors at a time. We develop a version of AMP where any \( K \) super-critical sample eigenvectors of \( X \) may be chosen as the spectral initialization.

In the model (3.1), we fix the normalization

\[
\|u_k^*\|^2 = n, \quad u_j^* u_k^* = 0 \quad \text{for all } j \neq k \in \{1, \ldots, K'\}.
\]

We order the signal components such that the first \( K \) will correspond to the spectral initialization, and the remaining \( K' - K \) are ordered arbitrarily. (Thus \( \theta_1, \ldots, \theta_{K'} \) are not sorted, and they may
have arbitrary signs.) Supposing that $K_+$ values $\theta_1, \ldots, \theta_{K'}$ are positive and $K_- = K' - K_+$ values are negative, we denote by

$$\lambda_1(\mathbf{X}), \ldots, \lambda_{K'}(\mathbf{X})$$

the largest $K_+$ and smallest $K_-$ sample eigenvalues of $\mathbf{X}$, sorted in the same order as $\theta_1, \ldots, \theta_{K'}$. We denote the associated sample eigenvectors of $\mathbf{X}$ by $\mathbf{f}_pca^1, \ldots, \mathbf{f}_{pca}^{K'}$, with the normalization and sign convention

$$\|\mathbf{f}_{\text{pca}}^k\|^2 = n, \quad \mathbf{f}_{\text{pca}}^k \top \mathbf{u}_* = 0, \quad \mathbf{f}_{\text{pca}}^j \top \mathbf{f}_{\text{pca}}^k = 0 \quad \text{for all } j \neq k \in \{1, \ldots, K'\}.$$ (3.4)

### 3.1 Preliminaries on sample eigenvectors and the $R$-transform

**Eigenvectors and spectral phase transition.** For the signal-plus-noise model (3.1), the quantitative behavior of the leading (positive and negative) sample eigenvalues/eigenvectors of $\mathbf{X}$ and associated phenomena of spectral phase transitions were studied in [BGN11], extending the work of [BBAP05, BS06, Pau07] for models with i.i.d. noise. These depend on the Cauchy-transform of the limit spectral distribution of $\mathbf{W}$, and we briefly review these results here.

Let $\Lambda$ be the limit spectral distribution of $\mathbf{W}$. Denote the Cauchy-transform of $\Lambda$ by

$$G(z) = \mathbb{E}[(z - \Lambda)^{-1}] \quad \text{for } z \in (\lambda_+, \infty) \cup (-\infty, \lambda_-)$$

where $\lambda_\pm$ are the endpoints of support of $\Lambda$ as defined in Assumption 2.1(c). $G(z)$ is strictly decreasing and positive on $(\lambda_+, \infty)$ and strictly decreasing and negative on $(-\infty, \lambda_-)$, and hence admits a functional inverse $G^{-1}(z)$ on $(0, G(\lambda_+)) \cup (G(\lambda_-), 0)$ where $G(\lambda_\pm) = \lim_{z \to \pm} G(z)$. For $k \in \{1, \ldots, K'\}$ such that $1/\theta_k$ belongs to this domain of $G^{-1}(z)$, define

$$\lambda_{\text{pca}, k} = G^{-1}(1/\theta_k), \quad \mu_{\text{pca}, k}^2 = \frac{-1}{\theta_k^2 G'(\lambda_{\text{pca}, k})}. \quad \text{ (3.5)}$$

The following theorem summarizes several results of [BGN11, Theorems 2.1 and 2.2], which establish the first-order behavior of the leading sample eigenvalues and eigenvectors.

**Theorem 3.1 (BGN11).** Suppose $\mathbf{W}$ satisfies Assumption 2.1, and $\theta_1, \ldots, \theta_{K'}$ are distinct and fixed as $n \to \infty$. Then for each $k \in \{1, \ldots, K'\}$ where $\theta_k > 1/G(\lambda_+) \geq 0$ or $\theta_k < 1/G(\lambda_-) \leq 0$, almost surely

$$\lim_{n \to \infty} \lambda_k(\mathbf{X}) = \lambda_{\text{pca}, k}, \quad \lim_{n \to \infty} \left(\frac{\mathbf{f}_{\text{pca}}^k \top \mathbf{u}_*/n}{\lambda_{\text{pca}, k}}\right)^2 = \mu_{\text{pca}, k}^2.$$  

For each other $k \in \{1, \ldots, K'\}, \lim_{n \to \infty} \lambda_k(\mathbf{X}) \in \{\lambda_-, \lambda_+\}$ and $\lim_{n \to \infty}(\mathbf{f}_{\text{pca}}^k \top \mathbf{u}_*/n)^2 = 0$. Also for each $j \neq k \in \{1, \ldots, K'\}, \lim_{n \to \infty}(\mathbf{f}_{\text{pca}}^j \top \mathbf{u}_*/n)^2 = 0$.

Thus, for “super-critical” signal eigenvalues $\theta_k$ exceeding the positive and negative phase transition thresholds $1/G(\lambda_+)$ and $1/G(\lambda_-)$, the corresponding sample eigenvalue $\lambda_k(\mathbf{X})$ converges to a deterministic value $\lambda_{\text{pca}, k}$ outside the interval $[\lambda_-, \lambda_+]$, the sample eigenvector $\mathbf{f}_{\text{pca}}^k$ achieves asymptotically non-vanishing alignment with its corresponding signal vector $\mathbf{u}_k$, and it has asymptotically 0 alignment with the other signal vectors $\mathbf{u}_j$. For “sub-critical” signal eigenvalues $\theta_k$ below these phase transition thresholds, $\lambda_k(\mathbf{X})$ converges to the spectral edges $\lambda_\pm$ of the noise spectral distribution, and $\mathbf{f}_{\text{pca}}^k$ is asymptotically uninformative. Note that $G(\lambda_+)$ and $G(\lambda_-)$ may be infinite, in which case the phase transition thresholds are 0, and all signal eigenvalues are super-critical. Whether this occurs depends on the rates of decay of the distribution of $\Lambda$ at its spectral edges $\lambda_\pm$, and this is discussed further in [BGN11, Proposition 2.4].
**R-transform.** The above first-order limits may be re-expressed via the R-transform and free cumulants of Λ, which linearize free addition of independent random matrices. Define

\[ R(z) = G^{-1}(z) - \frac{1}{z} \]

on the same domain as \( G^{-1} \). Differentiating on both sides yields

\[ R'(z) = \frac{1}{G'(G^{-1}(z))} + \frac{1}{z^2}. \]

For \( z \) small enough, \( R(z) \) and its derivative have the convergent series expansions defined through the free cumulants \( \{\kappa_j\}_{j \geq 1} \) of \( \Lambda \),

\[
R(z) = \sum_{j=0}^{\infty} \kappa_{j+1} z^j, \quad R'(z) = \sum_{j=0}^{\infty} (j+1)\kappa_{j+2} z^j; \tag{3.6}
\]

see e.g. \[MS17, \text{Theorem 17}\]. Thus for sufficiently large \(|\theta_k|\), the limits \( \lambda_{\text{pca},k} \) and \( \mu_{\text{pca},k}^2 \) in (3.5) also have the convergent series forms

\[
\lambda_{\text{pca},k} = \theta_k + R(1/\theta_k) = \theta_k + \sum_{j=0}^{\infty} \frac{\kappa_{j+1}}{\theta_k^j}, \quad 1 - \mu_{\text{pca},k}^2 = \frac{1}{\theta_k^2} R'(1/\theta_k) = \sum_{j=0}^{\infty} (j+1)\frac{\kappa_{j+2}}{\theta_k^{j+2}}. \tag{3.7}
\]

### 3.2 AMP algorithm

Isolating the first \( K \leq K' \) of the (unsorted) signal components for the spectral initialization, denote

\[
S = \text{diag}(\theta_1, \ldots, \theta_K) \in \mathbb{R}^{K \times K}, \quad S' = \text{diag}(\theta_1, \ldots, \theta_{K'}) \in \mathbb{R}^{K' \times K'},
\]

\[
F_{\text{pca}} = (f_{\text{pca}}^1, \ldots, f_{\text{pca}}^K) \in \mathbb{R}^{n \times K}, \quad U_* = (u_*^1, \ldots, u_*^{K'}) \in \mathbb{R}^{n \times K'}.
\]

We consider an AMP algorithm with iterates of dimension \( n \times K \), initialized spectrally at

\[
U_0 = F_{\text{pca}} S^{-1}, \quad F_0 = F_{\text{pca}}. \tag{3.8}
\]

For a sequence of Lipschitz functions \( u_1, u_2, \ldots \) where \( u_t : \mathbb{R}^{tK} \to \mathbb{R}^K \), this algorithm then iteratively computes for \( t \geq 1 \)

\[
U_t = u_t(F_0, F_1, \ldots, F_{t-1}), \tag{3.9}
\]

\[
F_t = XU_t - U_0 b_{t0}^T - U_1 b_{t1}^T - \ldots - U_t b_{tt}^T. \tag{3.10}
\]

Thus each \( u_t(\cdot) \) may depend on the preceding iterates \( F_0, \ldots, F_{t-1} \), including the spectral initialization. An additional matrix of side information may be incorporated into each \( u_t(\cdot) \) as in (2.2), but we omit this here for simplicity. To ease notation in the analysis, we have shifted the initialization index from 1 to 0.

**Debiasing coefficients.** For each fixed \( s \geq 1 \), define diagonal matrices \( \bar{\kappa}_s, \check{\kappa}_s \in \mathbb{R}^{K \times K} \) by the matrix series

\[
\bar{\kappa}_s = \sum_{j=0}^{\infty} \kappa_{j+s} S^{-j}, \quad \check{\kappa}_s = \sum_{j=0}^{\infty} (j+1)\kappa_{j+s} S^{-j}. \tag{3.11}
\]
For each $T \geq 1$, define the block-lower-triangular matrix

$$
\phi_T = \begin{pmatrix} \partial_s U_r \end{pmatrix}_{r,s \in \{0, \ldots, T\}} \in \mathbb{R}^{(T+1)K \times (T+1)K}
$$

with row blocks indexed by $r$ and column blocks by $s$, where $\langle \partial_s U_r \rangle = n^{-1} \sum_{i=1}^n \partial_s u_r(f_{0,i}, \ldots, f_{r-1,i})$, $\partial_s u_r \in \mathbb{R}^{K \times K}$ denotes the Jacobian of $u_r(f_{0}, \ldots, f_{r-1})$ in the argument $f_s$, and $\partial_s u_r = 0$ for $s \geq r$.

Define the matrix series

$$
b_T = \sum_{j=0}^\infty \kappa_{j+1} \phi_T^j, \quad \tilde{b}_T = \sum_{j=0}^\infty \phi_T^j \circ \tilde{\kappa}_{j+1},
$$

where we recall our notation $M \circ \tilde{\kappa}_s$ and $\tilde{\kappa}_s \circ M$ for the right- and left- multiplication of each block of $M$. Indexing blocks by $\{0, \ldots, T\}$ and writing $[t, s]$ to denote the $K \times K$ submatrix corresponding to row block $t$ and column block $s$, we set the debiasing coefficients of (3.10) up to iteration $T$ as

$$
b_{ts} = \begin{cases} 
\tilde{b}_T[t, s] & \text{if } s = 0, \\
b_T[t, s] & \text{otherwise}. 
\end{cases} \quad (3.12)
$$

**State Evolution.** The state of this algorithm up to iteration $T$ is characterized by

$$
\begin{pmatrix} \mu_0 \\ \vdots \\ \mu_T \end{pmatrix} \in \mathbb{R}^{(T+1)K \times K'}, \quad 
\Sigma_T = \begin{pmatrix} 
\sigma_{00} & \cdots & \sigma_{0T} \\
\vdots & \ddots & \vdots \\
\sigma_{T0} & \cdots & \sigma_{TT} 
\end{pmatrix} \in \mathbb{R}^{(T+1)K \times (T+1)K}
$$

and a corresponding joint law for random vectors $(U'_s, U_0, \ldots, U_{T+1}, F_0, \ldots, F_T)$, where $U'_s, U_t, F_t$ represent the limit distributions of $U'_s, U_t, F_t$, the matrix $\Sigma_T$ is the covariance of $(F_0, \ldots, F_T)$ conditional on $U'_s$, and $\mu_T$ relates the conditional mean of $(F_0, \ldots, F_T)$ to $U'_s$. These are defined recursively as follows:

Let $U'_* \in \mathbb{R}^{K'}$ be a random vector satisfying $\mathbb{E}[U'_* U'_*^\top] = \text{Id}$, representing the limit distribution of rows of $U'_s$ under the normalization (3.2). Set $\mu_{\text{pca}} = \text{diag}(\mu_{\text{pca},1}, \ldots, \mu_{\text{pca},K}) \in \mathbb{R}^{K \times K'}$, where the last $K' - K$ columns are 0. Then $\text{Id} - \mu_{\text{pca}} \Sigma_T \mu_{\text{pca}}^\top = \text{diag}(1 - \mu_{\text{pca},1}^2, \ldots, 1 - \mu_{\text{pca},K}^2) \in \mathbb{R}^{K \times K}$. We initialize

$$
\mu_0 = \mu_0 = \mu_{\text{pca}}, \quad \Sigma_0 = \sigma_0 = \text{Id} - \mu_{\text{pca}} \Sigma_{\text{pca}} \mu_{\text{pca}}^\top.
$$

Inductively, having defined $(\mu_{t-1}, \Sigma_{t-1})$, we define a joint law for $(U'_*, U_0, \ldots, U_t, F_0, \ldots, F_{t-1})$ by

$$
(F_0, \ldots, F_{t-1}) \mid U'_* \sim \mathcal{N}(\mu_{t-1} \cdot U'_*, \Sigma_{t-1}), \
U_0 = S^{-1} F_0, \quad U_s = u_s(F_0, \ldots, F_{s-1}) \text{ for } s = 1, \ldots, t. \quad (3.13)
$$

We then define the next mean transformation $\mu_s$ to have the blocks

$$
\mu_s = \mathbb{E}[U_s U'_* \Sigma'] S' \text{ for each } s = 0, \ldots, t. \quad (3.14)
$$

For $s = 0$, this may be checked to coincide with the above initialization $\mu_{\text{pca}}$. 

11
We decompose the second moment matrix of \((U_0, \ldots, U_t)\) into four parts,

\[
\Delta_t \overset{\text{def}}{=} \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & E[U_t U_t^\top] & \cdots & E[U_t U_0^\top] \\ \vdots & \vdots & \ddots & \vdots \\ 0 & E[U_0 U_t^\top] & \cdots & E[U_0 U_0^\top] \end{pmatrix} + \begin{pmatrix} 0 & E[U_0 U_t^\top] & \cdots & E[U_0 U_0^\top] \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}
\]

\[
\Delta_t + \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & E[U_t U_0^\top] & \cdots & E[U_t U_0^\top] \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},
\]  

(3.15)

set

\[
\Delta_t^{(j)} = \kappa_{j+2} \Delta_t + \tilde{\kappa}_{j+2} \odot \Delta_t + \Delta_t^\top \odot \tilde{\kappa}_{j+2} + \tilde{\kappa}_{j+2} \odot \tilde{\Delta}_t,
\]

and define analogously to (2.3)

\[
\Phi_t = \mathbb{E}[\partial_s u_r(F_0, \ldots, F_{r-1})] \in \mathbb{R}^{(t+1)K \times (t+1)K}.
\]

Then, recalling the function \(\Theta^{(j)}[\cdot, \cdot]\) from (2.6), we define the next covariance matrix \(\Sigma_t\) by

\[
\Sigma_t = \sum_{j=0}^{\infty} \Theta^{(j)}[\Phi_t, \Delta_t^{(j)}].
\]  

(3.16)

It may be checked from these definitions that the first \(t\) blocks of \(\mu_t\) coincide with \(\mu_{t-1}\), and the upper-left \(t \times t\) blocks of \(\Sigma_t\) coincide with \(\Sigma_{t-1}\).

Our main result in the context of model (3.1) shows that this state evolution provides a rigorous characterization of the AMP algorithm (3.9–3.10) with the spectral initialization (3.8), under the following assumptions.

**Assumption 3.2.**

(a) \(U'_s = (u'_1, \ldots, u'_K)\) is independent of \(O\), satisfies (3.2), and \(U'_s \xrightarrow{W_2} U'_s\) a.s. as \(n \to \infty\) where \(\mathbb{E}[U'_s U'_s^\top] = \text{Id}\).

(b) Each \(u_{t+1}^{(s)}\) is Lipschitz in all arguments. For each \(s = 0, \ldots, t\) and all \((\mu, \Sigma)\) in a sufficiently small open neighborhood of \((\mu_t, \Sigma_t)\) defined by (3.14) and (3.16), \(\partial_s u_{t+1}(F_0, \ldots, F_t)\) exists and is continuous on a set of probability 1 under the marginal law of \((F_0, \ldots, F_t)\) defined by \((F_0, \ldots, F_t) \mid U'_s \sim \mathcal{N}(\mu \cdot U'_s, \Sigma')\).

(c) \(\theta_1, \ldots, \theta_K\) are distinct. For each \(k \in \{1, \ldots, K\}\), either \(\theta_k > G(1/\lambda_+) \geq 0\) or \(\theta_k < G(1/\lambda_-) \leq 0\), and there exists some constant \(\iota \in (0, 1)\) such that

\[
\frac{\max(|\lambda_+|, |\lambda_-|)}{|\theta_k|} + \sum_{j=1}^{\infty} \frac{|\kappa_j|}{|\theta_k|^2 \cdot \iota^{j-1}} < 1.
\]  

(3.17)

Assumption 3.2(c) requires \(\theta_1, \ldots, \theta_K\) for the first \(K\) selected signals to be “super-critical”, as described by Theorem 3.1. Furthermore, (3.17) requires each \(|\theta_k|\) to exceed some constant depending only on the law of \(\Lambda\). (We believe that this additional requirement may be an artifact of the proof technique, and we do not optimize the value of this constant.) Note that this requirement (3.17) is sufficient to imply that the series (3.6) for \(R(z)\) and \(R'(z)\) are absolutely convergent at \(z = 1/\theta_k\), and hence the series (3.11) defining \(\tilde{\kappa}_s, \hat{\kappa}_s\) are also absolutely convergent.
Theorem 3.3. Consider the symmetric spiked model (3.1), where Assumptions 2.1 and 3.2 hold. For any \( T \geq 1 \), consider the spectrally initialized AMP algorithm (3.8–3.10) up to iteration \( T \), and define \((U'_s, U_0, \ldots, U_{T+1}, F_0, \ldots, F_T)\) by (3.13). Then almost surely as \( n \to \infty \),

\[
(U'_s, U_0, \ldots, U_{T+1}, F_0, \ldots, F_T) \xrightarrow{W_2} (U'_s, U_0, \ldots, U_{T+1}, F_0, \ldots, F_T).
\]

3.3 Proof idea

We provide the proof of Theorem 3.3 in Appendix A, and describe here the main idea. We construct an auxiliary AMP algorithm consisting of two phases. For some \( \tau \geq 1 \), we index the iterates in the first phase from \(-\tau\) to 0. We consider an initialization \( U^{(\tau)}_{-\tau} \) that is independent of \( W \), and apply linear AMP iterations with \( U^{(\tau)}_t = F^{(\tau)}_{t-1} S^{-1} \) for \( t = -\tau + 1, \ldots, 0 \). This has the effect of implementing a version of the power method to compute the sample eigenvectors \( F_{\text{pca}} \) of \( X \). A main step of the proof is to show that as \( \tau \to \infty \), the final iterate \( F^{(\tau)}_0 \) obtained from these linear updates approaches the spectral initialization \( F_0 = F_{\text{pca}} \).

In the second phase, starting from \( F^{(\tau)}_0 \), we then apply the same functions \( u_t(\cdot) \) as in the spectrally-initialized AMP algorithm for \( t = 1, 2, \ldots \). The combined auxiliary AMP algorithm can thus be summarized as follows:

\[
F^{(\tau)}_t = X U^{(\tau)}_t - \sum_{s=-\tau}^{t} U^{(\tau)}_s b^{(\tau)}_{ls}^\top, \quad U^{(\tau)}_t = u_t(F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_t)
\]

where

\[
u_{t+1}(F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_t) = \begin{cases} 
F^{(\tau)}_t S^{-1} & \text{for } -\tau + 1 \leq t < 0, \\
u_t(F^{(\tau)}_0, \ldots, F^{(\tau)}_t) & \text{for } 0 \leq t \leq T.
\end{cases}
\]

This AMP algorithm may be analyzed using Theorem 2.3 with side information \( E = U'_s \), to yield a state evolution for iterates \( t \geq 0 \) characterized by some \((\mu^{(\tau)}_t, \Sigma^{(\tau)}_t)\), which we describe explicitly in Corollary A.1. Then we prove that in the limit \( \tau \to \infty \), the iterates of this auxiliary AMP algorithm for \( t \geq 0 \) will be close to those of the spectrally-initialized AMP algorithm, in the sense that

\[
\limsup_{\tau \to \infty, n \to \infty} \frac{\|F^{(\tau)}_t - F_t\|_F}{\sqrt{n}} = 0, \quad \limsup_{\tau \to \infty, n \to \infty} \frac{\|U^{(\tau)}_t - U_t\|_F}{\sqrt{n}} = 0.
\]

We also prove that as \( \tau \to \infty \), the state evolution characterizing (3.18) converges to the state evolution described by (3.14) and (3.16), i.e. for each fixed \( t \geq 1 \),

\[
\lim_{\tau \to \infty} \mu^{(\tau)}_t = \mu_t \quad \text{and} \quad \lim_{\tau \to \infty} \Sigma^{(\tau)}_t = \Sigma_t.
\]

Combining the above two facts, we then can characterize the behavior of the AMP algorithm with spectral initialization.

4 Spectral initialization for the rectangular spiked model

In the rectangular setting, we consider analogously a rank-\( K' \) spiked signal-plus-noise model

\[
X = \sum_{k=1}^{K'} \frac{\theta_k}{\sqrt{mn}} u_k v_k^\top + W \in \mathbb{R}^{m \times n}.
\]
Here \( u_1^*, \ldots, u_{K'}^* \) and \( v_1^*, \ldots, v_{K'}^* \) are \( K' \) pairs of left and right signal singular vectors, normalized so that
\[
\|u_k^*\|^2 = m, \quad u_k^* \top u_k^* = 0, \quad \|v_k^*\|^2 = n, \quad v_k^* \top v_k^* = 0 \quad \text{for all } j \neq k \in \{1, \ldots, K'\}. \tag{4.2}
\]
We again order the signal singular values \( \theta_1, \ldots, \theta_{K'} \) (not necessarily in sorted order) so that the first \( K \) will correspond to the spectral initialization. We will assume \( W = \Omega \top \Lambda Q \) is bi-rotationally invariant in law.

We denote by
\[
\lambda_1(X), \ldots, \lambda_{K'}(X) \tag{4.3}
\]
the largest \( K' \) sample singular values of \( X \), sorted in the same order as \( \theta_1, \ldots, \theta_{K'} \). We denote their associated sample left singular vectors by \( f_{\text{pca}}^1, \ldots, f_{\text{pca}}^{K'} \) and right singular vectors by \( g_{\text{pca}}^1, \ldots, g_{\text{pca}}^{K'} \), normalized such that for all \( j \neq k \in \{1, \ldots, K'\} \),
\[
\|f_k^\top\|^2 = m, \quad f_k^\top u_k^* \geq 0, \quad f_k^\top f_k^* = 0, \quad \|g_k^\top\|^2 = n, \quad g_k^\top v_k^* \geq 0, \quad g_k^\top g_k^* = 0. \tag{4.4}
\]

### 4.1 Preliminaries on sample singular vectors and the rectangular \( R \)-transform

#### Singular vectors and spectral phase transition
We review results of [BGN12] that characterize the leading sample singular values/vectors of \( X \) and the associated spectral phase transition. (See [BGN12, Remark 2.6] for the equivalence with our setting.)

Let \( \Lambda \) be the limit singular value distribution of \( W \), and recall its largest point of support \( \lambda_+ \) defined in Assumption 2.4(c). Recall \( \gamma = m/n \leq 1 \). Following [BGN12], for \( z \in (\lambda_+, \infty) \), define
\[
\varphi(z) = E \left[ \frac{z}{z^2 - \Lambda^2} \right], \quad \bar{\varphi}(z) = \gamma \varphi(z) + \frac{1 - \gamma}{z}, \quad D(z) = \varphi(z) \bar{\varphi}(z). \tag{4.5}
\]
The function \( D(z) \) is strictly decreasing on \((\lambda_+, \infty)\). Let \( D^{-1}(z) \) be its functional inverse on \((0, D(\lambda_+))\), where \( D(\lambda_+) = \lim_{z \to \lambda_+} D(z) \). For each \( k \in \{1, \ldots, K'\} \) where \( 1/\theta_k^2 \) belongs to this domain of \( D^{-1}(z) \), define
\[
\lambda_{\text{pca},k} = D^{-1}(1/\theta_k^2), \quad \mu_{\text{pca},k}^2 = \frac{-2 \varphi(\lambda_{\text{pca},k})}{\theta_k^2 D'(\lambda_{\text{pca},k})}, \quad \nu_{\text{pca},k}^2 = \frac{-2 \bar{\varphi}(\lambda_{\text{pca},k})}{\theta_k^2 D'(\lambda_{\text{pca},k})}. \tag{4.6}
\]
The following theorem summarizes results of [BGN12, Theorems 2.8 and 2.9].

**Theorem 4.1** ([BGN12]). Suppose \( \theta_1, \ldots, \theta_{K'} \) are distinct, \( \gamma = m/n \leq 1 \), and these are fixed as \( m, n \to \infty \). Suppose \( W \) satisfies Assumption 2.4. Then for each \( k \in \{1, \ldots, K'\} \) where \( \theta_k > (D(\lambda_+))^{-1/2} \), almost surely
\[
\lim_{m,n \to \infty} \lambda_k(X) = \lambda_{\text{pca},k}, \quad \lim_{m,n \to \infty} \left( \frac{\tilde{f}_{\text{pca}}^\top u_k^*}{m} \right)^2 = \mu_{\text{pca},k}^2, \quad \lim_{m,n \to \infty} \left( \frac{\tilde{g}_{\text{pca}}^\top v_k^*}{n} \right)^2 = \nu_{\text{pca},k}^2.
\]
For each other \( k \in \{1, \ldots, K'\} \), almost surely \( \lim_{m,n \to \infty} \lambda_k(X) = \lambda_+ \), \( \lim_{m,n \to \infty} (\tilde{f}_{\text{pca}}^\top u_k^*/m)^2 = 0 \), and \( \lim_{m,n \to \infty} (\tilde{g}_{\text{pca}}^\top v_k^*/n)^2 = 0 \). Also for each \( j \neq k \in \{1, \ldots, K'\} \), \( \lim_{m,n \to \infty} (\tilde{f}_{\text{pca}}^\top u_k^*/m)^2 = 0 \) and \( \lim_{m,n \to \infty} (\tilde{g}_{\text{pca}}^\top v_k^*/n)^2 = 0 \).

This describes a spectral phase transition phenomenon analogous to Theorem 3.1, where signal singular values are “super-critical” if \( \theta_k > (D(\lambda_+))^{-1/2} \). If \( D(\lambda_+) = \infty \), then all signals are super-critical. Whether this occurs is again determined by the decay of the law of \( \Lambda \) at the spectral edge \( \lambda_+ \), c.f. [BGN12, Proposition 2.11].
Rectangular $R$-transform. To introduce the rectangular $R$-transform, first denote

$$T(z) = (1 + z)(1 + \gamma z), \quad U(z) = \frac{-\gamma - 1 + \sqrt{(1 + \gamma)^2 + 4\gamma z}}{2\gamma},$$

so that $T(U(z - 1)) = z$. Following [BGN12, Eq. (8)], the rectangular $R$-transform is defined for $z \in (0, D(\Lambda^+))$ by

$$R(z) = U(z(D^{-1}(z))^2 - 1). \tag{4.7}$$

Recalling the rectangular free cumulants $\{\kappa_{2j}\}_{j \geq 1}$ of $\Lambda$, for sufficiently small $z > 0$, $R(z)$ and its derivative admit the convergent series expansions

$$R(z) = \sum_{j=1}^{\infty} \kappa_{2j} z^j, \quad R'(z) = \sum_{j=0}^{\infty} (j + 1) \kappa_{2(j+1)} z^j. \tag{4.8}$$

See [BG09, Section 3.4], with the notational identification $H^\mu(z^{-2}) = D(z)$ and hence $(D^{-1}(z))^2 = 1/H^\mu_2^{-1}(z)$ as shown in [Fan20, Appendix C.3].

The identity $T(U(z - 1)) = z$ gives $T(R(z)) = (1 + \gamma R(z))(1 + R(z)) = z(D^{-1}(z))^2$. Then the definition of $\lambda_{\text{pca},k}$ in (4.6) may be written as

$$\lambda_{\text{pca},k}^2/\theta_k^2 = (1 + \gamma R(\theta_k^{-2}))(1 + R(\theta_k^{-2})).$$

Rearranging this identity, we may factor $\theta_k^2$ as the product of

$$\theta_{e,k} \overset{\text{def}}{=} \frac{\lambda_{\text{pca},k}}{\sqrt{\gamma(1 + R(\theta_k^{-2}))}} \quad \text{and} \quad \theta_{u,k} \overset{\text{def}}{=} \frac{\lambda_{\text{pca},k} \sqrt{\gamma}}{1 + \gamma R(\theta_k^{-2})}. \tag{4.9}$$

Then $\theta_{u,k}$ and $\theta_{e,k}$ satisfy

$$\theta_{u,k} \cdot \theta_{e,k} = \theta_k^2, \quad \frac{\lambda_{\text{pca},k}}{\sqrt{\gamma \theta_{e,k}}} - R(\theta_k^{-2}) = 1, \quad \frac{\lambda_{\text{pca},k} \sqrt{\gamma}}{\theta_{u,k}} - \gamma R(\theta_k^{-2}) = 1.$$

[Fan20, Appendix C.3] verifies that $\mu_{\text{pca},k}^2$ and $\nu_{\text{pca},k}^2$ in (4.6) may also be expressed via $R(\theta_k^{-2})$ and $R'(\theta_k^{-2})$ as

$$\mu_{\text{pca},k}^2 = \frac{T(R(\theta_k^{-2})) - \theta_k^{-2} T'(R(\theta_k^{-2})) R'(\theta_k^{-2})}{1 + \gamma R(\theta_k^{-2})}, \quad \nu_{\text{pca},k}^2 = \frac{T(R(\theta_k^{-2})) - \theta_k^{-2} T'(R(\theta_k^{-2})) R'(\theta_k^{-2})}{1 + \gamma R(\theta_k^{-2})}. \tag{4.10}$$

4.2 AMP algorithm

We again isolate the first $K \leq K'$ (unsorted) signal components for the spectral initialization, and define

$$S = \text{diag}(\theta_1, \ldots, \theta_K) \in \mathbb{R}^{K \times K}, \quad S' = \text{diag}(\theta_1, \ldots, \theta_{K'}) \in \mathbb{R}^{K' \times K'},$$
$$S_u = \text{diag}(\theta_{u,1}, \ldots, \theta_{u,K}) \in \mathbb{R}^{K \times K}, \quad S_v = \text{diag}(\theta_{v,1}, \ldots, \theta_{v,K}) \in \mathbb{R}^{K \times K},$$
$$U'_* = (u'_{1*}, \ldots, u'_{K'}*) \in \mathbb{R}^{m \times K'}, \quad V'_* = (v'_{1*}, \ldots, v'_{K'}*) \in \mathbb{R}^{n \times K'},$$
$$F_{\text{pca}} = (f_{\text{pca},1}, \ldots, f_{\text{pca},K}) \in \mathbb{R}^{n \times K}, \quad G_{\text{pca}} = (g_{\text{pca},1}, \ldots, g_{\text{pca},K}) \in \mathbb{R}^{n \times K}. $$
We consider an AMP algorithm with matrix-valued iterates of dimensions $n \times K$ and $m \times K$, initialized spectrally at

$$\begin{align*}
F_0 &= F_{\text{pca}}, & U_0 &= U_1 = F_{\text{pca}}S_u^{-1}, \\
G_0 &= G_1 = G_{\text{pca}}, & V_0 &= G_{\text{pca}}S_v^{-1}.
\end{align*}$$

(4.11)

Duplications of $U_1 = U_0$ and $G_1 = G_0$ are introduced here to simplify the expression of the state evolution to follow. For $t \geq 1$ and sequences of Lipschitz functions $v_t : \mathbb{R}^{tK} \to \mathbb{R}^K$ and $u_{t+1} : \mathbb{R}^{(t+1)K} \to \mathbb{R}^K$, this algorithm computes the iterations

$$\begin{align*}
V_t &= v_t(G_1, \ldots, G_t), \\
F_t &= XV_t - \sum_{j=0}^t U_j a_{tj}^\top, \\
U_{t+1} &= u_{t+1}(F_0, \ldots, F_t), \\
G_{t+1} &= X^\top U_{t+1} - \sum_{j=0}^t V_j b_{t+1,j}^\top.
\end{align*}$$

(4.12)

Thus each $v_t(\cdot)$ and $u_{t+1}(\cdot)$ may depend on all preceding iterates $G_s$ and $F_s$, including the spectral initializations $G_1$ and $F_0$.

**Debiasing coefficients.** We define two $(T + 1)K \times (T + 1)K$ block-lower-triangular matrices

$$\phi_T = \begin{pmatrix}
0 & 0 & \cdots & 0 & 0 \\
S_u^{-1} & 0 & \cdots & 0 & 0 \\
\langle \partial_0 U_2 \rangle & \langle \partial_1 U_2 \rangle & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\langle \partial_0 U_T \rangle & \langle \partial_1 U_T \rangle & \cdots & \langle \partial_{T-1} U_T \rangle & 0
\end{pmatrix}, \quad \psi_T = \begin{pmatrix}
S_v^{-1} & 0 & 0 & \cdots & 0 \\
0 & \langle \partial_1 V_1 \rangle & 0 & \cdots & 0 \\
0 & \langle \partial_1 V_2 \rangle & \langle \partial_2 V_2 \rangle & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \langle \partial_1 V_T \rangle & \langle \partial_2 V_T \rangle & \cdots & \langle \partial_T V_T \rangle
\end{pmatrix}.$$  

Here $S_u^{-1}, S_v^{-1}$ may be interpreted as $\langle \partial_0 U_1 \rangle, \langle \partial_0 V_0 \rangle$, and the first column of $\psi_T$ as $\langle \partial_0 V_t \rangle = 0$ for all $t \geq 1$. For each fixed $s \geq 1$, define $\tilde{\kappa}_{2s}, \tilde{\kappa}_{2s} \in \mathbb{R}^{K \times K}$ by the matrix series

$$\tilde{\kappa}_{2s} = \sum_{j=0}^\infty \kappa_{2(j+s)} S_v^{-2j}, \quad \tilde{\kappa}_{2s} = \sum_{j=0}^\infty (j+1) \kappa_{2(j+s)} S_v^{-2j}.$$  

(4.13)

Then define the $(T + 1)K \times (T + 1)K$ matrices

$$\begin{align*}
a_T &= \sum_{j=0}^\infty \kappa_{2(j+1)} \psi_T(\phi_T \phi_T)^j, & \tilde{a}_T &= \sum_{j=0}^\infty \psi_T(\phi_T \phi_T)^j \odot \tilde{\kappa}_{2(j+1)}, \\
b_T &= \gamma \sum_{j=0}^\infty \kappa_{2(j+1)} \phi_T(\psi_T \phi_T)^j, & \tilde{b}_T &= \gamma \sum_{j=0}^\infty \phi_T(\psi_T \phi_T)^j \odot \tilde{\kappa}_{2(j+1)}.
\end{align*}$$

Indexing blocks by $\{0, \ldots, T\}$ and writing $[t, s]$ to denote the $K \times K$ submatrix corresponding to row block $t$ and column block $s$, we set the debiasing coefficients of (4.12) up to iteration $T$ as

$$\begin{align*}
a_{ts} &= \begin{cases} 
\tilde{a}_T[t, s] & \text{if } s = 0, \\
a_T[t, s] & \text{otherwise.}
\end{cases} \\
b_{ts} &= \begin{cases} 
\tilde{b}_T[t, s] & \text{if } s = 0, \\
b_T[t, s] & \text{otherwise.}
\end{cases}
\end{align*}$$

(4.14)
**State Evolution.** The state of this AMP algorithm is described in terms of two recursively defined sequences of mean transformations

\[
\mu_T = \begin{pmatrix} \mu_0 \\
\vdots \\
\mu_T \end{pmatrix} \in \mathbb{R}^{(T+1)K \times K'}, \quad \nu_T = \begin{pmatrix} \nu_0 \\
\vdots \\
\nu_T \end{pmatrix} \in \mathbb{R}^{(T+1)K \times K'},
\]

and covariance matrices \( \Sigma_T = \{\sigma_{st}\}_{0 \leq s, t \leq T} \) and \( \Omega_T = \{\omega_{st}\}_{0 \leq s, t \leq T} \).

Let \( U'_s, V'_s \in \mathbb{R}^{K'} \) be random vectors satisfying \( \mathbb{E}[U'_sU'_s^\top] = \mathbb{E}[V'_sV'_s^\top] = \text{Id} \), representing the limit empirical distributions of rows of \( U'_s, V'_s \). Define \( \mu_{\text{pca}} = \text{diag}(\mu_{\text{pca},1}, \ldots, \mu_{\text{pca},K}) \in \mathbb{R}^{K \times K'} \) and \( \nu_{\text{pca}} = \text{diag}(\nu_{\text{pca},1}, \ldots, \nu_{\text{pca},K}) \in \mathbb{R}^{K \times K'} \), where the last \( K' - K \) columns are 0. We initialize

\[
\mu_0 = \mu_0 = \mu_{\text{pca}}, \quad \Sigma_0 = \sigma_{00} = \text{Id} - \mu_{\text{pca}}\mu_{\text{pca}}^\top,
\]

\[
\nu_1 = \begin{pmatrix} \nu_0 \\
\vdots \\
\nu_1 \end{pmatrix} = \begin{pmatrix} \nu_{\text{pca}} \\
\vdots \\
\nu_{\text{pca}} \end{pmatrix}, \quad \Omega_1 = \begin{pmatrix} \omega_{00} & \omega_{01} \\
\omega_{10} & \omega_{11} \end{pmatrix} = \begin{pmatrix} \text{Id} - \nu_{\text{pca}}\nu_{\text{pca}}^\top & \text{Id} - \nu_{\text{pca}}\nu_{\text{pca}}^\top \\
\text{Id} - \nu_{\text{pca}}\nu_{\text{pca}}^\top & \text{Id} - \nu_{\text{pca}}\nu_{\text{pca}}^\top \end{pmatrix}
\]

corresponding to \( F_0 \) and \( (G_0, G_1) \) in (4.11). For each \( t \geq 1 \), having defined \( (\mu_{t-1}, \Sigma_{t-1}, \nu_t, \Omega_t) \), the next state \( (\mu_t, \Sigma_t, \nu_{t+1}, \Omega_{t+1}) \) is constructed as follows:

To define \( (\mu_t, \Sigma_t) \), first define joint laws for random vectors \( (U'_s, U_0, \ldots, U_t, F_0, \ldots, F_{t-1}) \) and \( (V'_s, V_0, \ldots, V_t, G_0, \ldots, G_t) \) by

\[
\begin{align*}
(F_0, \ldots, F_{t-1}) | U'_s & \sim \mathcal{N}(\mu_{t-1} \cdot U'_s, \Sigma_{t-1}) \\
U_0 &= U_1 = S_u^{-1}F_0 \quad U_s = u_s(F_0, \ldots, F_{s-1}) \text{ for } s = 2, \ldots, t, \\
(G_0, \ldots, G_t) | V'_s & \sim \mathcal{N}(\nu_t \cdot V'_s, \Omega_t) \\
V_0 &= S_v^{-1}G_0 \quad V_s = v_s(G_1, \ldots, G_s) \text{ for } s = 1, \ldots, t.
\end{align*}
\]

Then define \( \mu_t \) to have the blocks

\[
\mu_s = \mathbb{E}[V'_sV'_s^\top] : \sqrt{\gamma} \text{ for each } s = 0, \ldots, t.
\]

For \( s = 0 \), it may be checked from (4.9) and (4.10) that \( \theta_{u,k}/\theta_{v,k} = \gamma(\mu_{\text{pca},k}/\nu_{\text{pca},k}) \), and hence that this coincides with the above initialization \( \mu_{\text{pca}} \). Next, decompose the second moment matrix of \( (U_0, \ldots, U_t) \) into four parts in the same way as (3.15),

\[
\Delta^\top_t \overset{\text{def}}{=} \begin{pmatrix} 0 & 0 & \cdots & 0 \\
0 & \mathbb{E}[U_0U_0^\top] & \cdots & \mathbb{E}[U_0V_0^\top] \\
\vdots & \vdots & \ddots & \vdots \\
0 & \mathbb{E}[U_tU_0^\top] & \cdots & \mathbb{E}[U_tV_0^\top] \end{pmatrix}
+ \begin{pmatrix} 0 & \mathbb{E}[U_0U_0^\top] & \cdots & \mathbb{E}[U_0V_0^\top] \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 \end{pmatrix}
+ \begin{pmatrix} \mathbb{E}[U_0U_0^\top] & 0 & \cdots & 0 \\
\mathbb{E}[U_0U_0^\top] & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\mathbb{E}[U_0U_0^\top] & 0 & \cdots & 0 \end{pmatrix} + \begin{pmatrix} \mathbb{E}[U_0U_0^\top] & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 \end{pmatrix}.
\]

Decompose the second moment matrix for \( (V_0, \ldots, V_t) \) analogously as \( \Gamma_t \overset{\text{def}}{=} \tilde{\Gamma}_t + \bar{\Gamma}_t + \tilde{\Gamma}_t^\top + \tilde{\Gamma}_t \). Set

\[
\Delta^{(j)}_t = \kappa_2(j+1) \Delta_t + \bar{k}_2(j+1) \otimes \tilde{\Delta}_t + \tilde{\Delta}_t^\top \otimes \bar{k}_2(j+1) + \bar{k}_2(j+1) \otimes \tilde{\Delta}_t,
\]

\[
\Gamma^{(j)}_t = \kappa_2(j+1) \Gamma_t + \bar{k}_2(j+1) \otimes \tilde{\Gamma}_t + \tilde{\Gamma}_t^\top \otimes \bar{k}_2(j+1) + \bar{k}_2(j+1) \otimes \tilde{\Gamma}_t.
\]
Define the large-\( n \) limits of \( \phi_t, \psi_t \) as

\[
\Phi_t = \left( \mathbb{E}[\partial_s u_{r}(F_0, \ldots, F_{r-1})] \right)_{r,s \in \{0, \ldots, t\}}, \quad \Psi_t = \left( \mathbb{E}[\partial_s v_{r}(G_1, \ldots, G_r)] \right)_{r,s \in \{0, \ldots, t\}}
\]

again with the identifications \( \partial_0 u_1 = S_u^{-1}, \partial_0 v_0 = S_v^{-1}, \) and \( \partial_0 v_t = 0 \) for \( t \geq 1 \). Then, recalling \( \Xi^{(j)}[\cdot, \cdot, \cdot, \cdot] \) from (2.14), define

\[
\Sigma_t = \sum_{j=0}^{\infty} \Xi^{(j)}[\Phi_t, \Psi_t, \Delta_t^{(j)}, \Gamma_t^{(j)}].
\] (4.17)

Next, to define \( (\nu_{t+1}, \Omega_{t+1}) \), first define from \( (\mu_t, \Sigma_t, \nu_t, \Omega_t) \) the joint laws for random vectors \( (U'_s, U_0, \ldots, U_{t+1}, F_0, \ldots, F_t) \) and \( (V'_s, V_0, \ldots, V_t, G_0, \ldots, G_t) \) according to (4.15). Then define \( \nu_{t+1} \) to have the blocks \( \nu_s = \mathbb{E}[U'_s U'_s^\top] \cdot S'_\sqrt{\gamma} \) for \( s = 0, \ldots, t + 1 \). For \( s = 0 \), this coincides with the initialization \( \nu_{\text{pca}} \). Then, extending the above definitions of \( \Phi_t, \Psi_t, \Delta_t^{(j)}, \Gamma_t^{(j)} \) from \( t \) to \( t + 1 \) and recalling \( \Theta^{(j)}[\cdot, \cdot, \cdot, \cdot] \) from (2.12), define

\[
\Omega_{t+1} = \sum_{j=0}^{\infty} \Theta^{(j)}[\Phi_{t+1}, \Psi_{t+1}, \Delta_{t+1}^{(j)}, \Gamma_{t+1}^{(j)}].
\]

The last row block of \( \Psi_{t+1} \) and the last row and column block of \( \Gamma_{t+1}^{(j)} \) are undefined, as we have not yet defined \( (V_{t+1}, G_{t+1}) \). As in Section 2.2, we permit this ambiguity because the last column block of \( \Phi_{t+1} \) is 0, so the matrix products defining \( \Theta^{(j)} \) do not depend on these undefined blocks. It may be checked from these recursive definitions that the first \( t \) blocks of \( \mu_t \) coincide with \( \mu_{t-1} \), and the upper-left \( t \times t \) blocks of \( \Sigma_t \) coincide with \( \Sigma_{t-1} \). The same holds for \( \nu_{t+1} \) and \( \Omega_{t+1} \).

Our main result shows that this state evolution provides a rigorous characterization of the AMP algorithm (4.11–4.12) with spectral initialization, under the following assumptions.

**Assumption 4.2.**

(a) \( U'_s = (u_1^s, \ldots, u_K^s) \) and \( V'_s = (v_1^s, \ldots, v_K^s) \) are independent of \( O \) and \( Q \), satisfy (4.2), and \( U'_s \xrightarrow{W_2} U'_r, \ V'_s \xrightarrow{W_2} V'_r \) a.s. as \( m, n \rightarrow \infty \) where \( \mathbb{E}[U'_s U'_s^\top] = \mathbb{E}[V'_s V'_s^\top] = \text{Id} \).

(b) Each \( u_{t+1}(\cdot) \) is Lipschitz in all arguments. For each \( s = 0, \ldots, t \) and all \( (\mu, \Sigma) \) in a sufficiently small open neighborhood of \( (\mu_t, \Sigma_t) \) defined by (4.16) and (4.17), \( \partial_s u_{t+1}(F_0, \ldots, F_t) \) exists and is continuous on a set of probability 1 under the marginal law of \( (F_0, \ldots, F_t) \) defined by \( (F_0, \ldots, F_t) \mid U'_s \sim \mathcal{N}(\mu \cdot U'_r, \Sigma) \). The same holds for each \( v_t(\cdot) \) with respect to \( (\nu_t, \Omega_t) \) and \( (G_1, \ldots, G_t) \).

(c) The values \( \theta_1, \ldots, \theta_K \) are distinct. For each \( k \in \{1, \ldots, K\} \), \( \theta_k \geq G(1/\lambda_+ > 0 \) and there exists some constant \( \lambda \in (0, 1) \) such that

\[
\frac{\lambda_+ \sqrt{\gamma}}{\theta_{u,k}} + \gamma \sum_{j=1}^{\infty} \frac{\kappa_{2j}}{\theta^2_{k} 2j-1} < 1 \quad \text{and} \quad \frac{\lambda_+}{\theta_{v,k} \sqrt{\gamma}} + \sum_{j=1}^{\infty} \frac{\kappa_{2j}}{\theta^2_{k} 2j-1} < 1.
\]

Assumption 4.2(c) requires the signal singular values \( \theta_1, \ldots, \theta_K \) for the first \( K \) selected signals to exceed a constant depending only \( \gamma = m/n \) and the law of \( \Lambda \). In particular, these signal values are super-critical in the sense of Theorem 2.6, the series (4.8) for \( R(z) \) and \( R'(z) \) are absolutely convergent at \( z = 1/\theta^2_{k} \), and the series (4.13) defining \( \kappa_{2s}, \kappa_{2s} \) are also absolutely convergent.
Theorem 4.3. Consider the rectangular spiked model (4.1), where Assumptions 2.4 and 4.2 hold. For any $T \geq 1$, consider the spectrally initialized AMP algorithm (4.11–4.12) up to iteration $T$, and define $(U'_s, U_0, \ldots, U_{T+1}, F_0, \ldots, F_T)$ and $(V'_s, V_0, \ldots, V_T, G_0, \ldots, G_T)$ by (4.15). Then almost surely as $m, n \to \infty$,

$$
(U'_s, U_0, \ldots, U_{T+1}, F_0, \ldots, F_T) \xrightarrow{W_2} (U'_s, U_0, \ldots, U_{T+1}, F_0, \ldots, F_T),
$$

$$
(V'_s, V_0, \ldots, V_T, G_0, \ldots, G_T) \xrightarrow{W_2} (V'_s, V_0, \ldots, V_T, G_0, \ldots, G_T).
$$

Remark 4.4. As in Remark 2.7, we have defined $b_{ts}, a_{ts}$ in (3.12) and (4.14) using the free cumulants $\{\kappa_j\}$ of the limit spectral distributions, as well as the true signal values $S = (\theta_1, \ldots, \theta_K)$. Theorems 3.3 and 4.3 then also hold when $b_{ts}, a_{ts}$ are replaced by $b'_{ts}, a'_{ts}$ such that $\|b_{ts} - b'_{ts}\| \to 0$ and $\|a_{ts} - a'_{ts}\| \to 0$ a.s., and in particular if $\{\kappa_j\}, \{\tilde{\kappa}_j\}$, and $\{\hat{\kappa}_j\}$ are replaced by consistent estimates of these quantities.

5 Orthogonal AMP for Bayesian PCA

Finally, we discuss in this section an application to estimating the signal vectors $u^k_s$ and $v^k_s$ in the preceding signal-plus-noise models.

The distributions of $U'_s, V'_s \in \mathbb{R}^{K'}$ for the row-wise limits of $U'_s, V'_s$ may be interpreted as Bayesian “priors” for these rows. Assuming these prior distributions are known, we describe a Bayes-OAMP method in Section 5.1 that uses Bayes posterior-mean denoisers as the non-linearities in the preceding AMP algorithms. We suggest ways of estimating the Onsager debiasing coefficients and state evolution parameters in Section 5.2, and illustrate the method in simulation in Section 5.3. In practice, the distributions of $U'_s, V'_s$ are typically also unknown, but may be estimated from the data $X$ using empirical Bayes ideas. We will not consider this additional complexity here, but refer readers to [ZSF20] for an example of this approach.

Let us describe the Bayes-OAMP method in a setting where the AMP dimension $K'$ and signal rank $K'$ satisfy $K \leq K'$, reflecting applications where $K'$ may be large, and one may wish to estimate smaller subsets of dependent PCs at a time. In this setting, we consider the following additional assumption for the laws of $U'_s$ and $V'_s$.

Assumption 5.1. The last $K' - K$ coordinates of $U'_s$ have mean 0, and are independent of the first $K$ coordinates. For the rectangular model, the same holds for $V'_s$.

In applications, the data $X$ is often centered to have row and column means approximately 0 before applying PCA, leading to PCs also having mean approximately 0. The components $u^k_s$ and $v^k_s$ should ideally be grouped into small subsets of dependent signals, with the signals within each subset estimated together to maximally leverage their joint structure. The above assumption of exact independence of one such subset $(1, \ldots, K)$ from the remaining signals $(K + 1, \ldots, K')$ is a modeling approximation for this grouping.

Assumption 5.1 ensures that when AMP is initialized with the first $K$ signal components, its state evolution will not depend on the remaining $K' - K$ components: In the symmetric setting of Section 3, let $U_s \in \mathbb{R}^K$ be the first $K$ coordinates of $U'_s$. Then it is inductively verified using Assumption 5.1 that each mean transformation $\mu_t \in \mathbb{R}^{(t+1)K \times K'}$ has last $K' - K$ columns equal to 0, and the law of each vector $F_t$ and $U_t$ depends on $U'_s$ only via $U_s$. We may then write the conditional law of $(F_0, \ldots, F_t)$ more simply as

$$(F_0, \ldots, F_t) \ | \ U'_s \sim \mathcal{N}(\mu_t \cdot U_s, \Sigma_t)$$

19
where, with slight abuse of notation, we write \( \nu_t \in \mathbb{R}^{(t+1)K \times K} \) for its first \( K \) columns. Similarly, in the rectangular setting, let us write

\[
(F_0, \ldots, F_t) \mid U_s \sim \mathcal{N}(\mu_t \cdot U_s, \Sigma_t), \quad (G_0, \ldots, G_t) \mid V_s \sim \mathcal{N}(\nu_t \cdot V_s, \Omega_t)
\]

where \( U_s, V_s \) are the first \( K \) coordinates of \( U_s', V_s' \), and \( \mu_t, \nu_t \in \mathbb{R}^{(t+1)K \times K} \).

### 5.1 Bayes-OAMP

In the symmetric spiked model (3.1), under Assumption 5.1, we consider an AMP algorithm which estimates the first \( K \) components \( U_s \in \mathbb{R}^{n \times K} \) of \( U_s' \), using as its non-linearities the posterior-mean denoising functions

\[
u_{t+1}(f_0, \ldots, f_t) = \mathbb{E}[U_s \mid (F_0, \ldots, F_t) = (f_0, \ldots, f_t)]
\]

computed under the above conditional law \( (F_0, \ldots, F_t) \mid U_s \sim \mathcal{N}(\mu_t \cdot U_s, \Sigma_t) \). Explicitly, writing as shorthand \( f_t = (f_0, \ldots, f_t) \in \mathbb{R}^{(t+1)K} \),

\[
u_{t+1}(f_t) = \frac{\mathbb{E}[U_s \exp(-\frac{1}{2}(f_t - \mu_t \cdot U_s)^T \Sigma_t^{-1}(f_t - \mu_t \cdot U_s))]}{\mathbb{E}[\exp(-\frac{1}{2}(f_t - \mu_t \cdot U_s)^T \Sigma_t^{-1}(f_t - \mu_t \cdot U_s))]} \in \mathbb{R}^K.
\]

We will refer to the AMP algorithm (3.8–3.10) using this choice of non-linearity as Bayes-OAMP.

If this posterior mean function \( u_{t+1}(-) \) is Lipschitz (which holds e.g. if \( U_s \) has bounded support or log-concave density) and the signal strengths \( |\theta_1|, \ldots, |\theta_K| \) are sufficiently large, then Theorem 3.3 implies that the mean-squared-error risk of the estimate \( U_t \) has the asymptotic limit

\[
\lim_{n \to \infty} \frac{1}{n} \|U_t - U_s\|_F^2 \overset{\text{def}}{=} \text{MSE}(U_t) = \mathbb{E}[\|U_s - \mathbb{E}[U_s \mid F_0, \ldots, F_t]\|^2] = \text{Tr Cov}[U_s \mid F_0, \ldots, F_t].
\]

Thus these errors satisfy

1. \( \text{MSE}(U_{t+1}) \leq \text{MSE}(U_t) \), by the property \( \text{Cov}[U_s \mid F_0, \ldots, F_{t+1}] \preceq \text{Cov}[U_s \mid F_0, \ldots, F_t] \). This implies also by the martingale property of \( U_1, U_2, \ldots \) that the algorithm is convergent in the sense

\[
\lim_{n \to \infty} \frac{1}{n} \|U_{t+1} - U_t\|_F^2 = \mathbb{E}[\|U_{t+1} - U_t\|^2] = \text{MSE}(U_t) - \text{MSE}(U_{t+1}) \overset{t \to \infty}{\to} 0.
\]

2. \( \text{MSE}(U_t) \leq \text{MSE}(F_{\text{pca}}) \) for any iterate \( t \geq 1 \), because

\[
\text{MSE}(F_{\text{pca}}) = \text{MSE}(F_0) = \mathbb{E}[\|U_s - F_0\|_2^2] \geq \mathbb{E}[\|U_s - \mathbb{E}[U_s \mid F_0, \ldots, F_t]\|_2^2] = \text{MSE}(U_t).
\]

**Remark 5.2.** This algorithm differs from the Bayes-AMP algorithms of [RF12, MV21c] for Gaussian noise and the single-iterate Bayes-OAMP algorithm that was studied in [Fan20, Section 3], which use instead \( u_{t+1}(f_0, \ldots, f_t) = \mathbb{E}[U_s \mid F_t = f_t] \) based only on the single preceding iterate \( F_t \).

When \( W \) is symmetric Gaussian noise and \( \Lambda \) is distributed as Wigner’s semicircle law, these two approaches coincide: Indeed, in this case \( \kappa_2 = 1 \) and \( \kappa_j = 0 \) for all \( j \geq 3 \), so (3.16) reduces to \( \Sigma_t = \Delta_t = \mathbb{E}[UU^T] \) where \( U = (U_0, \ldots, U_t) \), coinciding with the state evolution shown in [MV21c]. For any \( t \geq 1 \) and \( 0 \leq s \leq t \), from the martingale identity

\[
\mathbb{E}[U_sU_s^T] = \mathbb{E}[U_s\mathbb{E}[U_s^T \mid F_0, \ldots, F_t]] = \mathbb{E}[\mathbb{E}[U_sU_s^T \mid F_0, \ldots, F_t]] = \mathbb{E}[U_sU_s^T]
\]
and the definition of $\mu_t$ in (3.14), we observe that $\Sigma_t(e_t \odot S) = \mu_t$ and $\sigma_{tt}S = \mu_t$, where $S = \text{diag}(\theta_1, \ldots, \theta_K) \in \mathbb{R}^{K \times K}$, $e_t \odot S \in \mathbb{R}^{(t+1)K \times K}$ has first $t$ row blocks equal to 0 and last row block equal to $S$, and $\sigma_{tt}, \mu_t \in \mathbb{R}^{K \times K}$ denote the last blocks of $\Sigma_t$ and $\mu_t$. Then $\Sigma_t^{-1}\mu_t = e_t \odot S$ and $\sigma_{tt}^{-1}\mu_t = S$. So (5.1) is equivalent to the single-iterate posterior mean,

$$u_{t+1}(f_t) = \frac{\mathbb{E}[U_* \exp(f_t^\top SU_* - \frac{1}{2}U_*^\top \mu_t^\top SU_*)]}{\mathbb{E}[\exp(f_t^\top SU_* - \frac{1}{2}U_*^\top \mu_t^\top SU_*)]} = \frac{\mathbb{E}[U_* \exp(-\frac{1}{2}(f_t - \mu_tU_*)^\top \sigma_{tt}^{-1}(f_t - \mu_tU_*))]}{\mathbb{E}[\exp(-\frac{1}{2}(f_t - \mu_tU_*)^\top \sigma_{tt}^{-1}(f_t - \mu_tU_*))] = \mathbb{E}[U_* \mid F_t = f_t].$$

Outside of this setting where $\Lambda$ has Wigner semicircle law, these two approaches are different. The state evolution of Bayes-OAMP is more involved, and we will not pursue a theoretical characterization of its fixed point in this work, as was done for single-iterate Bayes-OAMP in [Fan20]. We observe in simulation that the above monotonicity property of $\text{MSE}(U_t)$ need not hold for single-iterate Bayes-OAMP in settings of small signal strength, as was observed also in [MV21b], and that Bayes-OAMP can substantially improve over the single-iterate approach in such settings.

For the rectangular model (4.1), we consider analogously the Bayes-OAMP algorithm which estimates the first $K$ components $U_* \in \mathbb{R}^{n \times K}$ and $V_* \in \mathbb{R}^{n \times K}$ of $U'$ and $V'$, using in (4.11–4.12) the denoisers

$$v_t(g_1, \ldots, g_t) = \mathbb{E}[V_* \mid (G_1, \ldots, G_t) = (g_1, \ldots, g_t)],$$

$$u_{t+1}(f_0, \ldots, f_t) = \mathbb{E}[U_* \mid (F_0, \ldots, F_t) = (f_0, \ldots, f_t)].$$

Recall that $G_1 = G_{\text{pca}}$ and $F_0 = F_{\text{pca}}$, so these posterior means are conditional also on the spectral initializations. As in the symmetric setting, the asymptotic mean-squared-errors satisfy

$$\text{MSE}(V_{t+1}) \leq \text{MSE}(V_t) \leq \text{MSE}(G_{\text{pca}}), \quad \text{MSE}(U_{t+1}) \leq \text{MSE}(U_t) \leq \text{MSE}(F_{\text{pca}}).$$

### 5.2 Estimating the debiasing corrections and state evolution

Numerical implementations of the Bayes-OAMP algorithms require estimating the debiasing coefficients and state evolution parameters that describe the conditional laws of $(F_0, \ldots, F_t)$ and $(G_0, \ldots, G_t)$. We describe here one approach for this estimation.

**Symmetric matrices.** We estimate the law of $\Lambda$ by the observed empirical eigenvalue distribution of $X$, with largest $K_+$ positive eigenvalues and smallest $K_-$ negative eigenvalues removed. This estimate is weakly consistent as $n \to \infty$ by Weyl’s eigenvalue interlacing inequality. We compute the empirical moments of this law and estimate the free cumulants $\{\kappa_s\}$ through the non-crossing moment-cumulant relations, see e.g. [Fan20, Section 2.3]. For each signal value $\theta_k$, based on (3.5) and (3.7), we estimate

$$\theta_k \text{ by } \frac{1}{G(\lambda_{\text{pca},k})}, \quad \mu_{\text{pca},k}^2 \text{ by } \frac{-1}{\theta_k^2 G'(\lambda_{\text{pca},k})}, \quad R(1/\theta_k) \text{ by } \lambda_{\text{pca},k} - \theta_k, \quad R'(1/\theta_k) \text{ by } \theta_k^2(1 - \mu_{\text{pca},k}^2)$$

where $\lambda_{\text{pca},k}$ is the observed eigenvalue of $X$, and the integrals defining $G(\cdot)$ and $G'(\cdot)$ are computed using the above empirical estimate for the law of $\Lambda$. In particular, this provides the estimate of $S = \text{diag}(\theta_1, \ldots, \theta_K)$. Comparing (3.6) and (3.11), we then estimate

$$\bar{\kappa}_1 \text{ by } \text{diag}(R(1/\theta_1), \ldots, R(1/\theta_K)), \quad \bar{\kappa}_2 \text{ by } \text{diag}(R'(1/\theta_1), \ldots, R'(1/\theta_K)),$$
and estimate the remaining matrices $\tilde{\kappa}_s$ and $\hat{\kappa}_s$ using the following recursions derived from (3.11),

$$\tilde{\kappa}_s = \sum_{j=0}^{\infty} \kappa_{s+j} S^{-j} = \kappa_s I + \sum_{j=0}^{\infty} \kappa_{s+1+j} S^{-j} \cdot S^{-1} = \kappa_s I + \tilde{\kappa}_{s+1} S^{-1} \Rightarrow \tilde{\kappa}_{s+1} = (\tilde{\kappa}_s - \kappa_s I) S,$$

$$\hat{\kappa}_s = \sum_{j=0}^{\infty} (j+1) \kappa_{s+j} S^{-j} = \hat{\kappa}_s + \sum_{j=0}^{\infty} (j+1) \cdot \kappa_{s+1+j} S^{-j} \cdot S^{-1} = \hat{\kappa}_s + \hat{\kappa}_{s+1} S^{-1} \Rightarrow \hat{\kappa}_{s+1} = (\hat{\kappa}_s - \hat{\kappa}_s) S.$$

Explicitly differentiating (5.1) to compute the matrices $\langle \partial_s U_t \rangle$ constituting $\Phi_t$, and combining with the above, we obtain consistent estimates of the debiasing coefficients $b_{ts}$. For the state evolution, we note that numerically evaluating the expectations defining $(\mu_t, \Sigma_t)$ may be prohibitive. We instead approximate the second-moment matrices $\mathbb{E}[U_s U_t^\top]$ by the empirical averages $n^{-1} U_s^\top U_t$. We approximate $\Phi_t$ also by the empirical average $\hat{\mu}_t$, and combine with the above estimates of $\kappa_s, \tilde{\kappa}_s, \hat{\kappa}_s$ according to the definition (3.16) to obtain a consistent estimate of $\Sigma_t$. We use the martingale identity $\mathbb{E}[U_t U_t^\top] = \mathbb{E}[U_t U_t^\top]$ for $t \geq 1$ to approximate $\mathbb{E}[U_t U_s^\top]$ by the empirical average $n^{-1} U_s^\top U_t$, and use this to consistently estimate $\mu_t$.

We remark that when the AMP algorithm approaches convergence, the covariance matrix $\Sigma_t$ becomes nearly singular, leading to potential instabilities in evaluating the posterior mean function (5.1) and its Jacobian. We use a simple early stopping rule of terminating the iterations when the smallest eigenvalue of our estimate for $\Sigma_t$ falls below a small threshold. Alternatively, a small ridge regularization may be used when computing $\Sigma_t^{-1}$.

**Rectangular matrices.** Our approach is analogous. We estimate the law of $\Lambda$ by the empirical singular value distribution of $X$, with leading $K$ singular values removed. We then compute the empirical moments, and estimate the rectangular free cumulants $\{\kappa_{2s}\}$ via the moment-cumulant relations, see e.g., [Fan20, Section 2.4]. Using this estimated law of $\Lambda$ to compute $\varphi(\cdot), \bar{\varphi}(\cdot), D(\cdot)$, and $\hat{D}(\cdot)$ in (4.5), we estimate $\theta_s^2$ by $1/\sqrt{D(\lambda_{pca,k})}$ and $D^{-1}(1/\theta_k)^2$ by $\lambda_{pca,k}$ based on (4.6). We then estimate $\mu^2_{pca,k}, \nu^2_{pca,k}$ using (4.6), $R(\theta_k^{-2})$ using (4.7), $R'(\theta_K^{-2})$ using (4.10) and $\theta_{a,k}, \theta_{v,k}$ using (4.9). Applying (4.13) and (4.8), we estimate $\tilde{\kappa}_2$ by $\text{diag}(\theta_1^2 R(\theta_1^{-2}), \ldots, \theta_K^2 R(\theta_K^{-2}))$ and $\hat{\kappa}_2$ by $\text{diag}(R'(\theta_1^{-2}), \ldots, R'(\theta_K^{-2}))$, and the remaining $\tilde{\kappa}_{2s}$ and $\hat{\kappa}_{2s}$ based on the recursions

$$\tilde{\kappa}_{2s} = \sum_{j=0}^{\infty} \kappa_{2(s+j)} S^{-2j} = \kappa_{2s} + \sum_{j=0}^{\infty} \kappa_{2(s+1+j)} S^{-2j} \cdot S^{-2} = \kappa_{2s} + \tilde{\kappa}_{2(s+1)} S^{-2},$$

$$\hat{\kappa}_{2s} = \sum_{j=0}^{\infty} (j+1) \kappa_{2(s+j)} S^{-2j} = \hat{\kappa}_{2s} + \sum_{j=0}^{\infty} (j+1) \cdot \kappa_{2(s+1+j)} S^{-2j} \cdot S^{-2} = \hat{\kappa}_{2s} + \hat{\kappa}_{2(s+1)} S^{-2}.$$

This yields consistent estimates of the debiasing coefficients $b_{ts}, a_{ts}$. The state evolution parameters $(\mu_t, \Sigma_t, \nu_t, \Omega_t)$ are then estimated using $\Phi_t, \psi_t$ and the empirical second-moment matrices $m^{-1} U_s^\top U_t$ and $n^{-1} V_s^\top V_t$ as in the symmetric setting.

**5.3 Simulation studies**

We compare the estimation accuracy of three AMP algorithms under various noise spectral distributions for $W$:

1. **Bayes-OAMP**, as described and implemented in Sections 5.1 and 5.2 above.
Figure 1: Estimation errors for AMP iterates $\mathbf{U}_t$ in the symmetric spiked model with $n = 4000$, rank-2 signal, and signal prior $U_* \sim \frac{1}{2} \delta_{(0,1)} + \frac{1}{4} \delta_{(\sqrt{2},-1)} + \frac{1}{4} \delta_{(-\sqrt{2},-1)}$. Boxes indicate the $\{25, 50, 75\}$-percentiles across 50 random trials, and whiskers indicate $1.5 \times$ inter-quartile range. Iteration 0 corresponds to the spectral initialization $\mathbf{U}_0$. The noise spectral distributions are (left) the semicircle law, (middle) Uniform$[-\sqrt{3}, \sqrt{3}]$, and (right) standarized Beta$$(3, 1)$.

2. **Single-iterate Bayes-OAMP**, using instead the single-iterate posterior mean non-linearities $u_{t+1}(f_0, \ldots, f_t) = \mathbb{E}[U_* | F_t = f_t]$ and $v_t(g_1, \ldots, g_t) = \mathbb{E}[V_* | G_t = g_t]$, with debiasing coefficients and state evolution parameters estimated as in Section 5.2.

3. **Gaussian Bayes-AMP**, using the debiasing coefficients and state evolution described in [MV21c] for i.i.d. noise. We also estimate the signal strengths $\theta_k$ and initial spectral alignments $\mu_{\text{pca},k}^2, \nu_{\text{pca},k}^2$ from (3.5) and (4.6) using functions $G(\cdot)$ and $D(\cdot)$ that correspond to the semicircle and Marcenko-Pastur laws for i.i.d. noise.

**Symmetric model.** We consider the symmetric model (3.1) with $n = 4000$, in the settings

- (Semicircle) $\mathbf{W} \sim \text{GOE}(n)$ has i.i.d. $\mathcal{N}(0, 1/n)$ entries above the diagonal, and the limit spectral law $\Lambda$ is the semicircle distribution supported on $[-2, 2]$.

- (Uniform) $\mathbf{W} = \mathbf{O}^\top \mathbf{A} \mathbf{O}$ where $\mathbf{A}$ has i.i.d. Uniform$[-\sqrt{3}, \sqrt{3}]$ diagonal entries, and $\mathbf{O}$ is uniformly random.

- (Centered Beta) $\mathbf{W} = \mathbf{O}^\top \mathbf{A} \mathbf{O}$ where $\mathbf{A}$ has i.i.d. $\sqrt{80/3} \cdot \text{Beta}(3, 1 - 3/4)$ diagonal entries, and $\mathbf{O}$ is uniformly random.

In all three settings, $\Lambda$ is normalized to have mean $\kappa_1 = 0$ and variance $\kappa_2 = 1$. Figure 1 compares estimation error across AMP iterations, for an example where $K' = K = 2$, the two signal strengths are $(\theta_1, \theta_2) = (2, 1.6)$, and the distribution for $U_*$ is the discrete three-point prior

$$\frac{1}{2} \delta_{(0,1)} + \frac{1}{4} \delta_{(\sqrt{2},-1)} + \frac{1}{4} \delta_{(-\sqrt{2},-1)}$$

(5.2)

The error is defined by the subspace distance $\|\Pi_{\mathbf{U}_*} - \Pi_{\mathbf{U}_t}\|$ where $\Pi_{\mathbf{U}} \in \mathbb{R}^{n \times n}$ is the orthogonal projector onto the two-dimensional column span of $\mathbf{U}$. From the left panel, for GOE noise, we
Figure 2: Distributions of iterates $F_0, F_1, F_2$ in the Centered Beta noise example of Figure 1. Shown are histograms of the empirical distributions of the two columns of each iterate $F_t$ (denoted PC1 and PC2), overlaid with the marginal density of the corresponding coordinate of the state evolution law $\mathcal{N}(\mu_t \cdot U_s, \Sigma_t)$. This density agrees closely for Bayes-OAMP, whereas a large discrepancy is observed for the state evolution prediction of Gaussian Bayes-AMP.

Observe that the three algorithms yield comparable per-iteration error, and there is negligible degradation in accuracy from estimating the spectral free cumulants and the more complex state parameters in Bayes-OAMP. In the middle panel, for Uniform noise spectrum, Gaussian Bayes-AMP remains reasonably robust to the misspecification of the noise distribution. In the right panel, for Centered Beta noise spectrum, we observe a significant improvement of Bayes-OAMP over the other two approaches, which both exhibit non-monotonic error across iterations.

Figure 2 depicts the accuracy of the state evolution predictions for the distribution of AMP iterates $F_t$ in the Centered Beta example. Marginal histograms of the two columns of $F_t$ are overlaid with the corresponding densities of $F_t$, computed using state evolution parameters $(\mu_t, \Sigma_t)$ empirically estimated as in Section 5.2. Note that these densities are exactly the ones assumed for posterior-mean denoising in (5.1), to obtain the next iterate $U_{t+1}$. We observe a close agreement for Bayes-OAMP, and a large discrepancy with the state evolution predictions of Gaussian Bayes-AMP.

Figure 3 illustrates estimation error across different signal strengths, in an example having rank-1 signal and the same Centered Beta noise. Note that since the density of Beta(3, 1) does not decay to 0 at its right edge, the spectral phase transition point for this example is 0, and any positive signal strength is super-critical. We observe that Bayes-OAMP remains effective and improves over the spectral initialization for any positive signal strength. In contrast, single-iterate Bayes-OAMP exhibits the following behavior: below some critical signal strength, it diverges from the informative spectral initialization to an uninformative solution. This type of behavior was also
Figure 3: Estimation errors for the sample eigenvector and three AMP algorithms, in a symmetric model with Centered Beta noise spectrum, \( n = 4000 \), rank-1 signal with prior \( U_* \sim \frac{1}{2} \delta_{1} + \frac{1}{2} \delta_{-1} \), and signal strength \( \theta \) varying from 0.1 to 2.0. The spectral “phase transition” occurs at \( \theta = 0 \).

reported in [MV21b, Section 4] for a different prior. Our implementation of Gaussian Bayes-AMP uses the Cauchy-transform \( G(\cdot) \) based on the semicircle law to infer \( \theta \), and thus is only applicable when the largest sample eigenvalue exceeds 2, corresponding roughly to \( \theta \geq 1.6 \) in this example. For these values of \( \theta \), we observe its accuracy to be close to that of single-iterate Bayes-OAMP.

Rectangular model. We consider the rectangular model (4.1) with \((m,n) = (3000,4000)\), again with \( K' = K = 2 \), signal strengths \((\theta_1, \theta_2) = (2, 1.5)\), and the discrete three-point prior (5.2) for both \( U_* \) and \( V_* \). We consider the settings

- (Marcenko-Pastur) \( W \) has i.i.d. \( \mathcal{N}(0,1/n) \) entries, and the limit singular value distribution \( \Lambda \) is the square-root of a Marcenko-Pastur law.

- (Uniform) \( W = O^\top \Lambda Q \) where \( \Lambda \) has i.i.d. Uniform\([\sqrt{3/7}, 2\sqrt{3/7}]\) diagonal entries, and \((O, Q)\) are uniformly random.

- (Beta) \( W = O^\top \Lambda Q \) where \( \Lambda \) has i.i.d. \( \sqrt{5/3} \cdot \text{Beta}(3,1) \) diagonal entries, and \((O, Q)\) are uniformly random.

In all three settings, \( \Lambda \) is normalized so that \( \kappa_2 = \mathbb{E}[\Lambda^2] = 1 \). Figure 4 compares per-iteration estimation errors, where we again observe that Bayes-OAMP achieves the same error as Gaussian Bayes-AMP and single-iterate Bayes-OAMP for i.i.d. noise, but improves over the other procedures for the remaining noise settings.

6 Conclusion

We have developed AMP algorithms for both symmetric and rectangular spiked random matrix models in the context of orthogonally invariant noise. These algorithms extend those of [Fan20] and [MV21b] by allowing for matrix-valued AMP iterates, multivariate non-linearities, and spectral initializations using super-critical sample eigenvectors or singular vectors of the observed data. We have derived the forms of the Onsager debiasing coefficients for a general class of such algorithms, and established rigorous Gaussian state evolutions for their iterates. These depend on the spectral distributions of the noise via their (symmetric or rectangular) free cumulants.

We developed one application of such algorithms to estimate the super-critical signal vectors in these models, by choosing Bayes posterior mean denoising functions as the non-linearities in AMP.
Figure 4: Estimation errors for AMP iterates $U_t$ and $V_t$ in the rectangular spiked model with $(m,n) = (3000,4000)$, rank-2 signal, signal priors $U_*, V_* \sim \frac{1}{2} \delta_{(0,1)} + \frac{1}{4} \delta_{\sqrt{2}-1} + \frac{1}{4} \delta_{-\sqrt{2}-1}$, and signal strengths $(\theta_1, \theta_2) = (2,1.5)$. Iterates $V_0$ and $U_1$ correspond to the spectral initializations. The noise spectral distributions are (left) square-root of the Marcenko-Pastur law, (middle) Uniform$[\sqrt{3/7}, 2\sqrt{3/7}]$, and (right) the $\sqrt{5/3} \cdot \text{Beta}(3,1)$ distribution.

This Bayes-OAMP approach is similar to the algorithms of [RF12, MV21c], but applies the posterior mean conditional on all preceding AMP iterates, instead of only the single preceding iterate. Computation of this posterior mean is enabled by the above state evolution, whose parameters may be estimated empirically. We observe in simulation that Bayes-OAMP can yield improved estimation accuracy over both the spectral initialization and single-iterate AMP algorithms, and can be more robust than standard AMP algorithms that are designed for white noise. These simulations suggest that the method may be sufficiently accurate and stable for practical use in PCA applications.
A Proof for symmetric square matrices

A.1 State evolution for auxiliary square AMP

As discussed in Section 3.3, in the setting of Theorem 3.3, we consider an auxiliary AMP algorithm starting at time index \(-\tau\), with an initialization \(U^{(\tau)}_{-\tau} \in \mathbb{R}^{n \times K}\) independent of \(W\) and having the iterates, for \(t = -\tau, -\tau + 1, -\tau + 2, \ldots\)

\[
F(t)^{\tau} = XU^{(\tau)}_t - \sum_{s = -\tau}^{t} U^{(\tau)}_s b^{(\tau)}_{ts}^T, \quad U^{(\tau)}_{t+1} = u_{t+1}(F^{\tau}_{-\tau}, \ldots, F^{\tau}_t). \tag{A.1}
\]

The coefficients \(b^{(\tau)}_{ts}\) are defined as follows: For \(T \geq 1\), we define

\[
\phi^{(\tau)}_{all,T} = \left(\partial_s U^{(\tau)}_t\right)_{r,s \in \{-\tau, \ldots, T\}}, \quad b^{(\tau)}_{all,T} = \sum_{j=0}^{\infty} \kappa_{j+1} (\phi^{(\tau)}_{all,T})^j
\]

where \(\{\kappa_j\}_{j \geq 1}\) are the free cumulants of the limit eigenvalue distribution \(\Lambda\) for \(W\). We take the above debiasing coefficients up to iteration \(T\) to be the blocks, for \(s, t \in \{-\tau, \ldots, T\}\),

\[
b^{(\tau)}_{ts} = b^{(\tau)}_{all,T}[t, s] \tag{A.2}
\]

Supposing that \((U'_s, U^{(\tau)}_{-\tau}) \xrightarrow{W_2} (U'_s, U^{(\tau)}_{-\tau})\), we define recursively the following state evolution:

Having defined the joint law of \((U'_s, U^{(\tau)}_{-\tau}, \ldots, U^{(\tau)}_t, F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_{t-1})\) for some \(t \geq -\tau\), we define \(\mu^{(\tau)}_{all,t} \in \mathbb{R}^{(t+\tau+1)K \times K'}\) having the blocks

\[
\mu^{(\tau)}_s = \mathbb{E}[U_s U^{t\top}_s] \cdot S' \in \mathbb{R}^{K \times K'} \text{ for each } s = -\tau, \ldots, t.
\]

Recalling the function \(\Theta^{[j],[\cdot, \cdot]}\) from (2.6), we define also the \((t + \tau + 1)K \times (t + \tau + 1)K\) matrices

\[
\Delta^{(\tau)}_{all,t} = \left(\mathbb{E}[U^{(\tau)}_{s} U^{(\tau)}_{s'}] \right)_{r,s \in \{-\tau, \ldots, t\}}, \quad \Phi^{(\tau)}_{all,t} = \left(\mathbb{E}[\partial_s u_r (F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_{t-1})]\right)_{r,s \in \{-\tau, \ldots, t\}},
\]

\[
\Sigma^{(\tau)}_{all,t} = \sum_{j=0}^{\infty} \Theta^{[j]}(\Phi^{(\tau)}_{all,t}, \kappa_{j+2} \Delta^{(\tau)}_{all,t}).
\]

Then we define the next joint law of \((U'_s, U^{(\tau)}_{-\tau}, \ldots, U^{(\tau)}_t, F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_t)\) by

\[
(F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_t) \mid U'_s \sim \mathcal{N} \left(\mu^{(\tau)}_{all,t} : U^{(\tau)}_s, \Sigma^{(\tau)}_{all,t}\right), \quad U^{(\tau)}_{s+1} = u_{s+1}(F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_s) \text{ for } s = -\tau, \ldots, t.
\]

Corollary A.1. In the symmetric spiked model (3.1), suppose Assumption 2.1 holds for \(W\). Suppose the initialization \(U^{(\tau)}_{-\tau} \in \mathbb{R}^{n \times K}\) is independent of \(W\), and \((U'_s, U^{(\tau)}_{-\tau}) \xrightarrow{W_2} (U'_s, U^{(\tau)}_{-\tau})\) a.s. as \(n \to \infty\). Suppose each function \(u_{t+1}(\cdot)\) is Lipschitz, and each derivative \(\partial_s u_{t+1}(F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_t)\) exists and is continuous on a set of probability 1 under the above law of \((F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_t)\). Then for any \(T \geq 1\), a.s. as \(n \to \infty\),

\[
(U'_s, U^{(\tau)}_{-\tau}, \ldots, U^{(\tau)}_{T+1}, F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_T) \xrightarrow{W_2} (U'_s, U^{(\tau)}_{-\tau}, \ldots, U^{(\tau)}_{T+1}, F^{(\tau)}_{-\tau}, \ldots, F^{(\tau)}_T).
\]

27
for each fixed \( t \) argument as in [Fan20, Theorem 3.1(a)], we obtain a.s. as \( t \to \infty \) that

\[
\mathbf{F}_t^{(\tau)} = \frac{1}{n} \mathbf{U}_s' \mathbf{S}' \mathbf{U}_s' \mathbf{U}_t^{(\tau)} + \mathbf{W} \mathbf{U}_t^{(\tau)} - \sum_{s=-\tau}^{t} \mathbf{U}_s^{(\tau)} \hat{b}_{ts}^{(\tau)\top}.
\]

Then, approximating \( n^{-1} \mathbf{S}' \mathbf{U}_s' \mathbf{U}_t^{(\tau)} \) by \( \mathbf{S}' \mathbf{W} \mathbb{E}[\mathbf{U}'_t \mathbf{U}_t^{(\tau)\top}] = \mu_t^{(\tau)\top} \), we consider an alternative AMP sequence with the same initialization \( \mathbf{U}_{-\tau} = \mathbf{U}_t^{(\tau)} \) and “side information” \( \mathbf{U}_s' \), defined by

\[
\mathbf{Z}_t = \mathbf{W} \mathbf{U}_t - \sum_{s=-\tau}^{t} \hat{U}_s \hat{b}_{ts},
\]

\[
\hat{\mathbf{F}}_t = \mathbf{Z}_t + \mathbf{U}_s' \mu_t^{(\tau)\top},
\]

\[
\hat{\mathbf{U}}_{t+1} = \hat{u}_{t+1}(\mathbf{Z}_{-\tau}, \ldots, \mathbf{Z}_t, \mathbf{U}_s') \overset{\text{def}}{=} u_{t+1}(\hat{\mathbf{F}}_{-\tau}, \ldots, \hat{\mathbf{F}}_t).
\]

Here, the debiasing coefficients are defined as

\[
\hat{b}_{ts} = \langle \partial_s \hat{u}_{t+1}(\mathbf{Z}_{-\tau}, \ldots, \mathbf{Z}_t, \mathbf{U}_s') \rangle = \langle \partial_s u_{t+1}(\hat{\mathbf{F}}_{-\tau}, \ldots, \hat{\mathbf{F}}_t) \rangle.
\]

Using Theorem 2.3 to analyze this AMP algorithm, we have a.s. as \( n \to \infty \) that

\[
(U'_s, \hat{U}_{-\tau}, \ldots, \hat{\mathbf{U}}_{t+1}, \hat{\mathbf{F}}_{-\tau}, \ldots, \hat{\mathbf{F}}_t) \overset{W_2}{\rightarrow} (U'_s, U_{-\tau}^{(\tau)}', \ldots, U_{t+1}^{(\tau)}', F_{-\tau}^{(\tau)}, \ldots, F_t^{(\tau)})
\]

for each fixed \( t \geq -\tau \), as described by the above state evolution. Then, applying the same inductive argument as in [Fan20, Theorem 3.1(a)], we obtain a.s. as \( n \to \infty \)

\[
n^{-1} \| \mathbf{U}_t^{(\tau)} - \hat{\mathbf{U}}_t \|_F^2 \to 0, \quad n^{-1} \| F_t^{(\tau)} - \hat{F}_t \|_F^2 \to 0
\]

for each fixed \( t \geq -\tau \), which implies this corollary. \( \square \)

As discussed in Section 3.3, we now specialize this auxiliary AMP algorithm (A.1) to the two-phase algorithm

\[
u_{t+1}(\mathbf{F}_{-\tau}^{(\tau)}, \ldots, \mathbf{F}_t^{(\tau)}) = \begin{cases} \mathbf{F}_t^{(\tau)} S^{-1} & \text{for } -\tau + 1 \leq t < 0, \\ u_{t+1}(\mathbf{F}_0^{(\tau)}, \ldots, \mathbf{F}_t^{(\tau)}) & \text{for } t \geq 0. \end{cases}
\]

The auxiliary algorithm is initialized at

\[
\mathbf{U}_{-\tau}^{(\tau)} = (u_{-\tau}^1, \ldots, u_{-\tau}^K) \text{ with } u_k^{(\tau)} = (\mu_{\text{pca},k} u_k + \sqrt{1-\mu_{\text{pca},k}^2} \cdot z_k) / \theta_k \text{ for each } k = 1, \ldots, K,
\]

where \( \mu_{\text{pca},k} \) is as defined in Theorem 3.1, and \( z_1, \ldots, z_K \) are independent standard Gaussian random vectors also independent of \( \mathbf{W} \).

For each \( t \geq 1 \), we adopt the following block decomposition of \( \phi_{all,t}^{(\tau)} \), where the first block corresponds to indices \( \{-\tau, \ldots, -1\} \) and the second to indices \( \{0, \ldots, t\} \):

\[
\phi_{all,t}^{(\tau)} = \begin{pmatrix} \phi_{-\tau}^{(\tau)} \\ \phi_{-}^{(\tau)} \end{pmatrix}, \quad \text{where } \phi_{-}^{(\tau)} \in \mathbb{R}^{T \times \tau K} \text{ and } \phi_{-\tau}^{(\tau)} \in \mathbb{R}^{(t+1)K \times (t+1)K}.
\]
Note that, due to the lower-triangular form of $\phi_{all,t}^{(\tau)}$ and the linear update rule for the first $\tau$ steps, we have $\phi_{all}^{(\tau)} = 0$ and

$$
\phi_{l-}^{(\tau)} = \begin{pmatrix}
0 & 0 & \cdots & 0 & 0 \\
S^{-1} & 0 & \cdots & 0 & 0 \\
0 & S^{-1} & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & S^{-1} & 0 \\
\end{pmatrix}, \quad \phi_{l+}^{(\tau)} = \begin{pmatrix}
0 & \cdots & 0 & S^{-1} \\
0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 \\
\end{pmatrix}.
$$

Applying Lemma D.2 with $A = \phi_{all,t}^{(\tau)}$ and $B = S^{-1}$, for all $r \in \{0, \ldots, t\}$ and $c \in \{1, \ldots, \tau\}$,

$$
(\phi_{all,t}^{(\tau)})^{j-r} = \begin{cases}
(\phi_{t}^{(\tau)})^{j-c}[r,0]S^{-c} & 1 \leq c \leq j, \\
0 & j < c.
\end{cases}
$$

Similarly, for the state evolution, we decompose

$$
\mu_{all,t}^{(\tau)} = \begin{pmatrix}
\mu_{l-}^{(\tau)} \\
\mu_{l+}^{(\tau)}
\end{pmatrix}, \quad \Delta_{all,t}^{(\tau)} = \begin{pmatrix}
\Delta_{l-}^{(\tau)} & \Delta_{l+}^{(\tau)} \\
\Delta_{l+}^{(\tau)} & \Delta_{l+}^{(\tau)}
\end{pmatrix}, \quad \Phi_{all,t}^{(\tau)} = \begin{pmatrix}
\Phi_{l-}^{(\tau)} & \Phi_{l-}^{(\tau)} \\
\Phi_{l+}^{(\tau)} & \Phi_{l+}^{(\tau)}
\end{pmatrix}, \quad \Sigma_{all,t}^{(\tau)} = \begin{pmatrix}
\Sigma_{l-}^{(\tau)} & \Sigma_{l-}^{(\tau)} \\
\Sigma_{l+}^{(\tau)} & \Sigma_{l+}^{(\tau)}
\end{pmatrix}.
$$

\textbf{A.2 Phase I – linear AMP}

We first establish the convergence of the iterates and the associated state evolution of the first $\tau$ steps of this auxiliary AMP algorithm, as $\tau \to \infty$. For notational convenience, in this section only, we re-index these iterates as $1, 2, \ldots, \tau$ and provide a standalone result for such a linear AMP algorithm.

Let $U_1 = (u_1^1, \ldots, u_1^K)$ with each $u_1^k = (\mu_{\text{pca},k}u_0^k + \sqrt{1 - \mu_{\text{pca},k}^2} \cdot z_k)/\theta_k$. The first $\tau$ iterates of the above auxiliary AMP algorithm has the structure of the following linear AMP:

$$
F_t = Xu_t - \sum_{i=1}^{t} \kappa_{t-i+1} U_i S^{-(t-i)}, \quad U_{t+1} = F_t S^{-1}.
$$

We write $F_0 = U_1 S$. Up to iterate $\tau$, let $\mu_\tau = (\mu_t)_{1 \leq t \leq \tau}$ and $\Sigma_\tau = (\sigma_{st})_{1 \leq s, t \leq \tau}$ be the parameters of the state evolution describing this linear AMP, where $\mu_t \in \mathbb{R}^{K \times K'}$ and $\sigma_{st} \in \mathbb{R}^{K \times K}$. Recall $\mu_{\text{pca}} = \text{diag}(\mu_{\text{pca},1}, \ldots, \mu_{\text{pca},K}) \in \mathbb{R}^{K \times K'}, \quad \text{Id} - \mu_{\text{pca}}\mu_{\text{pca}}^\top = \text{diag}(1 - \mu_{\text{pca},1}^2, \ldots, 1 - \mu_{\text{pca},K}^2) \in \mathbb{R}^{K \times K}$.

Then Corollary A.1 ensures

$$(U'_s, U_1, \ldots, U_\tau, F_1, \ldots, F_\tau) \overset{W_2}{\to} (U'_s, U_1, \ldots, U_\tau, F_1, \ldots, F_\tau)$$

where the limiting distribution is defined by

$$(F_1, \ldots, F_\tau) \mid U'_s \sim \mathcal{N}(\mu_\tau \cdot U'_s, \Sigma_\tau),$$

$$U_1 \sim S^{-1} \cdot \mathcal{N}(\mu_{\text{pca}} \cdot U'_s, \text{Id} - \mu_{\text{pca}}\mu_{\text{pca}}^\top), \quad U_t = S^{-1} F_{t-1} \text{ for } 2 \leq t \leq \tau.
$$

\textbf{Lemma A.2.} Under Assumptions \textit{2.1 and 3.2(a) and (c)}, the following hold for the linear AMP algorithm (A.5):

29
(a) \( \lim_{t \to \infty} \limsup_{n \to \infty} \| \mathbf{F}_t - \mathbf{F}_{\text{pca}} \|_F / \sqrt{n} = 0 \) a.s.

(b) The state evolution satisfies \( \mu_t = \mu_{\text{pca}} \) for every \( t \geq 1 \), and \( \lim_{\min(s,t) \to \infty} \sigma_{st} = \text{Id} - \mu_{\text{pca}} \mu_{\text{pca}}^\top \).

Proof. Recall the sample eigenvalues \( \lambda_k(\mathbf{X}) \) for \( k = 1, \ldots, K' \) from (3.3), constituting the \( K_+ \) largest and \( K_- \) smallest eigenvalues of \( \mathbf{X} \), with associated eigenvectors \( \mathbf{f}_{\text{pca}}^i \) satisfying (3.4). Denote the remaining eigenvalues and eigenvectors as \( \lambda_i(\mathbf{X}) \) and \( \mathbf{f}_{\text{pca}}^i \) for \( i = K' + 1, \ldots, n \) in any order, with the same normalization \( \| \mathbf{f}_{\text{pca}}^i \| = \sqrt{n} \). Denote \( \| \Lambda \|_\infty = \max(|\lambda_+|, |\lambda_-|) \), fix a small constant \( \delta > 0 \), and define the event

\[
\mathcal{E}_n = \{ |\lambda_i(\mathbf{X})| < \| \Lambda \|_\infty + \delta \text{ for all } i = K' + 1, \ldots, n \}.
\]

Then \( \mathcal{E}_n \) occurs almost surely for all large \( n \) by Assumption 2.1(c) and Weyl’s eigenvalue interlacing inequality.

Let \( \mathbf{f}_t^k \) be the \( k \)th column of the linear AMP iterate \( \mathbf{F}_t \). For part (a), it suffices to show

\[
\lim_{t \to \infty} \limsup_{n \to \infty} \| \mathbf{f}_t^k - \mathbf{f}_{\text{pca}}^k \| / \sqrt{n} = 0
\]

for each \( k \in \{1, \ldots, K\} \). Fixing any such \( k \), by the definition of linear AMP in (A.5), we have

\[
\mathbf{f}_t^k = \frac{1}{\theta_k} \cdot \mathbf{X} \mathbf{f}_{t-1}^k - \sum_{j=1}^{t} \frac{r_{t-j+1}}{\theta_k^{t-j+1}} \cdot \mathbf{f}_{j-1}^k.
\]  

(A.7)

We first show that the component of \( \mathbf{f}_t^k \) orthogonal to \( \mathbf{f}_{\text{pca}}^k \) vanishes a.s. in the limits \( t \to \infty \) and \( n \to \infty \). Define \( r_{t}^{k,i} = \mathbf{f}_{\text{pca}}^i \mathbf{f}_t^k / n \) for each \( i \in \{1, \ldots, n\} \). Then applying (A.7),

\[
\mathbf{r}_t^{k,i} = \frac{1}{\theta_k} \cdot \mathbf{f}_{\text{pca}}^i \mathbf{X} \frac{\mathbf{f}_t^k}{n} - \sum_{j=0}^{t-1} \frac{\kappa_{t-j}}{\theta_k^{t-j}} \cdot \mathbf{r}_j^{k,i}.
\]

(A.8)

For any \( i \in \{1, \ldots, K'\} \setminus \{k\} \), by the initialization \( \mathbf{r}_0^{k,i} = \theta_k \mathbf{u}_1^k = \mu_{\text{pca},k} \mathbf{u}_s^k + \sqrt{1 - \mu_{\text{pca},k}^2} \cdot z_k \), we have

\[
r_{0}^{k,i} = \frac{\mu_{\text{pca},k}}{n} \left( \mathbf{f}_{\text{pca}}^i \mathbf{u}_s^k + \frac{1 - \mu_{\text{pca},k}^2}{n} \right) \mathbf{z}_k \to 0
\]

a.s. as \( n \to \infty \), where the first term converges to 0 by Theorem 3.1, and the second term converges to 0 since \( \mathbf{f}_{\text{pca}}^i \mathbf{z}_k / n \sim \mathcal{N}(0, 1/n) \). Thus, it follows from the recursion (A.8) that

\[
\lim_{n \to \infty} r_{t}^{k,i} = 0 \text{ a.s. for each fixed } t \geq 0 \text{ and } i \in \{1, \ldots, K'\} \setminus \{k\}.
\]  

(A.9)

For any \( i \in \{K' + 1, \ldots, n\} \), consider a space \( \mathcal{X} \) of bounded infinite-dimensional vectors with elements in \([0, \infty)\). For each \( t \geq 0 \), we define an element \( \mathbf{g}_t^{k,i} \in \mathcal{X} \) by padding 0’s after \((r_t^{k,i}, r_{t-1}^{k,i}, \ldots, r_0^{k,i})\), i.e.,

\[
\mathbf{g}_t^{k,i} = (r_t^{k,i}, r_{t-1}^{k,i}, \ldots, r_0^{k,i}, 0, 0, \ldots).
\]

30
For some $\iota \in (0, 1)$ chosen as in Assumption 3.2(c), let us consider a norm $\| \cdot \|$ on $X$ defined by

$$\|(x_0, x_{-1}, x_{-2}, \ldots)\| = \sup_{j \geq 0} |x_{-j}| \cdot \iota^j.$$ 

Consider a map $g : X \to X$ defined as $g(\varrho_{t-1}^{k,i}) = \varrho_t^{k,i}$ for each $t \geq 1$. We verify that $g$ is contractive with respect to the norm $\| \cdot \|$. Let $\{\varrho_t\}_{t \geq 1}$ and $\{\tilde{\varrho}_t\}_{t \geq 1}$ be two sequences of vectors in $X$ given by

$$\begin{cases} 
\varrho_t = (r_t, r_{t-1}, \ldots, r_0, 0, \ldots) \\
\tilde{\varrho}_t = (\tilde{r}_t, \tilde{r}_{t-1}, \ldots, \tilde{r}_0, 0, \ldots)
\end{cases}$$

where both $\{r_t\}_{t \geq 1}$ and $\{\tilde{r}_t\}_{t \geq 1}$ satisfy the same recursion as in (A.8). Then we have

$$\|g(\varrho_{t-1}) - g(\tilde{\varrho}_{t-1})\| = \|\varrho_t - \tilde{\varrho}_t\| = \sup_{0 \leq j \leq t} |r_j - \tilde{r}_j| \cdot \iota^{t-j} = \max \{|r_t - \tilde{r}_t|, \iota \cdot \|\varrho_{t-1} - \tilde{\varrho}_{t-1}\|\}.$$ 

We then need to control $|r_t - \tilde{r}_t|$. It follows from (A.8) that

$$|r_t - \tilde{r}_t| \leq \left| \frac{\lambda_i(X)}{\theta_k} \cdot (r_{t-1} - \tilde{r}_{t-1}) \right| + \sum_{j=0}^{t-1} \frac{\kappa_{t-j}}{\theta_k} |r_j - \tilde{r}_j| \cdot \iota^{t-1-j}$$

$$\leq \left( \frac{\lambda_i(X)}{|\theta_k|} + \sum_{j=1}^{t} \frac{\kappa_j}{|\theta_k|^{1/2} \iota^{j-1}} \right) \cdot \max_{0 \leq j \leq t-1} |r_j - \tilde{r}_j| \cdot \iota^{t-1-j}$$

$$\leq \left( \frac{\|\Lambda\|_{\infty} + \delta}{|\theta_k|} + \sum_{j=1}^{\infty} \frac{\kappa_j}{|\theta_k|^{1/2} \iota^{j-1}} \right) \cdot \|\varrho_{t-1} - \tilde{\varrho}_{t-1}\|$$

where the last inequality holds on the event $E_n$ defined above. For sufficiently small $\delta > 0$, we have $\eta \in (0, 1)$ by Assumption 3.2(c). Then denoting $\rho = \max(\eta, \iota) \in (0, 1)$, we obtain

$$\|g(\varrho_{t-1}) - g(\tilde{\varrho}_{t-1})\| \leq \rho \cdot \|\varrho_{t-1} - \tilde{\varrho}_{t-1}\|.$$ 

Therefore, $g$ is a $\rho$-contraction. Applying this property to $\{\varrho_t^{k,i}\}_{t \geq 1}$ yields

$$|r_t^{k,i}| \leq \|(r_t^{k,i}, r_{t-1}^{k,i}, \ldots, r_0^{k,i}, 0, \ldots)\| \leq \rho^t \cdot \|(r_0^{k,i}, 0, \ldots) - (0, 0, \ldots)\| = \rho^t \cdot |r_0^{k,i}|. \quad (A.10)$$

This holds simultaneously for all $i \in \{K'+1, \ldots, n\}$ on the event $E_n$.

Now write $f_t^k = \xi_t^k f_{\text{pca}} + r_t^k$ where $r_t^k$ is orthogonal to $f_{\text{pca}}^k$. Since $\{f_{\text{pca}}^i/\sqrt{n}\}_{i=1,\ldots,n}$ is an orthonormal basis of $\mathbb{R}^n$, we can expand $r_t^k$ as

$$r_t^k = \sum_{i=1, i \neq k}^{n} \frac{f_{\text{pca}}^i}{n} \cdot f_t^k = \sum_{i=1, i \neq k}^{K'} K' r_t^{k,i} f_{\text{pca}}^i + \sum_{i=K'+1}^{n} L_3 r_t^{k,i} f_{\text{pca}}^i - L_2.$$
By (A.9), we have \( \lim_{n \to \infty} \| \mathcal{I}_1 \| / \sqrt{n} = 0 \) for each fixed \( t \geq 1 \). For \( \mathcal{I}_2 \), we apply (A.10) on the event \( \mathcal{E}_n \):

\[
\frac{\| \mathcal{I}_2 \|^2}{n} = \sum_{i=K'+1}^{n} (r_{t,i}^k)^2 \leq \rho^2 t \cdot \sum_{i=K'+1}^{n} (r_{0,0}^k)^2 \leq \rho^2 t \cdot \frac{\| r_0^k \|^2}{n} \leq \rho^2 t \cdot \frac{\| f_0^k \|^2}{n}.
\]

By the initialization \( f_0^k = \mu_{\text{pca},k} u_k^k + \sqrt{1 - \mu_{\text{pca},k}^2} \cdot z_k \), we have \( \lim_{n \to \infty} \frac{\| f_0^k \|^2}{n} = 1 \). Thus, as desired,

\[
\lim_{t \to \infty} \lim_{n \to \infty} \sup \frac{\| r_t^k \|}{\sqrt{n}} = 0. 
\] (A.11)

We now show that \( \xi_t^k \overset{\text{def}}{=} \frac{f_t^k \top}{\mu_{\text{pca}}} u_k^k / n \to 1 \) by using the state evolution (A.6) of linear AMP. By this state evolution, we have

\[
\mu_{t+1} = \mathbb{E}[U_{t+1} U_{t+1}^\top] \cdot S' = S^{-1} \cdot \mathbb{E}[F_t U_{t+1}^\top] \cdot S' = S^{-1} \cdot \mathbb{E}[(\mu_t U_{t+1} + Z_t) U_{t+1}^\top] \cdot S' = S^{-1} \mu_t S'
\]

where the second equality follows from \( U_{t+1} = S^{-1} F_t \) and the last equality is due to \( Z_t \perp U'_{t+1} \) and \( \mathbb{E}[U_{t+1} U_{t+1}^\top] = \text{Id} \). From the definition of \( U_1 \), it may be checked that \( \mu_1 = \mu_{\text{pca}} \in \mathbb{R}^{K \times K'} \). Then it follows from the above that \( \mu_t = \mu_{\text{pca}} \) for all \( t \geq 1 \). Thus for each \( k \in \{1, \ldots, K\} \), we have

\[
\mu_{\text{pca},k} = \lim_{n \to \infty} \frac{f_{t+1}^k \top}{\mu_{\text{pca}}} u_k^k / n, \quad \mu_{\text{pca},k} = \lim_{n \to \infty} \frac{f_t^k \top}{\mu_{\text{pca}}} u_k^k = \lim_{n \to \infty} \frac{\xi_t^k \mu_{\text{pca}}}{\mu_{\text{pca},k}} \frac{f_t^k \top}{\mu_{\text{pca}}} u_k^k + \frac{r_t^k \top}{\mu_{\text{pca},k}} u_k^k
\]

where the left equality follows from Theorem 3.1, and the right equality applies \( \mu_t = \mu_{\text{pca}} \). This further implies that

\[
\lim_{n \to \infty} \sup_{n \to \infty} \left| (\xi_t^k - 1) \frac{f_{t+1}^k \top}{\mu_{\text{pca}}} u_k^k / n \right| \leq \lim_{n \to \infty} \sup_{n \to \infty} \frac{\| r_t^k \|}{\sqrt{n}}.
\]

Since \( \lim_{n \to \infty} \frac{f_{t+1}^k \top}{\mu_{\text{pca}}} u_k^k / n = \mu_{\text{pca},k} \neq 0 \) a.s. by Theorem 3.1, taking the limit as \( t \to \infty \) on both sides, it follows from (A.11) that

\[
\lim_{n \to \infty} \lim_{n \to \infty} \sup_{n \to \infty} | \xi_t^k - 1 | = 0. \quad \text{(A.12)}
\]

Then, recalling \( f_t^k = \xi_t^k f_{t+1}^k + r_t^k \) and combining with (A.11),

\[
\lim_{n \to \infty} \lim_{n \to \infty} \sup_{n \to \infty} \frac{\| f_t^k - f_{t+1}^k \|}{\sqrt{n}} = 0.
\]

This shows both part (a) and the claim about \( \mu_t \) in part (b).

It remains to show the convergence of \( \sigma_{st} \) in part (b). By (A.6) and the above identities \( \mathbb{E}[F_t U_{t+1}^\top] = \mu_t = \mu_{\text{pca}} \), we have

\[
\sigma_{st} = \mathbb{E}[(F_s - \mu_s U_s')(F_t - \mu_t U_t')^\top] = \mathbb{E}[F_s F_t^\top] - \mu_{\text{pca}} H_{\text{pca}}^\top = \lim_{n \to \infty} \frac{F_t^\top F_s}{n} - \mu_{\text{pca}} H_{\text{pca}}^\top.
\]
Thus, using the notation \( (uv) = u^T v / n \) and recalling the decomposition \( f^k_t = \xi^k_t f^k_{\text{pca}} + r^k_t \), we can bound the difference between \( (1 - \mu^2_{\text{pca},k}) \mathbb{I}\{k = k'\} \) and the \((k, k')\)th entry of \( \sigma_{st} \) as

\[
|\sigma_{st}[k, k'] - (1 - \mu^2_{\text{pca},k}) \mathbb{I}\{k = k'\}| \\
\leq \limsup_{n \to \infty} \left| \xi^k_s \xi^k_t (f^k_{\text{pca}} r^k_{\text{pca}}) - \mathbb{I}\{k = k'\} \right| \\
\leq \limsup_{n \to \infty} \left| \xi^k_s \xi^k_t (r^k_{\text{pca}} r^k_{\text{pca}}) - \mathbb{I}\{k = k'\} \right| \\
\leq \limsup_{n \to \infty} \left| \xi^k_s \xi^k_t - 1 \right| + \left| \xi^k_s (f^k_{\text{pca}} r^k_{\text{pca}}) \right| + \left| \xi^k_t (f^k_{\text{pca}} r^k_{\text{pca}}) \right| + \left| (r^k_s r^k_t) \right|
\]

where the last inequality applies the triangle inequality and the orthogonality and normalization of \( f^k_{\text{pca}} \) in (3.4). Finally, by the convergence of \( \tau \) in (A.11) and that of \( \xi \) in (A.12) as \( t \to \infty \), we get

\[
\lim_{\min(s, t) \to \infty} \sigma_{st}[k, k'] = (1 - \mu^2_{\text{pca},k}) \cdot \mathbb{I}\{k = k'\},
\]

i.e. \( \lim_{\min(s, t) \to \infty} \sigma_{st} = \text{Id} - \mu_{\text{pca}} \mu_{\text{pca}}^T \).

The auxiliary AMP iterations \( U^{(r)}_x, F^{(r)}_\tau, U^{(r)}_x \) are defined by this linear AMP algorithm up to the iterates \( U^{(r)}_0 \) and \( F^{(r)}_0 \). Thus, translating Lemma A.2 back to the indexing of this auxiliary AMP algorithm, and recalling the spectral initializations \( F_0 = F_{\text{pca}} \) and \( U_0 = F_{\text{pca}} S^{-1} \), the lemma implies the following:

\[
\lim_{\tau \to \infty} \limsup_{n \to \infty} \frac{\|F^{(r)}_i - F^{(r)}_0\|_F}{\sqrt{n}} = 0, \quad \lim_{\tau \to \infty} \limsup_{n \to \infty} \frac{\|U^{(r)}_i - U^{(r)}_0\|_F}{\sqrt{n}} = 0 \text{ for all fixed } i \leq 0,
\]

\[
\mu^{(r)}_i = \mu_{\text{pca}}, \quad \lim_{\tau \to \infty} \Sigma^{(r)}_{all, i, j} = \text{Id} - \mu_{\text{pca}} \mu_{\text{pca}}^T \text{ for all fixed } i, j \leq 0. \quad (A.13)
\]

### A.3 Phase II - auxiliary AMP

Now we proceed to prove Theorem 3.3 which provides a precise characterization of the state evolution of the AMP algorithm with spectral initialization for symmetric matrices.

**Proof of Theorem 3.3.** We show by induction that the following statements hold a.s. for each \( t \geq 0 \). In particular, part (b) is the main result that we want to prove, and all of the other parts serve as a road map for the proof. Parts (a)–(d) will apply standard comparison arguments, and the specific definitions (3.12) and (3.16) will be used to verify parts (e) and (f).

(a) \( \lim_{\tau \to \infty} \limsup_{n \to \infty} \|U^{(r)}_t - U_t\|_F / \sqrt{n} = 0 \) and \( \lim_{n \to \infty} \|U_t\|_F / \sqrt{n} < C_t \) for a constant \( C_t > 0 \).

(b) \( (U'_x, U_0, \ldots, U_t, F_0, \ldots, F_{t-1}) \overset{W_2}{\to} (U'_x, U_0, \ldots, U_t, F_0, \ldots, F_{t-1}) \) where the limit distribution is defined by the recursion (3.13).

(c) \( \lim_{\tau \to \infty} \Phi^{(r)}_t = \Phi_t \) and \( \lim_{\tau \to \infty} \Delta^{(r)}_t = \Delta_t \).

(d) \( \lim_{\tau \to \infty} \lim_{n \to \infty} \|\Phi^{(r)}_t - \Phi_t\| = 0. \)

(e) \( \lim_{\tau \to \infty} \limsup_{n \to \infty} \|F^{(r)}_t - F_t\|_F / \sqrt{n} = 0 \) and \( \lim_{n \to \infty} \|F_t\|_F / \sqrt{n} < C_t \) for a constant \( C_t > 0 \).

(f) \( \lim_{\tau \to \infty} \mu^{(r)}_t = \mu_t \) and \( \lim_{\tau \to \infty} \Sigma^{(r)}_t = \Sigma_t. \)
Denote by $t^{(a)}, t^{(b)}, \ldots, t^{(f)}$ the claims of parts (a-f) at iteration $t$. We induct on $t$. For the base case $t = 0$, parts (a), (c), and (f) follow from (A.13) proved in the previous section, and the initializations $U_0 = F_{\text{pca}} S^{-1}$ and $F_0 = F_{\text{pca}}$. For (c) and (d), we have $\phi_0^{(r)} = \phi_0 = \Phi_0^{(r)} = \Phi_0 = 0$, while $\lim_{t \to \infty} \Delta_0^{(r)} = \Delta_0$ follows from part (a). For (b), $(U^*_s, U_0) \overset{W_2}{\to} (U^*_t, U_0)$ follows from the convergence $(U^*_s, U_0^{(r)}) \overset{W_2}{\to} (U^*_t, U_0^{(r)})$ for the auxiliary AMP algorithm, together with (A.13). Thus all statements hold for $t = 0$.

For the induction step, let $t \geq 1$, and assume $s^{(a-f)}$ holds for $0 \leq s \leq t - 1$. Let us then prove statements (a-f) for iteration $t$.

**Part (a)** By Assumption 3.2(b), $u_t$ is Lipschitz, so there exists some $L > 0$ such that

$$\frac{\|U(t) - U(t)\|_F}{\sqrt{n}} = \frac{\|u_t(F(0), \ldots, F(t)) - u_t(F(0), \ldots, F(t-1))\|_F}{\sqrt{n}} \leq L \cdot \sum_{s=0}^{t-1} \frac{\|F(s) - F_s\|_F}{\sqrt{n}}.$$

Then by the induction hypothesis $t - 1^{(e)}$, $\lim_{t \to \infty} \limsup_{n \to \infty} \frac{\|U(t) - U(t)\|_F}{\sqrt{n}} = 0$. Moreover, for a constant $C_t > 0$,

$$\lim_{n \to \infty} \frac{\|U(t)\|_F^2}{n} = \lim_{n \to \infty} \frac{\|u_t(F(0), \ldots, F(t-1))\|_F^2}{n} = \mathbb{E} \left[ \|u_t(F(0), \ldots, F(t-1))\|^2 \right] < C_t.$$

**Part (b)** Let $u_{s,i} \in \mathbb{R}^{K'}$ denote the $i$th row of $U_{s} \in \mathbb{R}^{n \times K'}$, and similarly for the other matrix variables. Let $g$ be any pseudo-Lipschitz function, i.e. $\|g(x) - g(y)\| \leq C(1 + \|x\| + \|y\|)\|x - y\|$ for a constant $C > 0$. Then there exist some constant $C' > 0$ such that

$$\frac{\sum_{s=0}^{t} \|u_{s,i}\| + \sum_{s=0}^{t-1} \|f_{s,i}\|}{\sqrt{n}} \leq C' \frac{\sum_{s=0}^{t} \|u_{s,i}\|^2 + \|f_{s,i}\|^2}{\sqrt{n}}.$$

**Part (c)** Let $\tau$ be a constant. Let $t \geq 1$, and assume $s^{(a-f)}$ holds for $0 \leq s \leq t - 1$. Let us then prove statements (a-f) for iteration $t$.
By the induction hypothesis $t - 1(f)$ and Assumption 3.2(b), for all sufficiently large $\tau$ and each $1 \leq s < r \leq t$, $\partial_s u_t(F_{r-1}^{(r)})$ exists and is continuous on a set of probability 1 under the law of $F_{r-1}^{(r)}$ prescribed by Corollary A.1. Thus, for all sufficiently large $\tau$, Corollary A.1 applies to this auxiliary AMP algorithm up to iteration $t$, to yield

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(u_{s,i}, u_{0,i}^{(r)}, \ldots, u_{t,i}^{(r)}, f_{0,i}^{(r)}, \ldots, f_{t-1,i}^{(r)}) = \mathbb{E}\left[ g(U_{s}^{(r)}, U_{0}^{(r)}, \ldots, U_{t}^{(r)}, F_{0}^{(r)}, \ldots, F_{t-1}^{(r)}) \right] .
$$

Recalling $(F_{0}^{(r)}, \ldots, F_{t}^{(r)}) \mid U_{t}^{r} \sim \mathcal{N}(\mu_{t-1}^{(r)} U_{t}^{r}, \Sigma_{t-1}^{(r)})$ and $(F_{0}, \ldots, F_{t-1}) \mid U_{t}^{r} \sim \mathcal{N}(\mu_{t-1}^{(r)}, \Sigma_{t-1}^{(r)})$, let us couple these laws by setting $(F_0^{(r)}, \ldots, F_{t}^{(r)}) = \mu_{t-1}^{(r)} U_{t}^{r} + (\Sigma_{t-1}^{(r)})^{1/2} Z$ and $(F_0, \ldots, F_{t-1}) = \mu_{t-1} U_{t}^{r} + \Sigma_{t-1}^{1/2} Z$ for a standard Gaussian vector $Z$ independent of $U_{t}^{r}$. Since $\lim_{t \to \infty} \mu_{t}^{(r)} = \mu_{t-1}$ and $\lim_{t \to \infty} \Sigma_{t}^{(r)} = \Sigma_{t-1}$ by $t - 1(f)$, and $g(\cdot)$ is pseudo-Lipschitz, by the dominated convergence theorem,

$$
\lim_{\tau \to \infty} \mathbb{E}\left[ g(U_{s}^{(r)}, U_{0}^{(r)}, \ldots, U_{t}^{(r)}, F_{0}^{(r)}, \ldots, F_{t-1}^{(r)}) \right] = \mathbb{E}\left[ g(U_{s}^{(r)}, U_{0}, \ldots, U_{t}, F_{0}, \ldots, F_{t-1}) \right] .
$$

Combining the above three displays, this shows for any pseudo-Lipschitz function $g$ that

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(u_{s,i}, u_{0,i}, \ldots, u_{t,i}, f_{0,i}, \ldots, f_{t-1,i}) = g(U_{s}^{(r)}, U_{0}, \ldots, U_{t}, F_{0}, \ldots, F_{t-1}),
$$

which implies the desired Wasserstein-2 convergence in part (b).

**Part (c)** Since the upper-left $tK \times tK$ submatrix of $\Phi_{t}^{(r)}$ is exactly equal to $\Phi_{t-1}^{(r)}$ and the last column block of $\Phi_{t}^{(r)}$ is 0, by the induction hypothesis $t - 1(c)$, it suffices to show that the last row block of $\Phi_{t}^{(r)}$ converges to that of $\Phi_{t}$, and similarly for $\Delta_{t}^{(r)}$ and $\Delta_{t}$. For any $s \in \{0, \ldots, t - 1\}$, we have

$$
\Phi_{t}^{(r)}[t, s] = \mathbb{E}\left[ \partial_s u_t(F_{0}^{(r)}, \ldots, F_{t}^{(r)}) \right] \quad \text{and} \quad \Phi_{t}[t, s] = \mathbb{E}\left[ \partial_s u_t(F_{0}, \ldots, F_{t-1}) \right] .
$$

By Assumption 3.2(b), $\partial_s u_t$ is bounded and continuous on a set of probability 1 under $(F_0, \ldots, F_{t-1})$, so by weak convergence of $(F_0^{(r)}, \ldots, F_{t}^{(r)})$ to $(F_0, \ldots, F_{t-1})$ under the above coupling,

$$
\lim_{\tau \to \infty} \Phi_{t}^{(r)}[t, s] = \Phi_{t}[t, s] .
$$

Similarly, for any $s \in \{0, \ldots, t\}$, $\lim_{\tau \to \infty} \Delta_{t}^{(r)}[t, s] = \Delta_{t}[t, s]$ by the induced coupling of $(U_0^{(r)}, \ldots, U_{t}^{(r)})$ and $(U_0, \ldots, U_{t})$ and the dominated convergence theorem.

**Part (d)** As above, we only need to show that the last row block of $\phi_{t}^{(r)}$ converges to that of $\phi_{t}$. For any $s \in \{0, \ldots, t - 1\}$, by the auxiliary AMP state evolution of Corollary A.1,

$$
\lim_{n \to \infty} \phi_{t}^{(r)}[t, s] = \lim_{n \to \infty} \langle \partial_s u_t(F_{0}^{(r)}, \ldots, F_{t}^{(r)}) \rangle = \mathbb{E}\left[ \partial_s u_t(F_{0}^{(r)}, \ldots, F_{t}^{(r)}) \right] = \Phi_{t}^{(r)}[t, s] \quad (A.14)
$$

where the second equality again follows from $\partial_s u_t$ being bounded and continuous on a set of probability 1 under $(F_0^{(r)}, \ldots, F_{t}^{(r)})$, for sufficiently large $\tau$. Similarly, by $t(b),

$$
\lim_{n \to \infty} \phi_{t}[t, s] = \mathbb{E}\left[ \partial_s u_t(F_{0}, \ldots, F_{t-1}) \right] = \Phi_{t}[t, s] . \quad (A.15)
$$

Combining (A.14) and (A.15), it then follows from $t(c)$ that

$$
\lim_{\tau \to \infty} \lim_{n \to \infty} \left( \phi_{t}^{(r)}[t, s] - \phi_{t}[t, s] \right) = 0 .
$$
Part (e) We now control the difference between $F_t^{(\tau)}$ and $F_t$. By their definitions in (A.1) and (3.10), applying the triangle inequality yields

$$
\frac{\|F_t^{(\tau)} - F_t\|_F}{\sqrt{n}} \leq \frac{\|X\|\|U_t^{(\tau)} - U_t\|_F}{\sqrt{n}} + \frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=-\tau}^{t} \left( U_i^{(\tau)} b_{t,i}^{(\tau)} - U_0^{(\tau)} b_{t,0}^{(\tau)} \right) \right) + \frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=-\tau}^{t} \left( U_i^{(\tau)} b_{t,i}^{(\tau)} - U_i^{(\tau)} b_{t,i}^{(\tau)} \right) \right).
$$

(A.16)

Since $\lim_{n \to \infty} ||X|| < C$ a.s. for a constant $C > 0$, the first term vanishes in the limit $n \to \infty$ by $t^{(a)}$. For the second term in (A.16), we can further decompose and bound it by

$$
\frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=-\tau}^{0} U_i^{(\tau)} b_{t,i}^{(\tau)} - \sum_{i=-\tau}^{0} U_i b_{t,i}^{(\tau)} \right) \leq \frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=-\tau}^{0} U_i^{(\tau)} b_{t,i}^{(\tau)} - U_0 b_{t,0}^{(\tau)} \right) + \frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=-\tau}^{0} ||U_i^{(\tau)} - U_0||_F \right).
$$

(A.17)

Term $I_1$. By the triangle inequality,

$$
\frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=-\tau}^{0} U_i^{(\tau)} b_{t,i}^{(\tau)} - U_0 b_{t,0}^{(\tau)} \right) \leq \left( \frac{1}{n} \sum_{i=-\tau}^{0} b_{t,i}^{(\tau)} - b_{t,0}^{(\tau)} \right) \frac{\|U_0\|_F}{\sqrt{n}} + \frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{i=-\tau}^{0} ||b_{t,i}^{(\tau)}|| \cdot \frac{\|U_i^{(\tau)} - U_0\|_F}{\sqrt{n}} \right).
$$

(A.18)

Since $\|U_0\|_F/\sqrt{n} = \|F_{\text{PCA}} S^{-1}\|_F/\sqrt{n} < C$ for a constant $C > 0$, it suffices to bound $I_{1,1}$ and $I_{1,2}$. To bound $I_{1,1}$, recall the definition of debiasing coefficients in (3.12):

$$
b_{t,0} = \tilde{B}_t[t,0] = \sum_{j=0}^{\infty} \phi_t^{(j)}[t,0] \kappa_{j+1} = \sum_{j=0}^{t} \phi_t^{(j)}[t,0] \kappa_{j+1},
$$

where the second equality applies $\phi_t^{(j)} = 0$ when $j > t$. Recall also the debiasing coefficients in the auxiliary AMP sequence from (A.2): For every $i \in \{-\tau, \ldots, 0\}$,

$$
b_{t,i}^{(\tau)} = B_t^{(\tau)}[t,i] = \sum_{j=0}^{\infty} \kappa_{j+1} \left( \phi_{\text{all},t}^{(j)} \right)^{j}[t,i] = \sum_{j=-i}^{\infty} \kappa_{j+1} \left( \phi_t^{(j)} \right)^{j+i}[t,0] S^i
$$

(A.19)

where the last equality follows from (A.3) and the fact that $(\phi_t^{(j)})^j = 0$ when $j > t$. Therefore,

$$
\sum_{i=-\tau}^{0} b_{t,i}^{(\tau)} = \sum_{i=-\tau}^{0} \sum_{j=-i}^{\infty} \kappa_{j+1} \left( \phi_t^{(j)} \right)^{j+i}[t,0] S^i
$$

$$
= \sum_{j=0}^{t} \left( \phi_t^{(j)} \right)^{j}[t,0] \left( \sum_{i=-\tau}^{0} \kappa_{j-i+1} S^i \right)
$$

$$
= \sum_{j=0}^{t} \left( \phi_t^{(j)} \right)^{j}[t,0] \kappa_{j+1} + \sum_{j=0}^{t} \left( \phi_t^{(j)} \right)^{j}[t,0] \sum_{i=\tau+1}^{\infty} \kappa_{j+i+1} S^{-i}
$$

and thus

$$
\sum_{i=-\tau}^{0} b_{t,i}^{(\tau)} - b_{t,0} = \sum_{j=0}^{t} \left( (\phi_t^{(j)})^j[t,0] - \phi_t^{(j)}[t,0] \right) \kappa_{j+1} + \sum_{j=0}^{t} \left( \phi_t^{(j)} \right)^{j}[t,0] \left( \sum_{i=\tau+1}^{\infty} \kappa_{j+i+1} S^{-i} \right).
$$
Taking the limits \( \tau \to \infty \) and \( n \to \infty \), the first term converges to 0 by \( t^{(d)} \). Since each \( u_t \) is Lipschitz, we have \( \|\phi_t^{(r)}j[t,0]\| \leq C_t \) for some constant \( C_t > 0 \) and all \( j = 0, \ldots, t \), and thus the second term will also converge to 0 as \( \tau \to \infty \) by the absolute convergence of the series defining \( \hat{\kappa}_{j+1} \) in (3.11), as a consequence of Assumption 3.2(c). We therefore conclude

\[
\lim_{\tau \to \infty} \lim_{n \to \infty} \sum_{i=-\tau}^{0} b_{t,i}^{(r)} - b_{t,0} = 0. \tag{A.20}
\]

To bound \( I_{1,2} \), we apply Lemma D.1 with \( x_i^{(r)} = \|b_{t,-i}^{(r)}\| \) and \( y_i^{(r)} = \limsup_{n \to \infty} \|U_i^{(r)} - U_0\|_F / \sqrt{n} \) for every \( i = 0, \ldots, \tau \). We need to verify that \( \{x_i^{(r)}\}_{i=0}^{\tau} \) and \( \{y_i^{(r)}\}_{i=0}^{\tau} \) satisfy the conditions of Lemma D.1. By (A.19) and Assumption 3.2(c), for some constants \( C_t, C_t' > 0 \) independent of \( i \) and \( \tau \),

\[
\sum_{i=0}^{\tau} |x_i^{(r)}| = \sum_{i=0}^{\tau} \|b_{t,-i}^{(r)}\| \leq \sum_{i=0}^{\tau} \sum_{j=0}^{i+t} |\kappa_{j+1}| \cdot \|\phi_t^{(r)}j-i[t,0]S^{-i}\| 
\leq C_t \sum_{i=0}^{\tau} \sum_{j=0}^{i+t} \frac{|\kappa_{i+j+1}|}{\min_{k \in \{1, \ldots, K\}} \theta_k} \leq C_t \sum_{j=0}^{\tau} \sum_{i=0}^{\infty} \frac{|\kappa_{i+j+1}|}{\min_{k \in \{1, \ldots, K\}} \theta_k} < C_t'.
\]

Therefore \( \{x_i^{(r)}\}_{i=0}^{\tau} \) is uniformly bounded. Furthermore,

\[
\sum_{i=\tau}^{t} |x_i^{(r)}| \leq C_t \sum_{j=0}^{t} \sum_{i=\tau}^{\infty} \frac{|\kappa_{i+j+1}|}{\min_{k \in \{1, \ldots, K\}} \theta_k},
\]

so for any \( \epsilon > 0 \), there exists some \( \tau > 0 \) such that \( \sum_{i=\tau}^{\infty} |x_i^{(r)}| < \epsilon \) for all \( \tau \geq \tau \). For \( y_i^{(r)} \), we have

\[
|y_i^{(r)}| \leq \lim_{n \to \infty} \frac{\|U_i^{(r)}\|_F}{\sqrt{n}} + \lim_{n \to \infty} \frac{\|U_0\|_F}{\sqrt{n}}.
\]

Lemma A.2 implies that the first limit exists, depends on \( (i, \tau) \) only via the difference \( \tau - i \), and \( \lim_{\tau \to \infty} \lim_{n \to \infty} \|U_i^{(r)}\|_F / \sqrt{n} < C \) for a constant \( C > 0 \). Then there is a constant \( C' > 0 \) independent of \( (i, \tau) \) for which \( |y_i^{(r)}| < \epsilon \), so \( \{y_i^{(r)}\}_{i=0}^{\tau} \) is uniformly bounded. By (A.13), \( \lim_{\tau \to \infty} y_i^{(r)} = 0 \) for each \( 0 \leq i \leq \tau \). Hence, applying Lemma D.1,

\[
\lim_{\tau \to \infty} \limsup_{n \to \infty} \sum_{i=-\tau}^{0} b_{t,i}^{(r)} \cdot \frac{\|U_i^{(r)} - U_0\|_F}{\sqrt{n}} = 0. \tag{A.21}
\]

Combining (A.18), (A.20) and (A.21), we obtain that

\[
\lim_{\tau \to \infty} \limsup_{n \to \infty} \frac{1}{\sqrt{n}} \cdot \left\| \sum_{i=-\tau}^{0} U_i^{(r)}b_{t,i}^{(r)\top} - U_0b_{t,0}^{\top} \right\|_F = 0. \tag{A.22}
\]

**Term** \( I_2 \). The convergence of the term \( I_2 \) in (A.17) is a more straightforward comparison: By the triangle inequality,

\[
\sum_{i=1}^{t} \frac{1}{\sqrt{n}} \left\| U_i^{(r)}b_{t,i}^{(r)\top} - U_i^{(r)}b_{t,0}^{\top} \right\|_F \leq \sum_{i=1}^{t} \|b_{t,i}^{(r)}\| \cdot \frac{\|U_i^{(r)} - U_i\|_F}{\sqrt{n}} + \sum_{i=1}^{t} \|b_{t,i}^{(r)} - b_{t,0}\| \cdot \frac{\|U_i\|_F}{\sqrt{n}}.
\]
For each \( i \in \{0, \ldots, t\} \), by the definition of the debiasing coefficients in (A.2), we have
\[
b_{t,i}^{(r)} = \sum_{j=0}^{\infty} \kappa_{j+1} \left( \phi_t^{(r)} \right)^j [t,i] = \sum_{j=0}^{t} \kappa_{j+1} \left( \phi_t^{(r)} \right)^j [t,i].
\]
Since each \( u_t \) is Lipschitz, this implies \( \|b_{t,i}^{(r)}\| < C_t \) for a constant \( C_t > 0 \). Recalling the definitions of \( b_{t,i}^{(r)} \) in (A.2) and \( b_{t,i} \) in (3.12), and applying \( t(d) \), also \( \lim_{\tau \to \infty} \lim_{n \to \infty} b_{t,i}^{(r)} - b_{t,i} = 0 \). Then combining with \( t(u) \),
\[
\lim \lim_{\tau \to \infty} \sup_{n \to \infty} \frac{1}{\sqrt{n}} \left\| U_i^{(r)} b_{t,i}^{(r)\top} - U_i b_{t,i}^{\top} \right\|_F = 0. \tag{A.23}
\]
Combining (A.16), (A.17), (A.22), and (A.23) shows \( \lim_{\tau \to \infty} \lim_{n \to \infty} \|F_t^{(r)} - F_t\|_F/\sqrt{n} = 0 \) as desired for part (c).

**Part (f)** Recall that \( \Sigma_t^{(r)} \) is the lower-right \((t+1)K \times (t+1)K\) submatrix of \( \Sigma_{all,t}^{(r)} \) as in (A.4). By the block decompositions of \( \Phi_t^{(r)} \) and \( \Delta_t^{(r)} \) in (A.4), we obtain
\[
\begin{align*}
\Sigma_t^{(r)} &= \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_{all,t}^{(r)} \right)^i \Delta_{all,t}^{(r)} (\Phi_{all,t}^{(r)\top})^{j-i} + \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_t^{(r)} \right)^i \Delta_{t}^{(r)} (\Phi_t^{(r)\top})^{j-i} \\
&\hspace{1cm}+ \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_{all,t}^{(r)} \right)^i \Delta_{all,t}^{(r)} (\Phi_{all,t}^{(r)\top})^{j-i} + \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_t^{(r)} \right)^i \Delta_{t}^{(r)} (\Phi_t^{(r)\top})^{j-i} \\
&= \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_t^{(r)} \right)^i \Delta_{t}^{(r)} (\Phi_t^{(r)\top})^{j-i} + \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_{all,t}^{(r)} \right)^i \Delta_{all,t}^{(r)} (\Phi_{all,t}^{(r)\top})^{j-i} \\
&\hspace{1cm}+ \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_t^{(r)} \right)^i \Delta_{t}^{(r)} (\Phi_t^{(r)\top})^{j-i} + \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_{all,t}^{(r)} \right)^i \Delta_{all,t}^{(r)} (\Phi_{all,t}^{(r)\top})^{j-i}.
\end{align*}
\tag{A.24}
\]
where we observe that the third term is the transpose of the second term. To show \( \lim_{\tau \to \infty} \Sigma_t^{(r)} = \Sigma_t \), we recall \( \Delta_t = \Delta_t + \Delta_t^\top + \Delta_t \), and correspondingly decompose \( \Sigma_t \) in (3.16) as follows:
\[
\Sigma_t = \sum_{j=0}^{\infty} \Phi_t^{(r)} \kappa_{j+2} \Delta_{t} + \kappa_{j+2} \Delta_t^\top + \kappa_{j+2} \Delta_t \] \[= \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_t^{(r)} \right)^i \Delta_{t}^{(r)} (\Phi_t^{(r)\top})^{j-i} + \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_{all,t}^{(r)} \right)^i \Delta_{all,t}^{(r)} (\Phi_{all,t}^{(r)\top})^{j-i} \\
+ \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_t^{(r)} \right)^i \Delta_{t}^{(r)} (\Phi_t^{(r)\top})^{j-i} + \sum_{j=0}^{\infty} \sum_{i=0}^{j} \left( \Phi_{all,t}^{(r)} \right)^i \Delta_{all,t}^{(r)} (\Phi_{all,t}^{(r)\top})^{j-i}.
\tag{A.25}
\]
We will show that, for any \( r, c \in \{0, 1, \ldots, t\} \),
\[
\lim_{\tau \to \infty} \Sigma_t^{(r)} [r,c] = \Sigma_t [r,c], \quad \lim_{\tau \to \infty} \left( \Sigma_t^{(r)} + \Sigma_t^{(r)\top} \right) [r,c] = \Sigma_t [r,c], \quad \lim_{\tau \to \infty} \Sigma_t^{(r)} [r,c] = \Sigma_t [r,c].
\]
Convergence of $\hat{\Sigma}_t^{(r)}$  

We have

\[
\hat{\Sigma}_t^{(r)}[r, c] = \sum_{j=0}^{\infty} \kappa_{j+2} \sum_{i=0}^{j} \sum_{\alpha, \beta = 1}^\tau (\Phi_{all,t}^{(r)})^i [r, -\alpha] \Delta_{all,t}^{(r)}[-\alpha, -\beta](\Phi_{all,t}^{(r)} \text{T})^{j-i}[-\beta, c] \\
= \sum_{j=0}^{\infty} \sum_{i=0}^{j} \sum_{\alpha, \beta = 1}^\tau \kappa_{j+2} (\Phi_t^{(r)})^{-i}[r, 0] S^{-\alpha} \Delta_{-}[\alpha, -\beta] S^{-\beta} (\Phi_t^{(r)} \text{T})^{j-i-\beta}[0, c]
\]

where the second equality follows from (A.3). Let us write $\Delta$, and the summation may be restricted to

\[
\sum_{\alpha, \beta = 1}^\tau \kappa_{j+2} (\Phi_t^{(r)})^{-i}[r, 0] S^{-\alpha} \Delta_{-}[\alpha, -\beta] S^{-\beta} (\Phi_t^{(r)} \text{T})^{j-i-\beta}[0, c]
\]

Fixing any $\tau \in [0, t]$, we apply Lemma D.1 with $\Phi_t^{(r)}[r, 0]$ is bounded by a constant and the summation may be restricted to $\alpha, \beta \in [0, t]$, there exists some constant $C_t > 0$ such that

\[
|\hat{Z}_t^{(r)}| \leq C_t \max_{\ell = 0}^{2\tau} \sum_{\alpha, \beta = 1}^{\tau} \frac{|\kappa_{\ell + \alpha + \beta + 2}|}{\min_{k \in \{1, \ldots, K\}} |\theta_k|^{\alpha + \beta}} \cdot \left| \Delta_{-}[\alpha, -\beta] - \Delta_{t}[0, 0] \right|
\]

Fixing any $\ell \in [0, 2\tau]$, we apply Lemma D.1 with $x_i^{(2\tau)} = i \cdot |\kappa_{\ell+i+2}| / \min_{k \in \{1, \ldots, K\}} |\theta_k|^i$ and

\[
y_{i\ell}^{(2\tau)} = \sup_{\alpha, \beta \geq 1, \alpha + \beta = i} \left| \Delta_{-}[\alpha, -\beta] - \Delta_{t}[0, 0] \right|
\]

By Assumption 3.2(c), $\{x_i^{(2\tau)}\}_{i=1, \ldots, 2\tau}$ satisfies the condition of Lemma D.1. By Lemma A.2, for any fixed $\alpha, \beta \geq 0$,

\[
\lim_{\tau \to \infty} \Delta_{-}[\alpha, -\beta] = \lim_{\tau \to \infty} \mathbb{E}[U_{-\alpha} U_{-\beta}^\text{T}] \\
= \lim_{\tau \to \infty} S^{-1} \mathbb{E}[F_{-\alpha-1} F_{-\beta-1}^\text{T}] S^{-1} \\
= S^{-1} \left( \lim_{\tau \to \infty} \Sigma_{all,t}[\alpha-1, -\beta-1] + \mu_{-\alpha-1} \cdot \mu_{-\beta-1}^\text{T} \right) S^{-1} = S^{-2}, \quad (A.26)
\]
Recall that $\Delta_t[0,0] = \mathbb{E}[U_0U_0^\top] = S^{-2}$

because $U_0 = F_{\text{pca}} S^{-1} W \xrightarrow{d} U_0$, this gives $\lim_{\tau \to \infty} y_{i}^{(2\tau)} = 0$ for every fixed $i$. Observe that $\mathbb{E}[\|F_{-\alpha}^{(r)}\|]$ depends on $(\alpha, \tau)$ only via the difference $\tau - \alpha$, and (A.13) implies $\lim_{\tau \to \infty} \mathbb{E}[\|F_{-\alpha}^{(r)}\|] < C$ for a constant $C > 0$. Then $\mathbb{E}[\|F_{-\alpha}^{(r)}\|] < C'$ for a constant $C' > 0$ and all $(\alpha, \tau)$, implying that $\{y_{i}^{(2\tau)}\}$ is uniformly bounded. Then it follows from Lemma D.1 that $\lim_{\tau \to \infty} \|\hat{I}_{1}^{(r)}\| = 0$.

Next, we show that $\lim_{\tau \to \infty} \hat{\Sigma}_{\tau}^{(r)} = \hat{\Sigma}_{\tau}$, taking $t \to \infty$, by the convergence of $\Phi_{t}^{(r)}$ to $\Phi_{t}$ in $t^{(c)}$,

$$
\lim_{\tau \to \infty} \hat{\Sigma}_{\tau}^{(r)} = \sum_{p,q=0}^{\infty} \Phi_{t}^{(r)}[r,0] \left( \sum_{\alpha,\beta=1}^{\infty} \kappa_{p+q+\alpha+\beta+2} S^{-\alpha} \Delta_{t}[0,0] S^{-\beta} \right) (\Phi_{t}^{(r)}\top)^{q}[0,c].
$$

Recall that $\Delta_{t}[0,0] = S^{-2}$, so this commutes with $S^{-\beta}$. Then, applying

$$
\sum_{\alpha,\beta=1}^{\infty} \kappa_{p+q+\alpha+\beta+2} S^{-\alpha} \Delta_{t}[0,0] S^{-\beta} = \sum_{\alpha,\beta=1}^{\infty} \kappa_{p+q+\alpha+\beta+2} S^{-\alpha} \Delta_{t}[0,0] - 2 \sum_{\alpha=0}^{\infty} \kappa_{p+q+\alpha+2} S^{-\alpha} \Delta_{t}[0,0] + \kappa_{p+q+2} \Delta_{t}[0,0]
$$

we identify this limit as $\hat{\Sigma}_{\tau}^{(r)}[r,c]$. So

$$
\lim_{\tau \to \infty} \hat{\Sigma}_{\tau}^{(r)} = \hat{\Sigma}_{\tau}.
$$

**Convergence of $\hat{\Sigma}_{\tau}^{(r)}$** Next, we determine the limit of $\hat{\Sigma}_{\tau}^{(r)}$. Applying (A.3) and re-indexing the summation similarly as above by setting $p = i$ and $q = j - i - \beta$,

$$
\hat{\Sigma}_{\tau}^{(r)}[r,c] = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \sum_{\alpha=0}^{\infty} \sum_{\beta=1}^{\infty} \kappa_{j+2} \Phi_{t}^{(r)}[r,\alpha] \Delta_{\tau}^{(r)}[\alpha, -\beta] (\Phi_{all,t}^{(r)}\top)^{i-j}[\beta, c]
$$

$$
= \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \sum_{\alpha=0}^{\infty} \sum_{\beta=1}^{\infty} \kappa_{j+2} \Phi_{t}^{(r)}[r,\alpha] \Delta_{\tau}^{(r)}[\alpha, -\beta] S^{-\beta} (\Phi_{t}^{(r)}\top)^{i-j-\beta}[0,c]
$$

$$
= \sum_{p,q=0}^{\infty} (\Phi_{t}^{(r)}p[r,\alpha] \sum_{\alpha=0}^{\infty} \sum_{\beta=1}^{\infty} \kappa_{p+q+\beta+2} \Delta_{\tau}[\alpha, -\beta] S^{-\beta} (\Phi_{t}^{(r)}\top)^{i-j-\beta}[0,c]
$$

$$
= \sum_{p,q=0}^{\infty} (\Phi_{t}^{(r)}p[r,\alpha] \sum_{\alpha=0}^{\infty} \sum_{\beta=1}^{\infty} \kappa_{p+q+\beta+2} \Delta_{\tau}[\alpha, -\beta] S^{-\beta} (\Phi_{t}^{(r)}\top)^{i-j-\beta}[0,c]
$$

$$
+ \sum_{p,q=0}^{\infty} (\Phi_{t}^{(r)}p[r,\alpha] \sum_{\alpha=0}^{\infty} \sum_{\beta=1}^{\infty} \kappa_{p+q+\beta+2} \Delta_{\tau}[\alpha, 0] S^{-\beta} (\Phi_{t}^{(r)}\top)^{i-j-\beta}[0,c].
$$

For each fixed $\alpha \in [0,t]$ and $\beta > 0$, note that by $I^{(r)}$,

$$
\lim_{\tau \to \infty} \mathbb{E}[\|U_{0}^{(r)} - U_{\alpha}\|^2] = 0
$$

40
and by \( t^{(a)} \) and (A.26),
\[
\lim_{\tau \to \infty} \mathbb{E} \| U_{-\beta}^{(\tau)} - U_0 \|^2 = \lim_{\tau \to \infty} \mathbb{E} \| U_{-\beta}^{(\tau)} - U_0^{(\tau)} \|^2
= \lim_{\tau \to \infty} \text{Tr} \left( \Delta_{\text{all},t}^{(\tau) \top}[-\beta, -\beta] - 2\Delta_{\text{all},t}^{(\tau)}[-\beta, 0] + \Delta_{\text{all},t}^{(\tau)}[0, 0] \right) = 0.
\]
So
\[
\lim_{\tau \to \infty} \Delta_{\text{all},t}^{(\tau)}[\alpha, -\beta] - \Delta_t[\alpha, 0] = \lim_{\tau \to \infty} \mathbb{E} \left[ U_{\alpha}^{(\tau)} U_{-\beta}^{(\tau) \top} \right] - \mathbb{E} \left[ U_{\alpha} U_{0}^{(\tau) \top} \right] = 0.
\]

Then by Lemma D.1 and a similar argument as used previously for \( \hat{I}_1^{(\tau)} \), this implies \( \lim_{\tau \to \infty} \hat{I}_1^{(\tau)} = 0 \). For \( \hat{I}_2^{(\tau)} \), we take \( \tau \to \infty \) and write \( \sum_{\beta=1}^{\infty} \kappa_{p+q+2} \beta S^{-\beta} = \kappa_{p+q+2} - \kappa_{p+q+2} \text{Id} \). Then
\[
\lim_{\tau \to \infty} \hat{\Sigma}_t^{(\tau)}[r, c] = \sum_{p, q=0}^{\infty} \sum_{\alpha=0}^{t} \Phi_t^p[r, \alpha] \left( \Delta_t[\alpha, 0] \left( \kappa_{p+q+2} - \kappa_{p+q+2} \text{Id} \right) \Phi_t^q \right)[0, c]
= \sum_{p, q=0}^{\infty} \left[ \Phi_t^p \left( (\Delta_t^\top + \Delta_t^\top) \circ (\kappa_{p+q+2} - \kappa_{p+q+2} \text{Id}) \Phi_t^q \right) \right][r, c].
\]
Summing this limit with its transpose, we obtain
\[
\lim_{\tau \to \infty} \left( \hat{\Sigma}_t^{(\tau)} + \hat{\Sigma}_t^{(\tau) \top} \right) = \bar{\Sigma}_t. \quad \text{(A.28)}
\]

**Convergence of \( \hat{\Sigma}_t^{(\tau)} \)** By the convergence of \( \Phi_t^{(\tau)}, \Delta_t^{(\tau)} \) to \( \Phi_t, \Delta_t \) in (c) and the fact that the summation in \( \hat{\Sigma}_t^{(\tau)} \) is finite and may be restricted to \( j \leq 2t \), we immediately have
\[
\lim_{\tau \to \infty} \hat{\Sigma}_t^{(\tau)} = \lim_{\tau \to \infty} \sum_{j=0}^{\infty} \kappa_{j+2} \sum_{i=0}^{j} \left( \Phi_t^i \right)^j \Delta_t^i \left( \Phi_t^i \right)^{j-1} = \sum_{j=0}^{\infty} \kappa_{j+2} \sum_{i=0}^{j} \Phi_t^i \Delta_t \left( \Phi_t^i \right)^{j-1} = \hat{\Sigma}_t. \quad \text{(A.29)}
\]
Collecting the decomposition of \( \hat{\Sigma}_t^{(\tau)} \) in (A.24), that of \( \hat{\Sigma}_t \) in (A.25), and the convergence of each component in (A.27), (A.28) and (A.29), we obtain \( \lim_{\tau \to \infty} \hat{\Sigma}_t^{(\tau)} = \hat{\Sigma}_t \) as desired for part (f). \( \square \)

**B Proof for rectangular matrices**

The proof strategy for the rectangular case is similar as that for the symmetric case, where an auxiliary AMP algorithm serves as a bridge between the actual algorithm and the desired state evolution.

**B.1 State evolution for auxiliary AMP**

In the setting of Theorem 4.3, consider the auxiliary AMP algorithm with initialization \( U_{-\tau}^{(\tau)} \in \mathbb{R}^{m \times K} \) independent of \( \mathbf{W} \), having the iterates for \( t = -\tau, -\tau + 1, -\tau + 2, \ldots \)
\[
\mathbf{G}_t^{(\tau)} = \mathbf{X}^{\top} U_t^{(\tau)} - \sum_{s=-\tau}^{-1} V_s^{(\tau)} U_t^{(\tau) \top}, \quad V_t^{(\tau)} = v_t(G_{-\tau}^{(\tau)}, \ldots, G_t^{(\tau)}),
\]
\[
\mathbf{F}_t^{(\tau)} = \mathbf{X} V_t^{(\tau)} - \sum_{s=-\tau}^{t} U_t^{(\tau)} A_{ts}^{(\tau) \top}, \quad U_{t+1}^{(\tau)} = u_{t+1}(\mathbf{F}_{-\tau}^{(\tau)}, \ldots, F_t^{(\tau)}).
\]
For \( T \geq 1 \), we define
\[
\phi^{(r)}_{all,T} = \left( \langle \partial_s U^{(r)}_s \rangle \right)_{r,s \in \{-\tau, \ldots, T\}}, \quad \psi^{(r)}_{all,T} = \left( \langle \partial_s V^{(r)}_s \rangle \right)_{r,s \in \{-\tau, \ldots, T\}},
\]
\[
a^{(r)}_{all,T} = \sum_{j=0}^{\infty} \kappa_2(j+1) \psi^{(r)}_{all,T} \phi^{(r)}_{all,T} \psi^{(r)}_{all,T}^j, \quad b^{(r)}_{all,T} = \gamma \sum_{j=0}^{\infty} \kappa_2(j+1) \phi^{(r)}_{all,T} \psi^{(r)}_{all,T} \phi^{(r)}_{all,T}^j
\]
where \( \{\kappa_2\}_{j \geq 1} \) are the rectangular free cumulants of the limit singular value distribution \( \Lambda \) for \( W \).

We set the above debiasing coefficients as the blocks
\[
a^{(r)}_{ts} = a^{(r)}_{all,T}[t, s], \quad b^{(r)}_{ts} = b^{(r)}_{all,T}[t, s]. \tag{B.2}
\]
Supposing that \( (U^{(r)}_s, U^{(r)}_s) \xrightarrow{W_2} (U^{(r)}_s, U^{(r)}_s) \), we define the following state evolution: Having defined joint laws \( (U^{(r)}_s, U^{(r)}_s, \ldots, U^{(r)}_s, F^{(r)}_s, \ldots, F^{(r)}_{s-1}) \) and \( (V^{(r)}_s, V^{(r)}_s, \ldots, V^{(r)}_{s-1}, G^{(r)}_s, \ldots, G^{(r)}_{s-1}) \), we define
\[
\nu^{(r)}_{all,t} = \begin{pmatrix} \nu^{(r)}_{-\tau} \\ \vdots \\ \nu^{(r)}_{t} \end{pmatrix}, \quad \text{where} \quad \nu^{(r)}_s = \mathbb{E}[U^{(r)}_s U^{(r)}_s^\top] \cdot S^{\prime} \sqrt{\gamma} \in \mathbb{R}^{K \times K^{\prime}} \text{ for each } s = -\tau, \ldots, t.
\]

We set
\[
\Delta^{(r)}_{all,t} = \left( \mathbb{E}[U^{(r)}_s U^{(r)}_s^\top]_{r,s \in \{-\tau, \ldots, t\}} \right)_{r,s \in \{-\tau, \ldots, t\}}, \quad \Phi^{(r)}_{all,t} = \left( \mathbb{E}[\partial_s u_r(F_1, \ldots, F_{r-1})] \right)_{r,s \in \{-\tau, \ldots, t\}},
\]
\[
\Gamma^{(r)}_{all,t} = \left( \mathbb{E}[V^{(r)}_s V^{(r)}_s^\top] \right)_{r,s \in \{-\tau, \ldots, t\}}, \quad \Psi^{(r)}_{all,t} = \left( \mathbb{E}[\partial_s v_r(G_1, \ldots, G_t)] \right)_{r,s \in \{-\tau, \ldots, t\}},
\]
leaving the last row and column blocks of \( \Gamma^{(r)}_{all,t} \) and \( \Psi^{(r)}_{all,t} \) momentarily undefined, and set
\[
\Omega^{(r)}_{all,t} = \gamma \sum_{j=0}^{\infty} \Theta^{(j)} \left[ \Phi^{(r)}_{all,t}, \Psi^{(r)}_{all,t}, \Phi^{(r)}_{all,t} \kappa_2(j+1) \Delta^{(r)}_{all,t}, \Psi^{(r)}_{all,t} \kappa_2(j+1) \Gamma^{(r)}_{all,t} \right],
\]
where \( \Theta^{(j)}[\cdot, \cdot, \cdot] \) is defined in (2.12). We define the joint law of \( (V^{(r)}_s, V^{(r)}_s, \ldots, V^{(r)}_{t}, G^{(r)}_s, \ldots, G^{(r)}_{t}) \) by
\[
(G^{(r)}_{-\tau}, \ldots, G^{(r)}_{t}) \mid V^{(r)}_s \sim \mathcal{N} \left( \nu^{(r)}_{all,t} \cdot V^{(r)}_s, \Omega^{(r)}_{all,t} \right),
\]
\[
V^{(r)}_s = v_s(G^{(r)}_{-\tau}, \ldots, G^{(r)}_{t}) \text{ for } s = -\tau, \ldots, t.
\]
Recalling \( \Xi^{(j)}[\cdot, \cdot, \cdot] \) from (2.14), we set
\[
\mu^{(r)}_{all,t} = \begin{pmatrix} \mu^{(r)}_{-\tau} \\ \vdots \\ \mu^{(r)}_{t} \end{pmatrix}, \quad \text{where} \quad \mu^{(r)}_s = \mathbb{E}[V^{(r)}_s V^{(r)}_s^\top] \cdot S^{\prime}/\sqrt{\gamma} \text{ for each } s = -\tau, \ldots, t,
\]
\[
\Sigma^{(r)}_{all,t} = \sum_{j=0}^{\infty} \Xi^{(j)} \left[ \Phi^{(r)}_{all,t}, \Psi^{(r)}_{all,t}, \Phi^{(r)}_{all,t} \kappa_2(j+1) \Delta^{(r)}_{all,t}, \Psi^{(r)}_{all,t} \kappa_2(j+1) \Gamma^{(r)}_{all,t} \right], \tag{B.3}
\]
and define the joint law of \( (U^{(r)}_s, U^{(r)}_s, \ldots, U^{(r)}_{t+1}, F^{(r)}_{-\tau}, \ldots, F^{(r)}_{t}) \) by
\[
(F^{(r)}_{-\tau}, \ldots, F^{(r)}_{t}) \mid U^{(r)}_s \sim \mathcal{N} \left( \mu^{(r)}_{all,t} \cdot U^{(r)}_s, \Sigma^{(r)}_{all,t} \right),
\]
\[
U^{(r)}_{s+1} = u_s(F^{(r)}_{-\tau}, \ldots, F^{(r)}_{s}) \text{ for } s = -\tau, \ldots, t.
\]
Corollary B.1. In the rectangular spiked model (4.1), suppose Assumption 2.4 holds for $W$. Suppose the initialization $U_{t-} \in \mathbb{R}^{m \times K}$ is independent of $W$, and $(U'_t, U_t) \overset{W_2}{\to} (U'_t, U_{t-})$ a.s. as $m, n \to \infty$. Suppose $u_t()$ and $v_t()$ are Lipschitz, and $\partial_t u_{t+1} (F_t^{(\tau)}, \ldots, F_t^{(\tau)})$ and $\partial_t v_t(G_t^{(\tau)}, \ldots, G_t^{(\tau)})$ exist and are continuous on a set of probability 1 under the above laws of $(F_t^{(\tau)}, \ldots, F_t^{(\tau)})$ and $(G_t^{(\tau)}, \ldots, G_t^{(\tau)})$ respectively. Then for any $T \geq 1$, a.s. as $m, n \to \infty$,

$$(U'_t, U_{t-}, \ldots, U_{T+1}, F_t^{(\tau)}, \ldots, F_T^{(\tau)}) \overset{W_2}{\to} (U'_t, U_{t-}, \ldots, U_{T+1}, F_{t-}, \ldots, F_T)$$

$$(V'_t, V_{t-}, \ldots, V_T, G_t^{(\tau)}, \ldots, G_T) \overset{W_2}{\to} (V'_t, V_{t-}, \ldots, V_T, G_{t-}, \ldots, G_T)$$

Proof. The proof is similar to that of Corollary A.1: Recalling $X = \frac{1}{\sqrt{m n}} U'_s S' V_s^T + W$, we write the iterations (B.1) as

$$G_t^{(\tau)} = \frac{1}{\sqrt{m n}} V'_s S' V_s'^T U_t^{(\tau)} + W^T U_t^{(\tau)} - \sum_{s=-\tau}^{t-1} V_s^{(\tau)} u_s^{(\tau)}$$

$$F_t^{(\tau)} = \frac{1}{\sqrt{m n}} U'_s S' V_s'^T V_t^{(\tau)} + W V_t^{(\tau)} - \sum_{s=-\tau}^{t} U_s^{(\tau)} a_s^{(\tau)}.$$

Then, approximating $S' U_t'^T U_t^{(\tau)}/\sqrt{m n}$ by $\sqrt{m} S' E [U_t'^T U_t^{(\tau)}] = \nu_t^{(\tau)T}$ and $S' V_s'^T V_t^{(\tau)}/\sqrt{m n}$ by $(1/\sqrt{\tau}) S' E [V_t'^T V_t^{(\tau)}] = \mu_t^{(\tau)T}$, we consider an alternative AMP sequence with the same initialization $\hat{U}_{t-} = U_{t-}$ and side information $U'_s$ and $V'_s$, defined by

$$\hat{Z}_t = W^T U_t - \sum_{s=-\tau}^{t-1} \hat{V}_s b_s^{t, s}, \quad \hat{G}_t = \hat{Z} + V'_s \nu_s^{(\tau)T}, \quad \hat{V}_t = \hat{v}_t (\hat{Z}_{t-}, \ldots, \hat{Z}_t, V'_s) \overset{\text{def}}{=} v_t (\hat{G}_{-\tau}, \ldots, \hat{G}_t),$$

$$\tilde{Y}_t = W \tilde{V}_t - \sum_{s=-\tau}^{t} \tilde{U}_s a_s^{t, s}, \quad \tilde{F}_t = \tilde{Y}_t + U'_s \mu_s^{(\tau)T}, \quad \tilde{U}_{t+1} = u_{t+1} (\tilde{Y}_{-\tau}, \ldots, \tilde{Y}_t, U'_s) \overset{\text{def}}{=} u_{t+1} (\tilde{F}_{-\tau}, \ldots, \tilde{F}_t).$$

We may apply Theorem 2.6 to analyze this AMP algorithm, together with an inductive argument as in [Fan20, Theorem 3.4(a)] to show

$$n^{-1} \| G_t^{(\tau)} - \hat{G}_t \|_{F}^2 \to 0, \quad n^{-1} \| V_t^{(\tau)} - \hat{V}_t \|_{F}^2 \to 0, \quad m^{-1} \| F_t^{(\tau)} - \tilde{F}_t \|_{F}^2 \to 0, \quad m^{-1} \| U_{t+1}^{(\tau)} - \tilde{U}_{t+1} \|_{F}^2 \to 0$$

for each fixed $t \geq -\tau$, which implies this corollary. \hfill \Box

Now we specialize the auxiliary AMP algorithm in (B.1) to the two-phased algorithm where

$$v_t (G_{-\tau}, \ldots, G_t^{(\tau)}) = \begin{cases} G_t^{(\tau)} S_u^{-1} & -\tau \leq t \leq 0, \\ v_t (G_1^{(\tau)}, \ldots, G_t^{(\tau)}) & t \geq 1 \end{cases} \quad \text{(B.4)}$$

and

$$u_{t+1} (F_{-\tau}, \ldots, F_t^{(\tau)}) = \begin{cases} F_t^{(\tau)} S_u^{-1} & -\tau \leq t \leq 0, \\ u_{t+1} (F_0^{(\tau)}, \ldots, F_t^{(\tau)}) & t \geq 1 \end{cases} \quad \text{(B.5)}$$

This auxiliary algorithm is initialized at

$$U_{-\tau}^{(\tau)} = (u_{-\tau}^1, \ldots, u_{-\tau}^k) \quad \text{with} \quad u_{-\tau}^k = (\mu_{\text{pca}, k} u_k^* + \sqrt{1 - \mu_{\text{pca}, k}^2} \cdot y_k) / \theta_{u,k} \text{ for each } k = 1, \ldots, K,$$
where each $\mu_{\text{pca},k}$ is defined in Theorem 4.1, and $y_1, \ldots, y_K$ are independent standard Gaussian random vectors also independent of $W$.

Similar as in the symmetric case, for each $t \geq 1$, we adopt the following block decomposition of $\phi_{all,t}^{(\tau)}$:

$$
\phi_{all,t}^{(\tau)} = \begin{pmatrix} \phi_{-t}^{(\tau)} & \phi_t^{(\tau)} \end{pmatrix}, \quad \text{where } \phi_{-t}^{(\tau)} \in \mathbb{R}^{\tau K \times \tau K} \text{ and } \phi_t^{(\tau)} \in \mathbb{R}^{(t+1)K \times (t+1)K}.
$$

Due to the linear update rule for the first $\tau$ steps, we have $\phi_{-t}^{(\tau)} = 0$ and

$$
\phi_{t}^{(\tau)} = \begin{pmatrix} 0 & 0 \cdots & 0 & 0 \\ S_u^{-1} & 0 & \cdots & 0 \\ 0 & S_u^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_u^{-1} \end{pmatrix} \in \mathbb{R}^{\tau K \times \tau K}, \quad \phi_{t}^{(\tau)} = \begin{pmatrix} 0 & \cdots & 0 & S_u^{-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{(t+1)K \times (t+1)K}. \quad (B.6)
$$

Similarly, we write

$$
p_{all,t}^{(\tau)} = \begin{pmatrix} p_{-t}^{(\tau)} & p_t^{(\tau)} \end{pmatrix}, \quad \text{where } p_{-t}^{(\tau)} = 0 \text{ and } p_t^{(\tau)} = \begin{pmatrix} S_v^{-1} & 0 \cdots & 0 \\ 0 & S_v^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_v^{-1} \end{pmatrix} \in \mathbb{R}^{\tau K \times \tau K}, \quad p_t^{(\tau)} = \begin{pmatrix} S_v^{-1} & 0 \cdots & 0 \\ \langle \partial_1 V_1 \rangle & \cdots & 0 \\ \vdots & \vdots & \ddots \vdots \\ \langle \partial_1 V_1 \rangle & \cdots & \langle \partial_1 V_1 \rangle \end{pmatrix} \in \mathbb{R}^{(t+1)K \times (t+1)K}. \quad (B.7)
$$

We first establish some important properties of $p_{all,t}^{(\tau)}$ and $\phi_{all,t}^{(\tau)}$ that will be useful in the analysis. By the block decomposition, we have

$$
p_{all,t}^{(\tau)} \phi_{all,t}^{(\tau)} = \begin{pmatrix} p_{-t}^{(\tau)} \phi_{-t}^{(\tau)} & p_t^{(\tau)} \phi_t^{(\tau)} \end{pmatrix}, \quad \phi_{all,t}^{(\tau)} p_{all,t}^{(\tau)} = \begin{pmatrix} \phi_{-t}^{(\tau)} p_{-t}^{(\tau)} & \phi_t^{(\tau)} p_t^{(\tau)} \end{pmatrix},
$$

where, recalling $S_u S_v = S^2$,

$$
p_{-t}^{(\tau)} \phi_{-t}^{(\tau)} = \phi_{-t}^{(\tau)} p_{-t}^{(\tau)} = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ S^{-2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \\ \end{pmatrix}, \quad p_t^{(\tau)} \phi_t^{(\tau)} = \phi_t^{(\tau)} p_t^{(\tau)} = \begin{pmatrix} 0 & \cdots & 0 & S^{-2} \\ 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}.
$$

Thus applying Lemma D.2 with $A = p_{all,t}^{(\tau)} \phi_{all,t}^{(\tau)}$ and $B = S^{-2}$ yields, for any $r \in \{0, \ldots, t\}$ and $c \in \{1, \ldots, \tau\}$,

$$
(p_{all,t}^{(\tau)} \phi_{all,t}^{(\tau)})^j [r, -c] = (p_t^{(\tau)} \phi_t^{(\tau)})^{j-c} [r, 0] S^{-2c} 1 \{1 \leq c \leq j\}. \quad (B.8)
$$
Similarly, we also have

\[(\phi_{\text{all},t}^{(\tau)} \psi_{\text{all},t}^{(\tau)})^j[r, -c] = (\phi_t^{(\tau)} \psi_t^{(\tau)})^{j-c}[r, 0] S^{-2c} \mathbb{1} \{1 \leq c \leq j\}\]

Moreover, we can also establish analogous properties of \((\psi_{\text{all},t}^{(\tau)} \phi_{\text{all},t}^{(\tau)})^j \psi_{\text{all},t}^{(\tau)}\) and \((\phi_{\text{all},t}^{(\tau)} \psi_{\text{all},t}^{(\tau)})^j \phi_{\text{all},t}^{(\tau)}\):

\[
(\psi_{\text{all},t}^{(\tau)} \phi_{\text{all},t}^{(\tau)})^j \psi_{\text{all},t}^{(\tau)}[r, -c] = \sum_{i=-\tau}^{t} (\psi_{\text{all},t}^{(\tau)} \phi_{\text{all},t}^{(\tau)})^j[r, i] \psi_{\text{all},t}^{(\tau)}[i, -c]
\]

\[
= (\psi_{\text{all},t}^{(\tau)} \phi_{\text{all},t}^{(\tau)})^j[r, -c] \psi_{\text{all},t}^{(\tau)}[-c, -c]
\]

\[
= (\psi_{\text{all},t}^{(\tau)} \phi_{\text{all},t}^{(\tau)})^{j-c}[r, 0] S^{-2c} S^{-1}_u \mathbb{1} \{1 \leq c \leq j\}
\]

\[
= (\psi_{\text{all},t}^{(\tau)} \phi_{\text{all},t}^{(\tau)})^{j-c} \psi_{\text{all},t}^{(\tau)}[r, 0] S^{-2c} \mathbb{1} \{1 \leq c \leq j\}
\]

where the second and fourth equalities follow from (B.7), and the third equality is due to (B.8). Similarly,

\[
(\phi_{\text{all},t}^{(\tau)} \psi_{\text{all},t}^{(\tau)})^j \phi_{\text{all},t}^{(\tau)}[r, -c] = \sum_{i=-\tau}^{t} (\phi_{\text{all},t}^{(\tau)} \psi_{\text{all},t}^{(\tau)})^j[r, i] \phi_{\text{all},t}^{(\tau)}[i, -c]
\]

\[
= (\phi_{\text{all},t}^{(\tau)} \psi_{\text{all},t}^{(\tau)})^j[r, -c + 1] \phi_{\text{all},t}^{(\tau)}[-c + 1, c]
\]

\[
= (\phi_{\text{all},t}^{(\tau)} \psi_{\text{all},t}^{(\tau)})^{j-c+1}[r, 0] S^{-2(c-1)} S^{-1}_u \mathbb{1} \{1 \leq c \leq j\}
\]

\[
= (\phi_{\text{all},t}^{(\tau)} \psi_{\text{all},t}^{(\tau)})^{j-c} \phi_{\text{all},t}^{(\tau)}[r, 0] S^{-2(c-1)} S^{-1}_u \mathbb{1} \{1 \leq c \leq j\}
\]

\[
= (\phi_{\text{all},t}^{(\tau)} \psi_{\text{all},t}^{(\tau)})^{j-c} \phi_{\text{all},t}^{(\tau)}[r, 0] S^{-2c} \mathbb{1} \{1 \leq c \leq j\}
\]

where the second and fourth equalities follow from (B.6) and (B.7), and the third equality is due to (B.9). All the above identities hold for \(\Psi_{\text{all},t}^{(\tau)}\) and \(\Phi_{\text{all},t}^{(\tau)}\) as well.

Finally, for the state evolution, we decompose

\[
\nu_{\text{all},t}^{(\tau)} = \begin{pmatrix} \nu_{\text{all},t}^{(\tau)} \\ \nu_{\text{all},t}^{(\tau)} \end{pmatrix}, \quad \mu_{\text{all},t}^{(\tau)} = \begin{pmatrix} \mu_{\text{all},t}^{(\tau)} \\ \mu_{\text{all},t}^{(\tau)} \end{pmatrix}
\]

\[
\Delta_{\text{all},t}^{(\tau)} = \begin{pmatrix} \Delta_{\text{all},t}^{(\tau)} \\ \Delta_{\text{all},t}^{(\tau)} \end{pmatrix}, \quad \Phi_{\text{all},t}^{(\tau)} = \begin{pmatrix} \Phi_{\text{all},t}^{(\tau)} \\ \Phi_{\text{all},t}^{(\tau)} \end{pmatrix}, \quad \Sigma_{\text{all},t}^{(\tau)} = \begin{pmatrix} \Sigma_{\text{all},t}^{(\tau)} \\ \Sigma_{\text{all},t}^{(\tau)} \end{pmatrix}
\]

\[
\Gamma_{\text{all},t}^{(\tau)} = \begin{pmatrix} \Gamma_{\text{all},t}^{(\tau)} \\ \Gamma_{\text{all},t}^{(\tau)} \end{pmatrix}, \quad \Psi_{\text{all},t}^{(\tau)} = \begin{pmatrix} \Psi_{\text{all},t}^{(\tau)} \\ \Psi_{\text{all},t}^{(\tau)} \end{pmatrix}, \quad \Omega_{\text{all},t}^{(\tau)} = \begin{pmatrix} \Omega_{\text{all},t}^{(\tau)} \\ \Omega_{\text{all},t}^{(\tau)} \end{pmatrix}
\]

**B.2 Phase I: Linear AMP for rectangular matrices**

As in the symmetric case, we first establish the convergence of the iterates and the associated state evolution of the first \(\tau\) steps of the auxiliary AMP algorithm, in the limit as \(\tau \to \infty\). We again reindex these iterates as \(1, 2, \ldots, \tau\).

Specifically, let \(U_1 = (u_1^1, \ldots, u_1^{K})\) with each \(u_1^k = (\mu_{\text{pca},k} u_k^k + \sqrt{1 - \mu_{\text{pca},k}^2} y_k)/\theta_{u,k}\). We write \(F_0 = U_1 S_u\). Then the first \(\tau\) iterates of the above auxiliary AMP algorithm have the structure of
the following linear AMP:
\[
G_t = X^\top F_{t-1} S_u^{-1} - \gamma \sum_{j=1}^{t-1} \kappa_{2j} G_{t-j} S^{-2j},
\]
(B.12)
\[
F_t = X G_t S_v^{-1} - \sum_{j=1}^t \kappa_{2j} F_{t-j} S^{-2j}.
\]

Up to iterate \( \tau \), let \( \mu_\tau = (\mu_t)_{1 \leq t \leq \tau}, \nu_\tau = (\nu_t)_{1 \leq t \leq \tau} \), \( \Sigma_\tau = (\sigma_{st})_{1 \leq s,t \leq \tau} \) and \( \Omega_\tau = (\omega_{st})_{1 \leq s,t \leq \tau} \) be the parameters of the state evolution describing this linear AMP, where each \( \mu_t, \nu_t \in \mathbb{R}^{K \times K'} \) and \( \sigma_{st}, \omega_{st} \in \mathbb{R}^{K \times K} \). Then it follows from Corollary B.1 that
\[
(F_1, \ldots, F_\tau) \overset{W_2}{\rightarrow} \mu_\tau \cdot U_\tau' + (Y_1, \ldots, Y_\tau) \sim \mathcal{N}(0, \Sigma_\tau) \perp U_\tau',
\]
\[
(G_1, \ldots, G_\tau) \overset{W_2}{\rightarrow} \nu_\tau \cdot V_\tau' + (Z_1, \ldots, Z_\tau) \sim \mathcal{N}(0, \Omega_\tau) \perp V_\tau'.
\]

Recall
\[
\mu_{\text{pca}} = (\mu_{\text{pca},1}, \ldots, \mu_{\text{pca},K}) \in \mathbb{R}^{K \times K'}, \quad \nu_{\text{pca}} = (\nu_{\text{pca},1}, \ldots, \nu_{\text{pca},K}) \in \mathbb{R}^{K \times K'}.
\]

**Lemma B.2.** Under Assumptions 2.4 and 4.2(a) and (c), the following holds for the linear AMP algorithm (B.12):

(a) \( \lim_{t \to \infty} \limsup_{m,n \to \infty} \| F_t - F_{\text{pca}} \|_F / \sqrt{m} = 0 \) and \( \lim_{t \to \infty} \limsup_{m,n \to \infty} \| G_t - G_{\text{pca}} \|_F / \sqrt{n} = 0 \) a.s.

(b) The state evolution satisfies \( \mu_t = \mu_{\text{pca}} \) and \( \nu_t = \nu_{\text{pca}} \) for every \( t \geq 1 \), and \( \lim_{\min(s,t) \to \infty} \sigma_{st} = \text{Id} - \mu_{\text{pca}} \mu_{\text{pca}} \top \) and \( \lim_{\min(s,t) \to \infty} \omega_{st} = \text{Id} - \nu_{\text{pca}} \nu_{\text{pca}} \top \).

**Proof.** Recall the \( K' \) largest sample singular values of \( X \) in (4.3) and the associated singular vectors in (4.4). We denote the remaining singular values and vectors as \( \lambda_i(X), f^i_{\text{pca}} \) and \( g^i_{\text{pca}} \) for \( i = K' + 1, \ldots, m \) in any order, with the same normalization that \( \| f^i_{\text{pca}} \| = \sqrt{m} \) and \( \| g^i_{\text{pca}} \| = \sqrt{n} \). Thus \( X g^i_{\text{pca}} / \sqrt{n} = \lambda_i(X) f^i_{\text{pca}} / \sqrt{m} \) and \( f^i_{\text{pca}} \top X / \sqrt{m} = \lambda_i(X) g^i_{\text{pca}} / \sqrt{n} \) for \( i = 1, \ldots, m \). Denote \( \| \Lambda \|_\infty = \lambda_+ \), fix a small constant \( \delta > 0 \), and define the event
\[
\mathcal{E}_{m,n} = \{ \lambda_i(X) \leq \| \Lambda \|_\infty + \delta \text{ for all } i = K' + 1, \ldots, m \}.
\]

Then \( \mathcal{E}_{m,n} \) occurs almost surely for all large \( (m,n) \) by Weyl’s singular value interlacing inequality.

Let \( f^k_t \) and \( g^k_t \) be the \( k \)th columns of the linear AMP iterates \( F_t \) and \( G_t \). For part (a), it suffices to show that
\[
\lim_{t \to \infty} \limsup_{m,n \to \infty} \frac{\| f^k_t - f^k_{\text{pca}} \|}{\sqrt{m}} + \frac{\| g^k_t - g^k_{\text{pca}} \|}{\sqrt{n}} = 0
\]
for each \( k = 1, \ldots, K \). Fixing any such \( k \), by the definition of linear AMP in (B.12), we have
\[
g^k_t = \frac{1}{\theta_{u,k}} \cdot X \top f^k_{t-1} - \gamma \sum_{j=1}^{t-1} \kappa_{2j} g^k_{t-j}, \quad f^k_t = \frac{1}{\theta_{v,k}} \cdot X g^k_t - \sum_{j=1}^t \kappa_{2j} f^k_{t-j}.
\]
(B.14)

We first show that the component of \( f^k_t \) orthogonal to \( f^k_{\text{pca}} \) and that of \( g^k_t \) orthogonal to \( g^k_{\text{pca}} \) vanish a.s. in the limits \( t \to \infty \) and \( m, n \to \infty \). Note that this linear AMP update ensures that \( g^k_t \)
is always in the span of \( \{ g^i \}_{i=1}^{m} \). For each \( t \geq 0 \) and \( i \in \{ 1, \ldots, m \} \), define \( \ell_t^{k,i} = (= \mathbf{f}_m^i) / m \).

For each \( t \geq 1 \) and \( i \in \{ 1, \ldots, m \} \), define \( r_t^{k,i} = (g^i)^T g^k / n \). We further extend these definitions by setting \( r_t^{k,i} \) for \( t = 0 \) and all \( i \in \{ 1, \ldots, m \} \). Then applying (B.14), for all \( i \in \{ 1, \ldots, m \} \),

\[
\ell_t^{k,i} = \frac{1}{\theta_{o,k}} \cdot \mathbf{f}_m^i \mathbf{x}^k_m \left\| \mathbf{f}_m^i \mathbf{x}^k_m \right\| \sum_{j=1}^{t} \frac{\kappa_{2j}^{k,i}}{\theta_{k}^{j}} - \sum_{j=1}^{t} \frac{\kappa_{2j}^{k,i}}{\theta_{k}^{j}} \ell_{t-j}^{k,i}
\]

Similarly, for all \( i \in \{ 1, \ldots, m \} \),

\[
r_t^{k,i} = \frac{1}{\theta_{u,k}} \cdot \mathbf{g}_m^i \mathbf{x}^k_m \left\| \mathbf{g}_m^i \mathbf{x}^k_m \right\| \sum_{j=1}^{t} \frac{\kappa_{2j}^{k,i}}{\theta_{k}^{j}} - \sum_{j=1}^{t} \frac{\kappa_{2j}^{k,i}}{\theta_{k}^{j}} r_{t-j}^{k,i}
\]

where the second equality applies our convention \( r_0^{k,i} = 0 \). First, for any \( i \in \{ 1, \ldots, K' \} \setminus \{ k \} \), by the initialization \( f_0 = \theta_{u,k} \mathbf{u}^k = \mu_{\text{pca},k} \mathbf{u}^k + \sqrt{1 - \mu_{\text{pca},k}^2} \cdot \mathbf{y}_k \), we have

\[
\ell_0^{k,i} = \frac{\mu_{\text{pca},k} \mathbf{f}_m^i \mathbf{u}^k}{m} + \frac{\sqrt{1 - \mu_{\text{pca},k}^2}}{m} \mathbf{f}_m^i \mathbf{y}_k \rightarrow 0
\]

a.s. as \( m, n \rightarrow \infty \), where the first term converges to 0 by Theorem 4.1, and the second term converges to 0 since \( \mathbf{f}_m^i \mathbf{x}^k_m / m \sim \mathcal{N}(0, 1/m) \). Thus, it follows from the recursions (B.15) and (B.16) that

\[
\lim_{m,n \rightarrow \infty} \ell_t^{k,i} = \lim_{m,n \rightarrow \infty} r_t^{k,i} = 0 \text{ a.s. for each fixed } t \geq 0 \text{ and } i \in \{ 1, \ldots, K' \} \setminus \{ k \}.
\]

Next, for each \( i \in \{ K' + 1, \ldots, m \} \), consider a space \( \mathcal{X} \) of bounded infinite-dimensional vectors with elements in \( [0, \infty) \). For each \( t \geq 1 \), we define two elements \( \phi_{t,i}^{k,i}, \varphi_{t,i}^{k,i} \in \mathcal{X} \) as

\[
\phi_{t,i}^{k,i} = (\ell_t^{k,i}, r_t^{k,i}, \ldots, \ell_0^{k,i}, r_0^{k,i}, 0, \ldots), \quad \varphi_{t,i}^{k,i} = (r_t^{k,i}, \ell_{t-1}^{k,i}, r_{t-1}^{k,i}, \ldots, \ell_0^{k,i}, r_0^{k,i}, 0, \ldots).
\]

Let \( \iota \in (0, 1) \) be chosen as in Assumption 4.2(c), and let us consider a norm \( \| \cdot \| \) on \( \mathcal{X} \) defined by

\[
\|(x_0, x_{-1}, x_{-2}, \ldots)\| = \sup_{k \geq 0} |x_k| \cdot \iota^k.
\]

Consider a map \( g : \mathcal{X} \rightarrow \mathcal{X} \) defined as \( g(\phi_{t-1}) = \varphi_{t,i}^{k,i} \) and another map \( h : \mathcal{X} \rightarrow \mathcal{X} \) given by \( h(\varphi_{t,i}^{k,i}) = \mathbf{g}_m^{i} \). Then we have \( \mathbf{g}_m^{i} = (h \circ g)(\mathbf{g}_m^{i-1}) \). We verify that both \( g \) and \( h \) are contractive with
respect to the norm \( \| \cdot \| \) on \( \mathcal{X} \). Let \( \{(\vartheta_t, \varphi_t)\}_{t \geq 1} \) and \( \{({\tilde{\vartheta}}_t, {\tilde{\varphi}}_t)\}_{t \geq 1} \) be two sequences given by

\[
\begin{align*}
\vartheta_t &= (\ell_t, r_t, \ldots, \ell_0, r_0, 0, \ldots) \\
\varphi_t &= (r_t, \ell_{t-1}, r_{t-1}, \ldots, \ell_0, r_0, 0, \ldots)
\end{align*}
\]

\[
\begin{align*}
{\tilde{\vartheta}}_t &= (\ell_t, {\tilde{\ell}}_t, \ldots, \ell_0, {\tilde{r}}_0, 0, \ldots) \\
{\tilde{\varphi}}_t &= (r_t, {\tilde{\ell}}_{t-1}, {\tilde{r}}_{t-1}, \ldots, \ell_0, {\tilde{r}}_0, 0, \ldots)
\end{align*}
\]

where both \( \{(\ell_t, r_t)\}_{t \geq 1} \) and \( \{(\ell_t, \tilde{r}_t)\}_{t \geq 1} \) satisfy the same recursions as in (B.15) and (B.16). Note that

\[
\|g(\vartheta_{t-1}) - g(\tilde{\vartheta}_{t-1})\| = \|\varphi_t - \tilde{\varphi}_t\| = \max \left\{ |r_t - \tilde{r}_t|, \max_{1 \leq j \leq t} |\ell_{t-j} - \tilde{\ell}_{t-j}|2^{j-1}, \max_{1 \leq j \leq t} |r_{t-j} - \tilde{r}_{t-j}|2^j \right\} \\
= \max \{|r_t - \tilde{r}_t|, t \cdot \|\vartheta_{t-1} - \tilde{\vartheta}_{t-1}\|\}.
\]

Then we need to control \(|r_t - \tilde{r}_t|\). It follows from (B.16) that

\[
|r_t - \tilde{r}_t| = \frac{\lambda_t(X)\sqrt{\gamma}}{\theta_{u,k}}(\ell_{t-1} - \tilde{\ell}_{t-1}) - s \sum_{j=1}^{t} \frac{\kappa_{2j}}{2^3t^2} (r_{t-j} - \tilde{r}_{t-j}) \\
\leq \frac{\lambda_t(X)\sqrt{\gamma}}{\theta_{u,k}} |\ell_{t-1} - \tilde{\ell}_{t-1}| + s \sum_{j=1}^{t} \frac{\kappa_{2j}}{2^3t^2} |r_{t-j} - \tilde{r}_{t-j}|2^{j-1} \\
\leq \left( \frac{\lambda_t(X)\sqrt{\gamma}}{\theta_{u,k}} + s \sum_{j=1}^{t} \frac{\kappa_{2j}}{2^3t^2} \right) \cdot \|\vartheta_{t-1} - \tilde{\vartheta}_{t-1}\|.
\]

where the last inequality holds on the above event \( \mathcal{E}_{m,n} \). For sufficiently small \( \delta \), we have \( \eta_1 \in (0,1) \) by Assumption 4.2(c). Similarly,

\[
\|h(\varphi_t) - h(\tilde{\varphi}_t)\| = \|\varphi_t - \tilde{\varphi}_t\| = \max \left\{ |\ell_t - \tilde{\ell}_t|, \max_{1 \leq j \leq t} |\ell_{t-j} - \tilde{\ell}_{t-j}|2^{j-1}, \max_{0 \leq j \leq t} |r_{t-j} - \tilde{r}_{t-j}|2^j \right\} \\
= \max \{|\ell_t - \tilde{\ell}_t|, t \cdot \|\varphi_t - \tilde{\varphi}_t\|\}.
\]

By (B.15), we have on the event \( \mathcal{E}_{m,n} \) that

\[
|\ell_t - \tilde{\ell}_t| = \frac{\lambda_t(X)}{\theta_{v,k}\sqrt{\gamma}} (r_t - \tilde{r}_t) - s \sum_{j=1}^{t} \frac{\kappa_{2j}}{2^3t^2} (r_{t-j} - \tilde{r}_{t-j}) \\
\leq \frac{\lambda_t(X)}{\theta_{v,k}\sqrt{\gamma}} |r_t - \tilde{r}_t| + s \sum_{j=1}^{t} \frac{\kappa_{2j}}{2^3t^2} |r_{t-j} - \tilde{r}_{t-j}|2^{j-1} \\
\leq \left( \frac{\|A\|_{\infty} + \delta}{\theta_{v,k}\sqrt{\gamma}} + \sum_{j=1}^{\infty} \frac{\kappa_{2j}}{2^3t^2} \right) \cdot \|\varphi_t - \tilde{\varphi}_t\|.
\]
We also have $\eta_2 \in (0,1)$ for sufficiently small $\delta$ by Assumption 4.2(c). Combining (B.18), (B.19), (B.20) and (B.21), we get

$$
\| (h \circ g)(\hat{\mathbf{t}}_{t-1}) - (h \circ g)(\hat{\mathbf{t}}_{t-1})_0 \| \leq \max\{\eta_1, t\} \cdot \max\{\eta_2, t\} \| \mathbf{t}_{t-1} - \hat{\mathbf{t}}_{t-1} \|.
$$

Therefore, $h \circ g$ is a $\rho$-contraction for some $\rho \in (0,1)$. Applying this property to $\{\mathbf{t}_t^{k,i}\}_{t \geq 1}$ yields

$$
\| \ell_t^{k,i} \| + \| r_t^{k,i} \| \leq \frac{1}{t} \| (\ell_t^{k,i}, r_t^{k,i}, \ldots, \ell_t^{k,i}, r_t^{k,i}, 0 \ldots) - (0, 0, \ldots) \|
$$

$$
\leq \frac{\rho}{t} \| (\ell_t^{k,i}, r_t^{k,i}, 0 \ldots) - (0, 0, \ldots) \|$

$$
= \frac{\rho}{t} (|\ell_t^{k,i}| + |r_t^{k,i}|).
$$

(B.22)

This holds simultaneously for all $i \in \{K' + 1, \ldots, m\}$ on the event $\mathcal{E}_m,n$.

Now write $\ell_t^k = c_t^k f_{t, \text{pca}}^k + \ell_t^k$ and $\mathbf{g}_t^k = c_t^k \mathbf{g}_{t, \text{pca}}^k + r_t^k$ where $\ell_t^k \perp f_{t, \text{pca}}^k$ and $r_t^k \perp g_{t, \text{pca}}^k$. Recalling that $\mathbf{g}_t^k$ belongs to the span of $\{f_{t, \text{pca}}^i\}_{i=1,\ldots,m}$, we can expand $\ell_t^k$ and $r_t^k$ as

$$
\ell_t^k = \sum_{i=1,i \neq k}^m \frac{f_{t, \text{pca}}^i \top f_{t, \text{pca}}^k m}{f_{t, \text{pca}}^k} \quad \text{and} \quad r_t^k = \sum_{i=1,i \neq k}^m \frac{g_{t, \text{pca}}^i \top g_{t, \text{pca}}^k n}{g_{t, \text{pca}}^k}.
$$

Thus, we further have

$$
\frac{\| \ell_t^k \|^2}{m} + \frac{\| r_t^k \|^2}{n} = \sum_{i=1,i \neq k}^K I_1 \left( (\ell_t^{k,i})^2 + (r_t^{k,i})^2 \right) + \sum_{i=K'+1}^m \left( (\ell_t^{k,i})^2 + (r_t^{k,i})^2 \right).
$$

By (B.17), we have $\lim_{n \to \infty} I_1 = 0$. For $I_2$, it follows from (B.22) and the definition $r_t^{0,k} \equiv 0$ that

$$
I_2 \leq \frac{2\rho^2}{t^2} \sum_{i=K'+1}^m \left( (\ell_t^{0,i})^2 + (r_t^{0,i})^2 \right) \leq \frac{2\rho^2}{t^2} \cdot \frac{\| \ell_t^0 \|^2}{m} \leq \frac{2\rho^2}{t^2} \cdot \frac{\| r_t^0 \|^2}{m}.
$$

By the initialization $\mathbf{f}_0^k = \mu_{\text{pca},k} \mathbf{u}_*^k + \sqrt{1 - \mu_{\text{pca},k}^2} \cdot \mathbf{y}_k$, we have $\lim_{m,n \to \infty} \| \mathbf{f}_0^k \|^2 / m = 1$. Therefore, it follows that

$$
\lim_{t \to \infty} \limsup_{m,n \to \infty} \frac{\| \ell_t^k \|}{\sqrt{m}} + \frac{\| r_t^k \|}{\sqrt{n}} = 0.
$$

(B.23)

Next we show that as $t \to \infty$, $c_t^k \stackrel{\text{def}}{=} \left( f_{t, \text{pca}}^k / \tau \right)^k$ and $s_t^k \stackrel{\text{def}}{=} (\mathbf{g}_{t, \text{pca}}^k) / n \to 1$ by using the state evolution (B.13) of the linear AMP. For any $t \geq 1$, we have

$$
\nu_{t+1} = \mathbb{E}[[U_{t+1} U_*^T] S^T S^T = S_u^{-1} \mathbb{E}[F_{t+1} U_*^T] S^T S^T = S_u^{-1} \mathbb{E}[(\mu U_* + Y_0) U_*^T] S^T S^T = S_u^{-1} \mu_t S^T S^T
$$

where the last equality is due to $Y_0 \perp U_*'$, $\mathbb{E}[Y_0] = 0$, and $\mathbb{E}[U_*' U_*^T] = \text{Id}$. Setting $U_1 = S_u^{-1} F_0$, this holds also for $t = 0$ upon identifying $\mu_0 = \mu_{\text{pca}}$ to match the initialization of $\mathbf{F}_0$. We also have, for $t \geq 1$,

$$
\mu_t = \frac{1}{\sqrt{\gamma}} \mathbb{E}[V_t V_*^T] S^T = \frac{1}{\sqrt{\gamma}} S_u^{-1} \mathbb{E}[G_t V_*^T] S^T = \frac{1}{\sqrt{\gamma}} S_u^{-1} \mathbb{E}[(\nu_t V_*' + Z_t) V_*^T] S^T = \frac{1}{\sqrt{\gamma}} S_u^{-1} \nu_t S^T
$$

49
where the last equality is due to $Z_t \perp V_t$, $\mathbb{E}[Z_t] = 0$, and $\mathbb{E}[V_tV_t^\top] = \text{Id}$. Combining the above two equalities yields

$$
\mu_{t+1} = S_v^{-1}S_u^{-1}\mu_t(S_t')^2.
$$

Since $\mu_0 = \mu_{\text{pca}}$ and $S_uS_v = S^2$, we have $\mu_t = \mu_{\text{pca}}$ for all $t \geq 1$. Thus also $\nu_t \equiv S_u^{-1}\mu_{\text{pca}}S_t\sqrt{\tau} = \nu_{\text{pca}}$ for all $t \geq 1$, by (4.9) and (4.10). Hence, for each $k \in \{1, \ldots, K\}$, we have

$$
\mu_{\text{pca},k} = \lim_{m,n \to \infty} \frac{F_{\text{pca}}^k}{m} u_*^k \quad \text{and} \quad \mu_{\text{pca},k} = \lim_{m,n \to \infty} \frac{f_t^k}{m} u_*^k = \lim_{m,n \to \infty} \frac{\xi_t^k (F_{\text{pca}}^k) u_*^k}{m} + \frac{(f_t^k) u_*^k}{m},
$$

where the left equality follows from Theorem 4.1 and the right equality applies $\mu_t = \mu_{\text{pca}}$. This further implies

$$
\limsup_{m,n \to \infty} \left| (\xi_t^k - 1) \frac{f_t^k}{m} u_*^k \right| \leq \limsup_{m,n \to \infty} \frac{\|f_t^k\|}{\sqrt{m}}.
$$

Since $\lim_{m,n \to \infty} F_{\text{pca}}^k u_*^k / m = \mu_{\text{pca},k} \neq 0$ a.s. by Theorem 4.1, taking the limit as $t \to \infty$ on both sides, it follows from (B.23) that

$$
\lim \limsup_{m,n \to \infty} |\xi_t^k - 1| = 0.
$$

Combining with (B.23), we have

$$
\lim \limsup_{t \to \infty} \frac{\|f_t^k - F_{\text{pca}}^k\|}{\sqrt{m}} = 0.
$$

Applying the same argument, we also get $\lim_{t \to \infty} \limsup_{m,n \to \infty} \|g_t^k - g_{\text{pca}}^k\|/\sqrt{n} = 0$.

Finally, the convergence of $\sigma_{st}$ and $\omega_{st}$ can be derived using the exact same argument as we have used for the linear AMP for symmetric matrices. This completes the proof.

Returning to the auxiliary AMP iterations $U_{\tau}^{(\tau)}$, $G_{\tau}^{(\tau)}$, $V_{\tau}^{(\tau)}$, $F_{\tau}^{(\tau)}$, . . . defined by (B.4) and (B.5), this linear AMP algorithm characterizes these iterates up to $U_1^{(\tau)}$, $G_1^{(\tau)}$. Thus, translating Lemma B.2 back to this indexing and recalling the spectral initializations $F_0 = F_{\text{pca}}$, $G_0 = G_1 = G_{\text{pca}}$, $U_0 = U_1 = F_{\text{pca}}S_u^{-1}$, and $V_0 = G_{\text{pca}}S_v^{-1}$, we have

$$
\lim \limsup_{\tau \to \infty} m,n \to \infty \frac{\|F_{\tau}^{(\tau)} - F_0\|}{\sqrt{m}} = 0, \quad \lim \limsup_{\tau \to \infty} m,n \to \infty \frac{\|V_{\tau}^{(\tau)} - V_0\|}{\sqrt{n}} = 0,
$$

$$
\mu_{\tau} = \mu_{\text{pca}}, \quad \lim_{\tau \to \infty} \gamma_{\tau} = \text{Id} - \mu_{\text{pca}}M_{\text{pca}}^T, \quad \text{for all fixed } i, j \leq 0,
$$

and

$$
\lim \limsup_{\tau \to \infty} m,n \to \infty \frac{\|G_{\tau}^{(\tau)} - G_0\|}{\sqrt{n}} = 0, \quad \lim \limsup_{\tau \to \infty} m,n \to \infty \frac{\|G_{\tau}^{(\tau)} - G_1\|}{\sqrt{n}} = 0,
$$

$$
\lim \limsup_{\tau \to \infty} m,n \to \infty \frac{\|U_{\tau}^{(\tau)} - U_0\|}{\sqrt{m}} = 0, \quad \lim \limsup_{\tau \to \infty} m,n \to \infty \frac{\|U_{\tau}^{(\tau)} - U_1\|}{\sqrt{m}} = 0,
$$

$$
\nu_{\tau} = \nu_{\text{pca}}, \quad \lim_{\tau \to \infty} \Omega_{\tau} = \text{Id} - \nu_{\text{pca}}M_{\text{pca}}^T, \quad \text{for all fixed } i, j \leq 1.
$$



50
B.3 Phase II: auxiliary AMP for rectangular matrices

Proof of Theorem 4.3. We show by induction that the following statements hold a.s. for each \( t \geq 0 \).

\((G.a)\) \( \lim_{t \to \infty} \limsup_{m,n \to \infty} \|U_{t}^{(r)} - U_{t}\|_{F}/\sqrt{m} = 0 \) and \( \lim_{m,n \to \infty} \|U_{t}\|_{F}/\sqrt{m} < C_{t} \) for a constant \( C_{t} > 0 \).

\((G.b)\) \( (U'_{r}, U_{0}, U_{1}, \ldots, U_{t}, F_{0}, \ldots, F_{t-1}) \xrightarrow{W_{2}} (U'_{r}, U_{0}, U_{1}, \ldots, U_{s}, F_{0}, \ldots, F_{t-1}) \), where the limiting distribution is defined in (4.15).

\((G.c)\) \( \lim_{t \to \infty} \Phi_{t}^{(r)} = \Phi_{t} \) and \( \lim_{t \to \infty} \Delta_{t}^{(r)} = \Delta_{t} \).

\((G.d)\) \( \lim_{t \to \infty} \lim_{m,n \to \infty} \|\phi_{t}^{(r)} - \phi_{t}\| = 0 \).

\((G.e)\) \( \lim_{t \to \infty} \limsup_{m,n \to \infty} \|G_{t}^{(r)} - G_{t}\|_{F}/\sqrt{n} = 0 \) and \( \lim_{m,n \to \infty} \|G_{t}\|_{F}/\sqrt{n} < C_{t} \) for a constant \( C_{t} > 0 \).

\((G.f)\) \( \lim_{t \to \infty} \Omega_{t}^{(r)} = \Omega_{t} \) and \( \lim_{t \to \infty} \nu_{t}^{(r)} = \nu_{t} \).

and

\((F.a)\) \( \lim_{t \to \infty} \limsup_{m,n \to \infty} \|V_{t}^{(r)} - V_{t}\|/\sqrt{n} = 0 \) and \( \lim_{m,n \to \infty} \|V_{t}\|_{F}/\sqrt{n} < C_{t} \) for a constant \( C_{t} > 0 \).

\((F.b)\) \( (V'_{r}, V_{0}, V_{1}, \ldots, V_{t}, G_{0}, \ldots, G_{t}) \xrightarrow{W_{2}} (V'_{r}, V_{0}, V_{1}, \ldots, V_{t}, G_{0}, \ldots, G_{t}) \), where the limiting distribution is defined in (4.15).

\((F.c)\) \( \lim_{t \to \infty} \Psi_{t}^{(r)} = \Psi_{t} \) and \( \lim_{t \to \infty} \Gamma_{t}^{(r)} = \Gamma_{t} \).

\((F.d)\) \( \lim_{t \to \infty} \limsup_{m,n \to \infty} \|\psi_{t}^{(r)} - \psi_{t}\| = 0 \).

\((F.e)\) \( \lim_{t \to \infty} \limsup_{m,n \to \infty} \|F_{t}^{(r)} - F_{t}\|/\sqrt{m} = 0 \) and \( \limsup_{m,n \to \infty} \|F_{t}\|_{F}/\sqrt{m} < C_{t} \) for a constant \( C_{t} > 0 \).

\((F.f)\) \( \lim_{t \to \infty} \Sigma_{t}^{(r)} = \Sigma_{t} \) and \( \lim_{t \to \infty} \mu_{t}^{(r)} = \mu_{t} \).

Denote by \( t^{(G)}, t^{(F)} \) the claims of parts G and F at iteration \( t \). The base cases of \( 0^{(G)}, 0^{(F)} \) and \( 1^{(G)} \) may be verified from the implications (B.24) and (B.25) of Lemma B.2, similar to the proof of Theorem 3.3.

Let us take \( t \geq 1 \), suppose by induction that \( s^{(F)} \) holds for all \( 0 \leq s \leq t - 1 \) and \( s^{(G)} \) holds for all \( 0 \leq s \leq t \), and show \( t^{(F)} \). The proofs of \( t^{(F,a-F,d)} \) apply the same arguments as in the proof of Theorem 3.3, and we omit these for brevity. We provide below the proofs of \( t^{(F,e)} \) and \( t^{(F,f)} \).

Part (F.e) We control the difference between \( F_{t}^{(r)} \) and \( F_{t} \). By their definitions in (B.1) and (4.12), applying the triangle inequality yields

\[
\frac{\|F_{t}^{(r)} - F_{t}\|_{F}}{\sqrt{m}} \leq \frac{\|X\|}{\sqrt{m}} \frac{\|V_{t}^{(r)} - V_{t}\|}{\sqrt{m}} + \frac{1}{\sqrt{m}} \left\| \sum_{i=-r}^{t} U_{i}^{(r)} a_{i}^{(r)}_{t,i} - \sum_{i=0}^{t} U_{i} a_{i}^{T}_{t,i} \right\|_{F}.
\]
The first term vanishes in the limit \( m, n \to \infty \) by \( \ell^{(F,a)} \), as \( \lim_{m,n \to \infty} \|X\| \) is finite. To show the second term will also converge to zero, we may write analogously to (A.17)

\[
\frac{1}{\sqrt{m}} \left\| \sum_{i=-\tau}^{t} U_i^{(r)} a_{t,i}^T - \sum_{i=0}^{t} U_i a_{t,i}^T \right\|_F \leq \frac{1}{\sqrt{m}} \left( \sum_{i=-\tau}^{0} U_i^{(r)} a_{t,i}^T - U_0 a_{t,0}^T \right) + \sum_{i=1}^{\infty} \frac{1}{\sqrt{m}} \left\| U_i^{(r)} a_{t,i}^T - U_i a_{t,i}^T \right\|_F.
\]

The term \( I_2 \) converges to 0 by \( \ell^{(G,a)}, \ell^{(G,d)}, \ell^{(F,d)} \), and the same argument as in the proof of Theorem 3.3. For \( I_1 \), let us further write, analogously to (A.18),

\[
\frac{1}{\sqrt{m}} \left\| \sum_{i=-\tau}^{0} U_i^{(r)} a_{t,i}^T - U_0 a_{t,0}^T \right\|_F \leq \left\| \sum_{i=-\tau}^{0} a_{t,i}^T - a_{t,0} \right\|_F \frac{\|U_0\|_F}{\sqrt{m}} + \sum_{i=-\tau}^{\infty} \left\| a_{t,i}^T \right\| \frac{\|U_i^{(r)} - U_0\|_F}{\sqrt{m}}.
\]

The term \( I_{1,2} \) converges to 0 by (B.25) and Lemma D.1, following again the same argument as in the proof of Theorem 3.3. Thus it remains to verify the convergence of \( I_{1,1} \).

Recall our definition of debiasing coefficients in (4.14) and (B.2):

\[
a_{t,0} = \bar{A}_t[t,0] = \sum_{j=0}^{\infty} \psi_t(\phi_t \psi_t)^j[t,0] \odot \bar{\kappa}_{2(j+1)},
\]

\[
a_{t,i}^{(r)} = A_{all,t}[t,i] = \sum_{j=0}^{\infty} \kappa_{2(j+1)}(\phi_{all,t}^{(r)} \psi_{all,t}^{(r)})^j[t,i] = \sum_{j=-i}^{-i+t} \kappa_{2(j+1)}(\phi_t^{(r)} \psi_t^{(r)})^{j+i}[t,0] S_{2i}
\]

where the last equality comes from (B.10). Therefore, we have

\[
\lim_{\tau \to \infty} \lim_{m,n \to \infty} \sum_{i=-\tau}^{0} a_{t,i}^{(r)} - a_{t,0} = \lim_{\tau \to \infty} \lim_{m,n \to \infty} \sum_{j=-i}^{-i+t} \kappa_{2(j+1)}(\phi_t^{(r)} \psi_t^{(r)})^{j+i}[t,0] S_{2i} - a_{t,0}
\]

\[
= \sum_{j=0}^{t} \psi_t(\phi_t \psi_t)^j[t,0] \left( \sum_{i=0}^{\infty} \kappa_{2(j+i+1)} S_{2i} \right) - a_{t,0} = 0
\]

by the expression of \( a_{t,0} \) and the definition of \( \bar{\kappa}_{2(j+1)} \). This shows \( \ell^{(F,c)} \).

**Part (F.f) By the definition of \( \Sigma_{all,t}^{(r)} \) in (B.3) and its block decomposition, we have**

\[
\Sigma_{t}^{(r)} = \sum_{j=0}^{\infty} \kappa_{2(j+1)} \sum_{i=0}^{j} \left( \left( \phi_{all,t}^{(r)} \phi_{all,t}^{(r)} \right)^i \Gamma_{all,t}^{(r)} (\phi_{all,t}^{(r)} \psi_{all,t}^{(r)})^{j-i} \right)_{t}
\]

\[
+ \sum_{j=0}^{\infty} \kappa_{2(j+1)} \sum_{i=0}^{j-1} \left( \left( \phi_{all,t}^{(r)} \phi_{all,t}^{(r)} \right)^i \psi_{all,t}^{(r)} \Delta_{all,t}^{(r)} (\phi_{all,t}^{(r)} \psi_{all,t}^{(r)})^{j-i} \right)_{t}
\]

where \([ \cdot ]_t\) means taking the lower-right \((t+1)K \times (t+1)K\) submatrix. Similarly, we may decompose \( \Sigma_t \) defined in (4.17) as \( \Sigma_t = I_1 + I_2 \). To show \( \lim_{\tau \to \infty} \Sigma_{t}^{(r)} = \Sigma_t \), it suffices to show \( \lim_{\tau \to \infty} I_{1}^{(r)} = I_1 \) and \( \lim_{\tau \to \infty} I_{2}^{(r)} = I_2 \).
For $\mathcal{I}_1^{(r)}$, by the block decompositions of $\Phi_{all,t}^{(r)}$, $\Psi_{all,t}^{(r)}$, and $\Delta_{all,t}^{(r)}$, we obtain

\[
\mathcal{I}_1^{(r)} = \sum_{j=0}^{\infty} \kappa_{2(j+1)} \sum_{i=0}^{j} (\Psi_{all,t}^{(r)} \Phi_{all,t}^{(r)})^i \Gamma^{(r)} (-\Phi_{all,t}^{(r)T} \Phi_{all,t}^{(r)})^{j-i} + \sum_{j=0}^{\infty} \kappa_{2(j+1)} \sum_{i=0}^{j} (\Psi_{all,t}^{(r)} \Phi_{all,t}^{(r)})^i \Gamma^{(r)} (-\Psi_{all,t}^{(r)T} \Phi_{all,t}^{(r)})^{j-i} 
\]

\[
\sum_{j=0}^{\infty} \kappa_{2(j+1)} \sum_{i=0}^{j} (\Psi_{all,t}^{(r)} \Phi_{all,t}^{(r)})^i \Gamma^{(r)} (-\Phi_{all,t}^{(r)T} \Phi_{all,t}^{(r)})^{j-i} + \sum_{j=0}^{\infty} \kappa_{2(j+1)} \sum_{i=0}^{j} (\Psi_{all,t}^{(r)} \Phi_{all,t}^{(r)})^i \Gamma^{(r)} (-\Psi_{all,t}^{(r)T} \Phi_{all,t}^{(r)})^{j-i} 
\]

where we observe that the third term is the transpose of the second term. Recalling $\Gamma_t = \tilde{\Gamma}_t + \tilde{\Gamma}_t + \tilde{\Gamma}_t$, we correspondingly decompose

\[
\mathcal{I}_1^{(r)} = \sum_{j=0}^{\infty} (\Psi_t \Phi_t)^j \left( \tilde{\kappa}_{2(j+1)} \odot \tilde{\Gamma}_t - \tilde{\kappa}_{2(j+1)} \odot \tilde{\Gamma}_t - \tilde{\Gamma}_t + \kappa_{2(j+1)} \tilde{\Gamma}_t \right) (\Psi_t \Phi_t)^{j-i} 
\]

\[
\sum_{j=0}^{\infty} (\Psi_t \Phi_t)^j \left( \tilde{\kappa}_{2(j+1)} \odot \tilde{\Gamma}_t - \tilde{\kappa}_{2(j+1)} \odot \tilde{\Gamma}_t + \kappa_{2(j+1)} \tilde{\Gamma}_t \right) (\Psi_t \Phi_t)^{j-i} 
\]

\[
+ \sum_{j=0}^{\infty} \kappa_{2(j+1)} (\Psi_t \Phi_t)^j \Gamma_t (\Psi_t \Phi_t)^{j-i} 
\]

\[
\sum_{j=0}^{\infty} \kappa_{2(j+1)} (\Psi_t \Phi_t)^j \Gamma_t (\Psi_t \Phi_t)^{j-i} 
\]

\[
\lim_{\tau \to \infty} \mathbf{\tilde{S}}_t^{(r)}[r,c] = \tilde{\mathbf{S}}_t[r,c], \quad \lim_{\tau \to \infty} \left( \mathbf{\tilde{S}}_t^{(r)} + \mathbf{\tilde{S}}_t^{(r)T} \right)[r,c] = \tilde{\mathbf{S}}_t[r,c], \quad \lim_{\tau \to \infty} \mathbf{\tilde{\Sigma}}_t^{(r)}[r,c] = \mathbf{\Sigma}_t[r,c].
\]

**Convergence of $\tilde{\mathbf{S}}_t^{(r)}$.** First, for $\tilde{\mathbf{S}}_t^{(r)}$, we have

\[
\tilde{\mathbf{S}}_t^{(r)}[r,c] = \sum_{j=0}^{\infty} \kappa_{2(j+1)} \sum_{i=0}^{j} (\Psi_{all,t}^{(r)} \Phi_{all,t}^{(r)})^i (r,-\alpha) \Gamma^{(r)} [-\alpha,-\beta] (\Phi_{all,t}^{(r)T} \Psi_{all,t}^{(r)})^{j-i} [-\beta,c]
\]

\[
= \sum_{j=0}^{\infty} \kappa_{2(j+1)} \sum_{i=0}^{j} (\Psi_{all,t}^{(r)} \Phi_{all,t}^{(r)})^i (r,-\alpha) \Gamma^{(r)} [-\alpha,-\beta] S^{-2\beta} (\Phi_{all,t}^{(r)T} \Psi_{all,t}^{(r)})^{j-i} [-\beta,0]
\]

\[
= \sum_{p+q=0}^{\alpha} (\Psi_{all,t}^{(r)} \Phi_{all,t}^{(r)})^p (r,0) \left( \sum_{\alpha+\beta=1}^{\tau} \kappa_{2(p+q+\alpha+\beta+1)} S^{-2\alpha} \Gamma^{(r)} [-\alpha,-\beta] S^{-2\beta} (\Phi_{all,t}^{(r)T} \Psi_{all,t}^{(r)})^{q} [0,c] \right)
\]

where the second equality follows from (B.8) and the third equality re-indexes the summation by setting $p = i - \alpha$ and $q = j - i - \beta$. By following the same argument for the convergence of the
corresponding \( \hat{\Sigma}_t^{(r)} \) in the proof of Theorem 3.3, we obtain that

\[
\lim_{\tau \to \infty} \sum_{\alpha, \beta=1}^{\tau} \kappa_2(p+q+\alpha+\beta+1)S^{-2\alpha} \Gamma_t^{(r)} [-\alpha, -\beta] S^{-2\beta} = \sum_{\alpha, \beta=1}^{\infty} \kappa_2(p+q+\alpha+\beta+1)S^{-2\alpha} \Gamma_t[0,0] S^{-2\beta}
\]

where the second equality is due to the fact that \( \Gamma_t[0,0] = \mathbb{E}[V_0V_0^T] = S_{\nu}^{-2} \) is diagonal, and the third equality follows from the definitions of \( \hat{\kappa} \) and \( \tilde{\kappa} \) in (3.11). Combining the above limit with the convergence of \( \Psi_t^{(r)} \) and \( \Phi_t^{(r)} \) in \( t^{(G,c)} \) and \( t^{(F,c)} \), since \( \Gamma_t[0,0] = \hat{\Gamma}_t[0,0] \) by definition, we get

\[
\lim_{\tau \to \infty} \hat{\Sigma}_t^{(r)}[r,c] = \sum_{p,q=0}^{\infty} (\Psi_t\Phi_t)^p[r,0] (\hat{\kappa}_2(j+1) - 2\tilde{\kappa}_2(j+2) + \kappa_2(j+1)\text{Id}) \hat{\Gamma}_t[0,0](\Phi_t^T\Psi_t^T)^q[0,c] = \hat{\Sigma}_t[r,c].
\]

**Convergence of \( \hat{\Sigma}_t^{(r)} \).** Applying (B.8) and re-indexing the summation similarly as above,

\[
\hat{\Sigma}_t^{(r)}[r,c] = \sum_{j=0}^{\infty} \kappa_2(j+1) \sum_{i=0}^j \sum_{\alpha, \beta=1}^{\tau} (\Psi_t^{(r)}\Phi_t^{(r)})^i[r,\alpha]\Gamma_t^{(r)}(\alpha,-\beta)(\Phi_t^{(r)})^T(\Psi_t^{(r)})^i[-\beta,c]
\]

Applying the same argument for the convergence of the corresponding \( \hat{\Sigma}_t^{(r)} \) in Theorem 3.3,

\[
\lim_{\tau \to \infty} \hat{\Sigma}_t^{(r)}[r,c] = \sum_{p,q=0}^{\infty} (\Psi_t\Phi_t)^p[r,\alpha] \sum_{\beta=1}^{\infty} \kappa_2(p+q+\alpha+\beta+1)\Gamma_t[\alpha,0] S^{-2\beta} (\Phi_t^T\Psi_t^T)^q[0,c]
\]

where the second equality follows from the definition of \( \hat{\kappa} \) in (3.11). Similarly, we also have

\[
\lim_{\tau \to \infty} \hat{\Sigma}_t^{(r)}[r,c] = \sum_{p,q=0}^{\infty} (\Psi_t\Phi_t)^p(\hat{\kappa}_2(p+q+1) - \kappa_2(p+q+1)\text{Id}) \circ (\hat{\Gamma}_t + \tilde{\Gamma}_t)(\Phi_t^T\Psi_t^T)^q[0,c]
\]

**Convergence of \( \Sigma_t^{(r)} \).** It directly follows from the convergence of \( \Phi_t^{(r)} \), \( \Psi_t^{(r)} \), and \( \Gamma_t^{(r)} \) in \( t^{(G,c)} \) and \( t^{(F,c)} \) that \( \lim_{\tau \to \infty} \Sigma_t^{(r)} = \Sigma_t \).
We describe the modifications of the arguments in [Fan20] that are needed to show Theorem 2.3.

C Proof for independent initialization

Suppose Assumptions 2.1, 2.2, and C.1 hold. Then almost surely for each fixed $t \geq 1$, we may establish similarly $t+1^{(G.a-G.f)}$. The proof is symmetric to the above, using the identities (B.9) and (B.11) in place of (B.8). Hence, we conclude that $\lim_{t \to \infty} \Sigma_t^{(\tau)} = \Sigma_t$, which shows $t^{(F.f)}$.

Now supposing by induction that $s^{(F)}$ and $s^{(G)}$ hold for $0 \leq s \leq t$, we may establish similarly $t+1^{(G.a-G.f)}$. The proof is symmetric to the above, using the identities (B.9) and (B.11) in place of (B.8) and (B.10), and we omit this for brevity. This completes the induction.

C Proof for independent initialization

We describe the modifications of the arguments in [Fan20] that are needed to show Theorem 2.3. Write the AMP iterations (2.1–2.2) as

\[ R_t = O \Sigma_t, \quad S_t = O^\top A R_t, \quad Z_t = S_t - U_t b_{t-1}^\top - \cdots - U_t b_{tt}^\top, \quad U_{t+1} = u_{t+1}(Z_1, \ldots, Z_t, E). \]

For each $t \geq 1$ and $k \geq 0$, define the $tK \times tK$ matrix

\[ L_t^{(k)} = \sum_{j=0}^\infty c_{k,j} \Theta_t^{(j)}, \quad \Theta_t^{(j)} = \Theta^{(j)}[\Phi_t, \kappa_j + 2 \Delta_t] \]

where \{c_{k,j}\}_{k,j \geq 0} are the partial moment coefficients defined in [Fan20, Section A.1], and $\Theta^{(j)}[\cdot, \cdot]$ is as defined in (2.6). We write

\[ B_t = \sum_{j=0}^\infty c_{k,j} \Phi_t^j \]

as the large-$n$ limit of $b_t$, replacing $\phi_t$ by its large-$n$ limit $\Phi_t$. (To simplify notation, we are using $\phi_t, b_t^\top$ and $\Phi_t, B_t^\top$ in place of $\Phi_t, B_t$ and $\Phi^\infty_t, B^\infty_t$ respectively in [Fan20].) We have also defined $\Sigma_t, \Theta_t^{(j)}, L_t^{(k)}$ directly in the large-$n$ limit, instead of using the notation $\Sigma_t^\infty, \Theta_t^{(j)}, L_t^{(k,\infty)}$. The identities and proofs of [Fan20, Lemmas A.2 and A.3] on the relations between $B_t^\infty, \Sigma_t, \Theta_t^{(j)}, L_t^{(k)}$ then hold without any modification, as they rely only on the definitions of these matrices in terms of $\Delta_t, \Phi_t$ and on the combinatorial relations between $\{c_{k,j}\}_{k,j \geq 0}$.

We first show an extended form of Theorem 2.3 under the following additional assumption.

**Assumption C.1.** The random variables $\Lambda$ and $U_1$ satisfy $\text{Var}[\Lambda] > 0$ and $E[U_1 U_1^\top] > 0$ strictly. Defining $(U_1, \ldots, U_{t+1}, Z_1, \ldots, Z_t)$ by the state evolution (2.7), for each $t \geq 1$ and any deterministic vector $v \in \mathbb{R}^K$, there do not exist constants $\alpha_1, \ldots, \alpha_t, \beta_1, \ldots, \beta_t$ for which $v^\top U_{t+1}$ is almost surely equal to $\alpha_1 \cdot v^\top U_1 + \ldots + \alpha_t \cdot v^\top U_t + \beta_1 \cdot v^\top Z_1 + \ldots + \beta_t \cdot v^\top Z_t$.

**Lemma C.2.** Suppose Assumptions 2.1, 2.2, and C.1 hold. Then almost surely for each fixed $t \geq 1$,

(a) $\lim_{n \to \infty} \langle (U_t^\top U_s) \rangle_{t,s=1}^l = \Delta_t$ and $\lim_{n \to \infty} \phi_t = \Phi_t$.

(b) For some random vectors $R_1, \ldots, R_t \in \mathbb{R}^K$ having finite second moment, $(R_1, \ldots, R_t, \Lambda) \xrightarrow{W_2} (R_1, \ldots, R_t, \Lambda)$. Furthermore, for each $k \geq 0$, $E[(R_1, \ldots, R_t)^\top \Lambda^k (R_1, \ldots, R_t)] = L_t^{(k)}$.

(c) $(U_1, \ldots, U_{t+1}, Z_1, \ldots, Z_t, E) \xrightarrow{W_2} (U_1, \ldots, U_{t+1}, Z_1, \ldots, Z_t, E)$ as described in Theorem 2.3.
(d) The matrix
\[
\begin{pmatrix}
\Delta_t & \Phi_t \Sigma_t \\
\Sigma_t \Phi_t^T & \Sigma_t
\end{pmatrix}
\]
is non-singular.

**Proof.** The proof is an extension of that of [Fan20, Lemma A.4]. We denote \(\ell(a),(b),(c),(d)\) for the claims of (a–d) up to iteration \(t\).

Applying Lemma D.5(a) to \(R_1 = OU_1\), we have \((\mathbf{X}, R_1) \xrightarrow{W_2} (\Lambda, R_1)\) where \(R_1 \sim \mathcal{N}(0, \mathbb{E}[U_1 U_1^T])\) is independent of \(\Lambda\). Then applying Lemma D.3, \(n^{-1} R_1^T \Lambda^k R_1 \to m_k \cdot \mathbb{E}[U_1 U_1^T] = m_k \cdot \Delta_1\) where \(m_k = \mathbb{E}[\Lambda^k]\). Then \(1^{(a)}\) and \(1^{(b)}\) follow as in [Fan20, Lemma A.4]. Conditional on \(R_1 = OU_1\), we may represent the law of \(S_1\) as
\[
S_1 = S_{\|} + S_{\perp}, \quad S_{\|} = U_1(R_1^T R_1)^{-1} R_1^T \Lambda R_1, \quad S_{\perp} = \Pi U_1 \tilde{O} \Pi R_1^T \Lambda R_1
\]
where \(\Pi U_1^+, \Pi R_1^+ \in \mathbb{R}^{n \times (n-K)}\) are projections orthogonal to the column spans of \(U_1, R_1\), and \(\tilde{O}\) is a Haar-orthogonal matrix of dimension \(n-K\). We have \(n^{-1} R_1^T R_1 \to \mathbb{E}[U_1 U_1^T] = \Delta_1\), which is invertible by Assumption C.1, and
\[
\begin{align*}
n^{-1}(\Pi R_1^+ \Lambda R_1)^T (\Pi R_1^+ \Lambda R_1) &= n^{-1} R_1^+ \Lambda^2 R_1 - n^{-1} R_1^+ \Lambda R_1 \cdot (n^{-1} R_1^+ R_1)^{-1} \cdot n^{-1} R_1^+ \Lambda R_1
\end{align*}
\]
Then applying the same arguments as [Fan20, Lemma A.4], using Lemma D.5(a) to analyze \(S_{\perp}\), we obtain
\[
(W_2, (Z_1, U_1, E)) \xrightarrow{W_2} (Z_1, U_1, E), \quad (U_1, E) \perp \perp Z_1 \sim \mathcal{N}(0, \kappa_2 \Delta_1).
\]
Here, \(\kappa_2 \Delta_1 = \Sigma_1\). Since \(u_2(Z_1, E)\) is Lipschitz, Lemma D.4 then implies \(1^{(c)}\). Assumption C.1 and the identities \(\Sigma_1 = \kappa_2 \Delta_1\) and \(\Phi_1 = 0\) then imply \(1^{(d)}\).

Now suppose \(\ell(a),(b),(c),(d)\) hold. For \(t+1^{(a)}\), convergence to \(\Delta_{t+1} \) in \(t+1^{(a)}\) is immediate from \(W_2\)-convergence of \((U_1, \ldots, U_{t+1})\) in \(t^{(c)}\) and Lemma D.3 (applied with \(k = 0\)). Since each \(u_\lambda(\cdot)\) is Lipschitz, its derivatives are bounded and also continuous with probability 1 with respect to the limit law of \((Z_1, \ldots, Z_t, E)\) in \(t^{(c)}\), by Assumption 2.2. Then the convergence \(\phi_t \to \Phi_t\) follows also from weak convergence of \((Z_1, \ldots, Z_t, E)\) in \(t^{(c)}\), which shows \(t+1^{(a)}\).

For \(t+1^{(b)}\), observe that by the \(W_2\)-convergence in \(t^{(c)}\), weak-differentiability of the Lipschitz function \(u_\lambda(\cdot)\), and [Fan20, Proposition E.5],
\[
n^{-1}(Z_1, \ldots, Z_t)^T (U_1, \ldots, U_t) \to \Sigma_t \cdot \Phi_t^T.
\]
Then conditioning \(O\) on \((R_1, \ldots, R_t) = O(U_1, \ldots, U_t)\) and \(O(Z_1, \ldots, Z_t) = \Lambda(R_1, \ldots, R_t) - (R_1, \ldots, R_t)B_t^T\), the same argument that shows [Fan20, Eq. (A.23)] yields \((R_1, \ldots, R_{t+1}, \Lambda) \xrightarrow{W_2} (R_1, \ldots, R_{t+1}, \Lambda)\) where
\[
R_{t+1} = \begin{pmatrix}
R_1 & \cdots & R_t & \Lambda R_1 & \cdots & \Lambda R_t
\end{pmatrix}
\mathbf{γ}_t^{-1} \begin{pmatrix}
\Delta_{t+1}[1 : t, t + 1] \\
\Phi_{t+1}[1 : t, t + 1]
\end{pmatrix} + R_{\perp}.
\]
Here
\[
\mathbf{γ}_t = \begin{pmatrix}
\Delta_t & \Delta_t B_t^T + \Phi_t \Sigma_t \\
\Phi_t B_t^T + \text{Id}
\end{pmatrix} \in \mathbb{R}^{2K \times 2K},
\]

56
and in the decomposition into blocks of size $K \times K$, $\Delta_{t+1}[1 : t, t + 1]$ and $\Phi_{t+1}[1 : t, t + 1]$ denote the first $t$ row blocks of the last column block $t+1$ of $\Delta_{t+1}$ and $\Phi_{t+1}[1 : t, t + 1]$. The random vector $R_\perp \in \mathbb{R}^K$ has the Gaussian law

$$R_\perp \sim \mathcal{N} \left( 0, \mathbb{E} [U_{t+1} U_{t+1}^\top] - \left( \begin{array}{c} \Delta_{t+1}[1 : t, t + 1] \\ \Sigma_t \Phi_{t+1}[1 : t, t + 1] \end{array} \right) \left( \begin{array}{c} \Delta_t \\ \Sigma_t \Phi_t \end{array} \right)^{-1} \left( \begin{array}{c} \Delta_{t+1}[1 : t, t + 1] \\ \Sigma_t \Phi_{t+1}[1 : t, t + 1] \end{array} \right) \right).$$

For any fixed $v \in \mathbb{R}^K$, we may check that $v^\top \text{Cov}[R_\perp]v$ is the residual variance of the $L_2$-projection of $v^\top U_{t+1}$ onto the linear span of $(v^\top Z_1, \ldots, v^\top Z_t, v^\top U_1, \ldots, v^\top U_t)$. Thus, by Assumption C.1, we have analogously to [Fan20, Eq. (A.24)] that

$$\text{Cov}[R_\perp] \succ 0 \quad \text{(C.1)}$$

Note that Lemma D.3 implies

$$n^{-1}(R_1, \ldots, R_{t+1})^\top \Lambda^K(R_1, \ldots, R_{t+1}) \rightarrow \mathbb{E}[(R_1, \ldots, R_{t+1})^\top \Lambda^K(R_1, \ldots, R_{t+1})].$$

Using [Fan20, Lemmas A.1 and A.2] and the same arguments as in [Fan20, Lemma A.4] to analyze $\mathbb{E}[(R_1, \ldots, R_t)^\top \Lambda^K R_{t+1}]$ and $\mathbb{E}[R_{t+1}^\top \Lambda^K R_{t+1}]$, we obtain $\mathbb{E}[(R_1, \ldots, R_{t+1})^\top \Lambda^K(R_1, \ldots, R_{t+1})] = L_{t+1}$ and hence conclude $t+1$.

For $t+1$, we define $\Phi_t = \Phi_{t+1}[1 : t + 1, 1 : t]$ and $\tilde{B}_t = B_{t+1}[1 : t + 1, 1 : t]$ as the first $t$ column blocks of $\Phi_{t+1}$ and $B_{t+1}$, and condition $O$ on

$$(R_1, \ldots, R_{t+1}) = O(U_1, \ldots, U_{t+1}) \text{ and } O(Z_1, \ldots, Z_t) = \Lambda(R_1, \ldots, R_{t+1}) - (R_1, \ldots, R_{t+1}) \tilde{B}_t^\top.$$

Applying again the $W_2$-convergence in $t$ and [Fan20, Proposition E.5], we have

$$n^{-1}(Z_1, \ldots, Z_t)^\top(U_1, \ldots, U_{t+1}) \rightarrow \Sigma_t \cdot \Phi_t^\top.$$

Observe that

$$\begin{pmatrix} \Delta_{t+1} \\ \Sigma_t \Phi_t \end{pmatrix} = \begin{pmatrix} \Delta_t \\ \Sigma_t \Phi_t \end{pmatrix} \left( \begin{array}{c} \Delta_t \\ \Sigma_t \Phi_t \end{array} \right)^{-1} \left( \begin{array}{c} \Delta_t \\ \Sigma_t \Phi_t \end{array} \right)$$

is invertible, as its submatrix removing row block $t + 1$ and column block $t + 1$ is invertible by $t^{(d)}$, and the Schur-complement of the $(t + 1, t + 1)$ block is invertible by (C.1). Then applying the same arguments as leading to [Fan20, Eq. (A.33)], we get $$(U_1, \ldots, U_{t+1}, Z_1, \ldots, Z_t, E, S_{t+1}) \xrightarrow{W_2} (U_1, \ldots, U_{t+1}, Z_1, \ldots, Z_t, E, S_{t+1})$$

where $S_{t+1} = (U_1 \cdots U_{t+1}) B_{t+1}[1 : t, t + 1] + (Z_1 \cdots Z_t) \Sigma_t^{-1} \cdot \Sigma_{t+1}[1 : t, t + 1] + S_\perp$ and

$$S_\perp \sim \mathcal{N} \left( 0, \left( \begin{array}{c} L_{t+1}^{(2)}[t + 1, t + 1] - L_{t+1}^{(1)}[1 : t + 1, t + 1] \\ L_{t+1}^{(0)} \end{array} \right) \left( \begin{array}{c} L_{t+1}^{(2)}[1 : t + 1, t + 1] \\ L_{t+1}^{(0)} \end{array} \right)^{-1} \left( \begin{array}{c} L_{t+1}^{(1)}[1 : t + 1, t + 1] \\ L_{t+1}^{(2)}[1 : t + 1, t + 1] \end{array} \right) \right).$$

Here again, we use $[1 : t, t + 1]$ to denote the first $t$ row blocks of the last column block $t+1$, and similarly for $[t+1, t + 1]$ and $[1 : t+1, t + 1]$. Since $Z_{t+1} = S_{t+1} - (U_1 \cdots U_{t+1}) B_{t+1}[1 : t, t + 1]$, the
some computations as in [Fan20, Lemma A.4] show convergence of \((U_1, \ldots, U_{t+1}, Z_1, \ldots, Z_{t+1}, E)\) as in \(t+1^{(c)}\), where the covariance of \((Z_1, \ldots, Z_{t+1})\) is exactly \(\Sigma_{t+1}\). Since \(U_{t+2} = u_{t+2}(Z_1, \ldots, Z_{t+1}, E)\) and \(u_{t+2}(\cdot)\) is Lipschitz, \(t+1^{(c)}\) follows from Lemma D.4.

Finally, for \(t + 1^{(d)}\), observe from the definition of \(L^{(k)}_t\) that for any fixed vector \(v \in \mathbb{R}^K\), \(v^\top \text{Cov}[S_{\perp}]v\) is the residual variance of the \(L_2\)-projection of \(v^\top AR_{t+1}\) onto the linear span of \((v^\top R_1, \ldots, v^\top R_{t+1}, v^\top AR_1, \ldots, v^\top AR_t)\). If \(v^\top \text{Cov}[S_{\perp}]v = 0\), then this would imply

\[
v^\top AR_{t+1} = \alpha_1 \cdot v^\top R_1 + \ldots + \alpha_{t+1} \cdot v^\top R_{t+1} + \beta_1 \cdot v^\top AR_1 + \ldots + \beta_t \cdot v^\top AR_t,
\]

and hence \((\Lambda - \alpha_{t+1})v^\top R_{\perp} = f(v^\top R_1, \ldots, v^\top R_t, \Lambda)\) for some function \(f\). This implies that \(R_{\perp}\) is constant conditional on \((R_1, \ldots, R_t, \Lambda)\) and the positive-probability event \(\Lambda \neq \alpha_{t+1}\). This contradicts that \(R_{\perp}\) is independent of \((R_1, \ldots, R_t, \Lambda)\) with a Gaussian law of positive variance. So \(v^\top \text{Cov}[S_{\perp}]v > 0\) implying \(\text{Cov}[S_{\perp}] > 0\) strictly. Then \(t + 1^{(d)}\) follows as in [Fan20, Lemma A.4], concluding the proof.

**Proof of Theorem 2.3.** We may remove Assumption C.1 by studying a perturbed AMP sequence and applying a continuity argument, as in [Fan20, Appendix D] and [BMN20]. Consider the perturbed AMP iterations

\[
Z_t^\epsilon = W^\epsilon U_t^\epsilon - U_t^\epsilon b_t^\epsilon^\top - \ldots - U_t^\epsilon b_t^\epsilon^\top, \quad U_{t+1}^\epsilon = u_{t+1}(Z_t^\epsilon, \ldots, Z_t^\epsilon, E) + \epsilon G_{t+1}
\]

where \(W^\epsilon = O^\top \text{diag}(\lambda + \epsilon \gamma)O\), the vector \(\gamma \in \mathbb{R}^n\) has i.i.d. Uniform\((-1, 1)\) entries, each \(b_t^\epsilon\) is the version of \(b_t\)s defined with the limit free cumulants of \(W^\epsilon\) in place of \(W\), and \(G_{t+1} \in \mathbb{R}^{n \times K}\) is an independent matrix with i.i.d. \(N(0, 1)\) entries in each iteration. Then the same argument as in [Fan20, Appendix D] shows

\[
\lim_{\epsilon \to 0} \limsup_{n \to \infty} n^{-1}||Z_t^\epsilon - Z_t||_F^2 = 0, \quad \lim_{\epsilon \to 0} \limsup_{n \to \infty} n^{-1}||U_t^\epsilon - U_t||_F^2 = 0,
\]

\[
\lim_{\epsilon \to 0} (\Delta_t^\epsilon, \Phi_t^\epsilon, B_t^\epsilon^\top, \Sigma_t^\epsilon) = (\Delta_t, \Phi_t, B_t^\top, \Sigma_t)
\]

where \(\Delta_t^\epsilon, \Phi_t^\epsilon, B_t^\epsilon, \Sigma_t^\epsilon\) are the matrices corresponding to this perturbed AMP sequence. Note that in [Fan20, Appendix D], the continuous-differentiability of each function \(u_t(\cdot)\) is used only to show the convergence \(\Phi_t \to \Phi_{\infty}\) at the end of the proof (i.e. \(\Phi_t \to \Phi_t\) in our notation), and the relaxed condition of Assumption 2.2(b) is sufficient for this statement. On the other hand, Assumption C.1 holds for this perturbed AMP sequence, so this perturbed sequence is characterized by Lemma C.2. Combining these shows Theorem 2.3.

**D Auxiliary lemmas**

**Lemma D.1.** For each \(\tau > 0\), let \(\{x_i^{(\tau)}\}_{0 \leq i \leq \tau}\) and \(\{y_i^{(\tau)}\}_{0 \leq i \leq \tau}\) be two sequences where \(|x_i^{(\tau)}|\) and \(|y_i^{(\tau)}|\) uniformly bounded by a constant \(C > 0\). Suppose, for any \(\epsilon > 0\), there exists some \(T > 0\) such that for all \(\tau > T\),

\[
\sum_{i=T}^{\tau} |x_i^{(\tau)}| \leq \epsilon \quad \text{and} \quad \lim_{\tau \to \infty} |y_i^{(\tau)}| = 0 \quad \text{for all} \quad 0 \leq i \leq T.
\]

Then \(\lim_{\tau \to \infty} \sum_{i=0}^{\tau} x_i^{(\tau)} y_i^{(\tau)} = 0\).
Proof. For any $\epsilon > 0$, let $T$ be as given. Then for all $\tau > T$, we have

$$
\left| \sum_{i=0}^{\tau} x_i(\tau) y_i(\tau) \right| \leq \max_{0 \leq i \leq T} |x_i(\tau)| \cdot \sum_{i=0}^{T} |y_i(\tau)| + \max_{T < i \leq \tau} |y_i(\tau)| \cdot \sum_{i=T}^{\tau} |x_i(\tau)|
$$

where $C$ is the given uniform bound. Taking the limit $\tau \to \infty$ on both sides, we obtain

$$
\limsup_{\tau \to \infty} \left| \sum_{i=0}^{\tau} x_i(\tau) y_i(\tau) \right| \leq C \epsilon.
$$

Then since $\epsilon$ is arbitrary, the desired result follows. \qed

Lemma D.2. For any integers $\tau, t, K \geq 1$, let $A \in \mathbb{R}^{(\tau+t+1)K \times (\tau+t+1)K}$ have the block decomposition

$$
A = \begin{pmatrix} A_- & 0 \\ A_+ & A_+ \end{pmatrix}
$$

where $A_- \in \mathbb{R}^{\tau K \times \tau K}$ and $A_+ \in \mathbb{R}^{(t+1)K \times (t+1)K}$. Suppose $A_- \text{ and } A_+$ have the further block decompositions

$$
A_- = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & B & 0 \end{pmatrix}, \quad A_+ = \begin{pmatrix} 0 & \cdots & 0 & B \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix}
$$

for some $B \in \mathbb{R}^{K \times K}$. Then, indexing the row and column blocks by $\{-\tau, -\tau + 1, \ldots, t\}$, for all $r = 0, \ldots, t$ and $c = 1, \ldots, \tau$,

$$
A_j[r, -c] = \begin{cases} A_+^c[r, 0] \cdot B^c & \text{if } 1 \leq c \leq j, \\ 0 & \text{if } j < c \end{cases}
$$

where $[r, -c]$ denotes the $K \times K$ submatrix corresponding to row block $r$ and column block $-c$.

Proof. Due to the special structure of $A$, it is straightforward to verify (by induction) that for any $j \geq 1$, we have

$$
A^j = \begin{pmatrix} A_- & A_+ \end{pmatrix} \left( \sum_{i=0}^{j-1} A_i A_+ A_-^j A_+^j \right).
$$

Moreover, observe that for each $j-i-1 \geq 0$,

$$
A_+ A_-^j = \begin{pmatrix} 0 & \cdots & 0 & B^{j-i} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}.
$$
Following the convention that $A^0_{++} = \text{Id}$, we further have
\[
\sum_{i=0}^{j-1} A^i_{++} A_{--} A^{j-i-1} = \begin{pmatrix} 0 & \ldots & 0 & A^0_{++}[;0] B^j & \ldots & A^{j-1}_{++}[;0] B \end{pmatrix},
\]
or equivalently, for all $r = 0, \ldots, t$ and $c = 1, \ldots, \tau$,
\[
A^j[r,-c] = \begin{cases} A^j_{++}[r,0] \cdot B^c & 1 \leq c \leq j, \\ 0 & j < c. \end{cases}
\]

Lemma D.3. If $\Lambda = \text{diag}(\lambda) \in \mathbb{R}^{n \times n}$ and the random variable $\Lambda$ satisfy Assumption 2.1, and $R \in \mathbb{R}^{n \times J}$ is such that $(R, \Lambda) \overset{W_2}{\to} (R, \Lambda)$ as $n \to \infty$ for a random vector $R \in \mathbb{R}^J$, then for any $k \geq 0$,
\[
\lim_{n \to \infty} \frac{1}{n} R^\top \Lambda^k R = \mathbb{E}[\Lambda^k \cdot RR^\top] \in \mathbb{R}^{J \times J}.
\]

Proof. Fix any $a, b \in \{1, \ldots, J\}$, let $C > \max(|\lambda_\cdot|, |\lambda_+|)$, and define the function $f_C(\lambda, r_a, r_b) = \min(\max(\lambda, -C), C)^k \cdot r_a r_b$. Then $|f_C(\lambda, r_a, r_b)| \leq C^k (r_a^2 + r_b^2)$, so by the given Wasserstein-2 convergence, a.s.
\[
\frac{1}{n} \sum_{i=1}^n f_C(\lambda_i, r_{a,i}, r_{b,i}) \to \mathbb{E}[f_C(\Lambda, R_a, R_b)].
\]
The left side coincides with the $(a, b)$ entry of $n^{-1} R^\top \Lambda^k R$ a.s. for all large $n$, while the right side coincides with that of $\mathbb{E}[\Lambda^k \cdot RR^\top]$. □

Lemma D.4. Suppose $Z \in \mathbb{R}^{n \times J}$ satisfies $Z \overset{W_2}{\to} Z$, and $u : \mathbb{R}^J \to \mathbb{R}^k$ is Lipschitz. Then $(Z, u(Z)) \overset{W_2}{\to} (Z, u(Z))$.

Proof. Let $L$ be the Lipschitz constant of $u$. For any continuous function $f : \mathbb{R}^{j+k} \to \mathbb{R}$ satisfying $|f(z, u)| \leq C(1 + \|z\|^2 + \|u\|^2)$, we have
\[
|f(z, u(z))| \leq C(1 + \|z\|^2 + (L\|z\| + u(0))^2) \leq C'(1 + \|z\|^2)
\]
for a different constant $C' > 0$. Then $n^{-1} \sum_i f(z_i, u(z_i)) \to \mathbb{E}[f(Z, u(Z))]$, so the result follows. □

Lemma D.5. Fix $J, L \geq 0$ and $K \geq 1$. Let $O \in \mathbb{R}^{(n-L) \times (n-L)}$ be a random Haar-uniform orthogonal matrix. Let $E \in \mathbb{R}^{n \times J}$ and $V \in \mathbb{R}^{(n-L) \times K}$ satisfy $E \overset{W_2}{\to} E$ and $n^{-1} V^\top V \to \Sigma \in \mathbb{R}^{K \times K}$. Let $\Pi \in \mathbb{R}^{n \times (n-L)}$ be any deterministic matrix with orthonormal columns. Then almost surely as $n \to \infty$,
\[
(\Pi O V, E) \overset{W_2}{\to} (Z, E)
\]
where $Z \sim \mathcal{N}(0, \Sigma) \in \mathbb{R}^K$ is independent of $E \in \mathbb{R}^J$.

Proof. Let $\hat{O}$ be the first $K$ columns of $O$. Writing the singular value decomposition $V = Q DU^\top$ and applying the equality in law $OQ \overset{L}{\overset{D}{\sim}} \hat{O}(V^\top V)^{1/2}$, where $(V^\top V)^{1/2} = UDU^\top$ is the positive-semidefinite matrix square-root. Then introducing $Z \in \mathbb{R}^{n \times K}$ with i.i.d. $\mathcal{N}(0, 1)$ entries, and applying $\hat{O} \overset{L}{\sim} \Pi^\top Z(n^{-1} Z^\top \Pi \Pi^\top Z)^{-1/2}$, also
\[
O V \overset{L}{\sim} \Pi^\top Z(n^{-1} Z^\top \Pi \Pi^\top Z)^{-1/2}(n^{-1} V^\top V)^{1/2}.
\]
So
\[ \Pi O V = (\text{Id} - \Pi^\perp) Z (n^{-1} Z^\top \Pi \Pi^\top Z)^{-1/2} (n^{-1} V^\top V)^{1/2} \]
\[ (D.1) \]
where \( \Pi^\perp = \text{Id} - \Pi \Pi^\top \in \mathbb{R}^{n \times n} \) is a projection onto a subspace of fixed dimension \( L \). Since \( n^{-1} Z^\top \Pi \Pi^\top Z \to \text{Id}_{K \times K} \) and \( n^{-1} V^\top V \to \Sigma \), and the matrix square-root is continuous, we obtain
\[ \left( Z (n^{-1} Z^\top \Pi \Pi^\top Z)^{-1/2} (n^{-1} V^\top V)^{1/2} , E \right) \overset{W_2}{\to} (Z, E) \]
\[ (D.2) \]
by [Fan20, Propositions E.1 and E.4]. Now write \( \Pi^\perp = UU^\top \) where \( U \in \mathbb{R}^{n \times L} \) has orthonormal columns. Then \( \Pi^\perp Z \overset{L}{=} UG \) where \( G \in \mathbb{R}^{L \times K} \) has i.i.d. \( \mathcal{N}(0,1) \) entries. Let \( u_i \) be the \( i \)th row of \( U \), and \( g_j \) be the \( j \)th column of \( G \). Then a.s.
\[ \frac{1}{n} \sum_{i=1}^{n} (u_i^\top g_j)^2 \leq \frac{1}{n} \sum_{i=1}^{n} \|u_i\|^2 \cdot \|g_j\|^2 = \frac{L}{n} \cdot \|g_j\|^2 \to 0. \]
This holds for each column \( j \) of \( \Pi^\perp Z \overset{L}{=} UG \), so \( \Pi^\perp Z \overset{W_2}{\to} 0 \). Then combining with (D.2) and applying this to (D.1), we obtain \((\Pi O V, E) \overset{W_2}{\to} (Z, E)\) by [Fan20, Proposition E.4].

Acknowledgments

We would like to thank Marco Mondelli for a helpful discussion about [MV21b]. This research was supported in part by NSF DMS-1916198.

References

[BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. The Annals of Probability, 33(5):1643–1697, 2005.

[BG09] Florent Benaych-Georges. Rectangular random matrices, related convolution. Probability Theory and Related Fields, 144(3-4):471–515, 2009.

[BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Advances in Mathematics, 227(1):494–521, 2011.

[BGN12] Floren Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. Journal of Multivariate Analysis, 111:120–135, 2012.

[BKM+19] Jean Barbier, Florent Krzakala, Nicolas Macris, Léon Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. Proceedings of the National Academy of Sciences, 116(12):5451–5460, 2019.

[BKRS20] Zhiqi Bu, Jason M Klusowski, Cynthia Rush, and Weijie J Su. Algorithmic analysis and statistical estimation of SLOPE via approximate message passing. IEEE Transactions on Information Theory, 67(1):506–537, 2020.
[BLM15] Mohsen Bayati, Marc Lelarge, and Andrea Montanari. Universality in polytope phase transitions and message passing algorithms. *The Annals of Applied Probability*, 25(2):753–822, 2015.

[BM11a] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[BM11b] Mohsen Bayati and Andrea Montanari. The LASSO risk for Gaussian matrices. *IEEE Transactions on Information Theory*, 58(4):1997–2017, 2011.

[BMDK20] Jean Barbier, Nicolas Macris, Mohamad Dia, and Florent Krzakala. Mutual information and optimality of approximate message-passing in random linear estimation. *IEEE Transactions on Information Theory*, 66(7):4270–4303, 2020.

[BMN20] Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *Information and Inference: A Journal of the IMA*, 9(1):33–79, 2020.

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

[Bol18] Erwin Bolthausen. A Morita type proof of the replica-symmetric formula for SK. In *International Conference on Statistical Mechanics of Classical and Disordered Systems*, pages 63–93. Springer, 2018.

[BS06] Jinho Baik and Jack W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408, 2006.

[CFM21] Michael Celentano, Zhou Fan, and Song Mei. Local convexity of the TAP free energy and AMP convergence for Z2-synchronization. *arXiv preprint arXiv:2106.11428*, 2021.

[CL21] Wei-Kuo Chen and Wai-Kit Lam. Universality of approximate message passing algorithms. *Electronic Journal of Probability*, 26:1–44, 2021.

[ÇO20] Burak Çakmak and Manfred Opper. A dynamical mean-field theory for learning in restricted Boltzmann machines. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(10):103303, 2020.

[DAM17] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2):125–170, 2017.

[DJM13] David L Donoho, Adel Javanmard, and Andrea Montanari. Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE Transactions on Information Theory*, 59(11):7434–7464, 2013.

[DM14] Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse PCA. In *2014 IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014.
David Donoho and Andrea Montanari. High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probability Theory and Related Fields*, 166(3):935–969, 2016.

Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems*, pages 424–432, 2016.

David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

David L Donoho, Arian Maleki, and Andrea Montanari. Message passing algorithms for compressed sensing: I. Motivation and construction. In 2010 *IEEE information theory workshop on information theory (ITW 2010, Cairo)*, pages 1–5. IEEE, 2010.

David L Donoho, Arian Maleki, and Andrea Montanari. Message passing algorithms for compressed sensing: II. Analysis and validation. In 2010 *IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, pages 1–5. IEEE, 2010.

Jian Ding and Nike Sun. Capacity lower bound for the Ising perceptron. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 816–827, 2019.

Zhou Fan. Approximate message passing algorithms for rotationally invariant matrices. *arXiv preprint arXiv:2008.11892*, 2020.

Oliver Y Feng, Ramji Venkataramanan, Cynthia Rush, and Richard J Samworth. A unifying tutorial on Approximate Message Passing. *arXiv preprint arXiv:2105.02180*, 2021.

Zhou Fan and Yihong Wu. The replica-symmetric free energy for Ising spin glasses with orthogonally invariant couplings. *arXiv preprint arXiv:2105.02797*, 2021.

Adel Javanmard and Andrea 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.

Yoshiyuki Kabashima. A CDMA multiuser detection algorithm on the basis of belief propagation. *Journal of Physics A: Mathematical and General*, 36(43):11111, 2003.

Florent Krzakala, Marc Mézard, François Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(08):P08009, 2012.

Lei Liu, Shunqi Huang, and Brian M Kurkoski. Memory approximate message passing. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1379–1384. IEEE, 2021.

Arian Maleki. *Approximate message passing algorithms for compressed sensing*. PhD thesis, Stanford University, 2010.
[MAYB13] Arian Maleki, Laura Anitori, Zai Yang, and Richard G Baraniuk. Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP). *IEEE Transactions on Information Theory*, 59(7):4290–4308, 2013.

[MP17] Junjie Ma and Li Ping. Orthogonal AMP. *IEEE Access*, 5:2020–2033, 2017.

[MS17] James A Mingo and Roland Speicher. *Free probability and random matrices*, volume 35. Springer, 2017.

[MV21a] Marco Mondelli and Ramji Venkataramanan. Approximate message passing with spectral initialization for generalized linear models. In *International Conference on Artificial Intelligence and Statistics*, pages 397–405. PMLR, 2021.

[MV21b] Marco Mondelli and Ramji Venkataramanan. PCA initialization for Approximate Message Passing in rotationally invariant models. *arXiv preprint arXiv:2106.02356*, 2021.

[MV21c] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics*, 49(1):321–345, 2021.

[OÇW16] Manfred Opper, Burak Çakmak, and Ole Winther. A theory of solving TAP equations for Ising models with general invariant random matrices. *Journal of Physics A: Mathematical and Theoretical*, 49(11):114002, 2016.

[Pau07] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[RF12] Sundeep Rangan and Alyson K Fletcher. Iterative estimation of constrained rank-one matrices in noise. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 1246–1250. IEEE, 2012.

[RSF19] Sundeep Rangan, Philip Schniter, and Alyson K Fletcher. Vector approximate message passing. *IEEE Transactions on Information Theory*, 65(10):6664–6684, 2019.

[RSFS19] Sundeep Rangan, Philip Schniter, Alyson K Fletcher, and Subrata Sarkar. On the convergence of approximate message passing with arbitrary matrices. *IEEE Transactions on Information Theory*, 65(9):5339–5351, 2019.

[SBC17] Weijie Su, Malgorzata Bogdan, and Emmanuel Candes. False discoveries occur early on the lasso path. *The Annals of Statistics*, 45(5):2133–2150, 2017.

[SCC19] Pragya Sur, Yuxin Chen, and Emmanuel J Candès. The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probability theory and related fields*, 175(1):487–558, 2019.

[SRF16] Philip Schniter, Sundeep Rangan, and Alyson K Fletcher. Vector approximate message passing for the generalized linear model. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1525–1529. IEEE, 2016.

[Tak17] Keigo Takeuchi. Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 501–505. IEEE, 2017.
[Tak19] Keigo Takeuchi. A unified framework of state evolution for message-passing algorithms. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 151–155. IEEE, 2019.

[Tak20a] Keigo Takeuchi. Bayes-optimal convolutional AMP. *arXiv preprint arXiv:2003.12245*, 2020.

[Tak20b] Keigo Takeuchi. Convolutional approximate message-passing. *IEEE Signal Processing Letters*, 27:416–420, 2020.

[VSR+15] Jeremy Vila, Philip Schniter, Sundeep Rangan, Florent Krzakala, and Lenka Zdeborová. Adaptive damping and mean removal for the generalized approximate message passing algorithm. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2021–2025. IEEE, 2015.

[Zie12] William P Ziemer. *Weakly differentiable functions: Sobolev spaces and functions of bounded variation*, volume 120. Springer Science & Business Media, 2012.

[ZSF20] Xinyi Zhong, Chang Su, and Zhou Fan. Empirical bayes pca in high dimensions. *arXiv preprint arXiv:2012.11676*, 2020.