

Rigorous Dynamics of Expectation-Propagation-Based Signal Recovery from Unitarily Invariant Measurements

Keigo Takeuchi, *Member, IEEE*

Abstract—Signal recovery from unitarily invariant measurements is investigated in this paper. A message-passing algorithm is formulated on the basis of expectation propagation (EP). A rigorous analysis is presented for the dynamics of the algorithm in the large system limit, where both input and output dimensions tend to infinity while the compression rate is kept constant. The main result is the justification of state evolution (SE) equations conjectured by Ma and Ping. This result implies that the EP-based algorithm achieves the Bayes-optimal performance that was originally derived via a non-rigorous tool in statistical physics and proved partially in a recent paper, when the compression rate is larger than a threshold. The proof is based on an extension of a conventional conditioning technique for the standard Gaussian matrix to the case of the Haar matrix.

Index Terms—Compressed sensing, expectation propagation, unitarily invariant measurements, state evolution, Haar matrices.

I. INTRODUCTION

A. Motivation

CONSIDER the recovery problem of an N -dimensional signal vector \mathbf{x} from a compressed noisy measurement vector $\mathbf{y} \in \mathbb{C}^M$ ($M \leq N$) [2], [3],

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}. \quad (1)$$

In (1), $\mathbf{A} \in \mathbb{C}^{M \times N}$ denotes a known measurement matrix. The signal vector \mathbf{x} is an unknown sparse¹ vector that is composed of independent and identically distributed (i.i.d.) elements. The noise vector $\mathbf{w} \in \mathbb{C}^M$ is independent of the other random variables. The goal of compressed sensing is to recover the sparse vector \mathbf{x} from the knowledge about \mathbf{y} and \mathbf{A} , as well as the statistics of all random variables.

A breakthrough for signal recovery is to construct message-passing (MP) that is Bayes-optimal in the large system limit, where the input and output dimensions N and M tend to infinity while the compression rate $\delta = M/N$ is kept constant.

Manuscript received April *, 2017. The author was in part supported by the Grant-in-Aid for Exploratory Research (JSPS KAKENHI Grant Number 15K13987), Japan and by the Grant-in-Aid for Scientific Research (B) (JSPS KAKENHI Grant Number 18H01441), Japan. The material in this paper was presented in part at 2017 IEEE International Symposium on Information Theory, Aachen, Germany, Jun. 2017 [1].

K. Takeuchi is with the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Aichi 441-8580, Japan (e-mail: takeuchi@ee.tut.ac.jp).

¹ In this paper, a signal $x \in \mathbb{R}$ is called sparse if the Rényi information dimension [4] of x is smaller than 1. If x is zero with probability $1 - p$, the information dimension is at most p . If x is discrete, it is zero.

The origin of this approach dates back to the Thouless-Anderson-Palmer (TAP) equation [5] in statistical physics. Motivated by the TAP approach, Kabashima [6] proposed an MP algorithm based on approximate belief propagation (BP) in the context of code-division multiple-access (CDMA) systems with i.i.d. zero-mean measurement matrices. When the compression rate is larger than the so-called BP threshold [7], the BP-based algorithm was numerically shown to achieve the Bayes-optimal performance in the large system limit, which was originally conjectured by Tanaka [8] via the replica method—a non-rigorous tool in statistical physics, and proved in [9], [10] for i.i.d. zero-mean Gaussian measurements. However, Kabashima [6] presented no rigorous analysis on the convergence property of the BP-based algorithm.

In order to resolve lack of a rigorous proof, approximate message-passing (AMP) was proposed in [11] and proved in [12] to achieve the optimal performance for i.i.d. zero-mean Gaussian measurements, when the compression rate is larger than the BP threshold. Spatially coupled measurement matrices are required for achieving the optimal performance in the whole regime [7], [13]–[15]. However, it is recognized that AMP fails to converge when the i.i.d. zero-mean assumption of measurement matrices is broken [16], unless damping [17] is employed.

As solutions to this convergence issue, since Opper and Winther’s pioneering work [18, Appendix D], as well as [19], several algorithms have been proposed on the basis of expectation propagation (EP) [20], expectation consistent (EC) approximations [18], [21], [22], S-transform [23], vector AMP [24], or turbo principle [25]–[28]. The EP-based algorithm [20] is systematically derived from Minka’s EP framework [29] by approximating the posterior distribution of \mathbf{x} with factorized Gaussian distributions. The EC-based algorithms [18], [21], [22] are iterative algorithms for solving a fixed point (FP) of the EC free energy. An algorithm in [23] is derived via the S-transform of $\mathbf{A}^H \mathbf{A}$. Rangan *et al.* [24] considered an EP-like approximation of the BP algorithm on a factor graph with vector-valued nodes. The algorithms [25]–[28] based on turbo principle are derived from a few heuristic assumptions. Interestingly, the algorithms in [18], [20], [22], [24], [27] are essentially equivalent, with the exception of [21], [23]. In this paper, these algorithms for signal recovery are simply referred to as EP-based algorithms, since we follow the EP-based derivation in [20].

Ma *et al.* [26], [27] derived state evolution (SE) equations of an EP-based algorithm under two heuristic assumptions. By investigating the properties of the SE equations, they conjectured that, for unitarily invariant measurement matrices, the FPs of the SE equations are the same as the extrema of an asymptotic energy function that describes the Bayes-optimal performance in the large system limit. The energy function was originally derived in [30], [31] via the replica method, and proved for bounded signals in [32]. In other words, the EP-based algorithm was conjectured to achieve the optimal performance in the large system limit, when the compression rate is larger than the BP threshold. Since the algorithm attempts to solve the minimum of the EC free energy [18], it is conjectured that the extrema of the EC free energy correspond to those of the Bayes-optimal one for unitarily invariant measurement matrices. The purpose of this paper is to present a rigorous proof for the conjecture.

B. Proof Strategy

The proof strategy is based on a conditioning technique used in [12], originally proposed by Bolthausen [33]. A challenging part in the proof is to evaluate the distribution of an estimation error in each iteration conditioned on estimation errors in all preceding iterations. Bayati and Montanari [12] evaluated the conditional distribution via the distribution of the measurement matrix \mathbf{A} conditioned on the estimation errors in all preceding iterations. When linear detection is employed as part of MP, the conditional distribution of \mathbf{A} can be regarded as the posterior distribution of \mathbf{A} given linear, noiseless, and compressed observations of \mathbf{A} , determined by the estimation errors in all preceding iterations. For i.i.d. Gaussian measurement matrices, it is well known that the posterior distribution is also Gaussian. The proof in [12] heavily relies on this well-known fact.

In order to present our proof strategy, assume $M = N$, and that \mathbf{A} is a Haar matrix [34], [35], which is uniformly distributed on the space of all possible $N \times N$ unitary matrices. Under appropriate coordinate rotations in the column spaces of \mathbf{A} , it is possible to show that the linear, noiseless, and compressed observation of \mathbf{A} is equivalent to observing *part* of the columns in \mathbf{A} . Since any Haar matrix is bi-unitarily invariant [34], the distribution of \mathbf{A} after the coordinate rotations is the same as the original one. Thus, evaluating the conditional distribution of \mathbf{A} reduces to analyzing the conditional distribution of a Haar matrix given part of its columns. This argument was implicitly used in [12].

Evaluation of this conditional distribution is an important part in this paper, while this part is not required for i.i.d. Gaussian measurements. For simplicity, let $N = 3$ and fix the first column of a Haar matrix $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. Evaluation of the conditional distribution is equivalent to characterizing \mathbf{a}_2 and \mathbf{a}_3 for given \mathbf{a}_1 . The two vectors must be on a plane perpendicular to \mathbf{a}_1 . From the orthonormality between \mathbf{a}_2 and \mathbf{a}_3 , the two vectors are on a unit circle that has the center at the intersection of the plane and $c\mathbf{a}_1$ for $c \in \mathbb{C}$. Intuitively, \mathbf{a}_2 and \mathbf{a}_3 should be Haar-distributed on this unit circle. Generalizing this intuition, we find that \mathbf{A} given its first t columns should have degrees of freedom that are equal to

those of an $(N - t) \times (N - t)$ Haar matrix. On the basis of this intuition, we evaluate the conditional distribution of \mathbf{A} .

C. Related Work

A similar paper [36] was posted on the arXiv a few months before posting the first version [37] of this paper. Short versions of the two papers were published in [1], [24]. The posted paper [36] addressed real-valued systems, while we consider complex-valued systems. Interestingly, the two papers share the common proof strategy based on [12]. However, there is a mathematically critical difference between them.

The main difference is in mathematical treatments on almost sure convergence. An empirical convergence based on pseudo-Lipschitz functions was considered in [36]. The approach allows us to analyze general Lipschitz-continuous decision functions and general pseudo-Lipschitz performance measures, as considered in [12]. However, Rangan *et al.* [36] omitted the proof of an important part on almost sure convergence—required in establishing the empirical convergence based on pseudo-Lipschitz functions—as pointed out in Appendix A-B.

In this paper, we present a rigorous proof of the part on almost sure convergence. Our approach relies on advanced results in probability theory, such as the strong law of large numbers for dependent random variables [38] and statistical properties of a Haar matrix. While this paper considers a Bayes-optimal decision function and the mean-square error (MSE), the part on almost sure convergence is proved in a general form, as considered in [36]. Thus, combining [36] and this paper establishes a rigorous proof of SE for general decision functions and general performance measures.

D. Contributions

The main contribution is the rigorous justification of the SE equations for the EP-based algorithm, conjectured in [27]. More precisely, we derive SE equations for individual elements of the signal vector in the large system limit. This implies the achievability of the Bayes-optimal performance proved in [32], when the compression rate is larger than the BP threshold, while the converse theorem is partially open, i.e. there are no algorithms outperforming the EP-based algorithm in the large system limit when unbounded signals are considered.

The technical novelty is in an extension of the conditioning technique in [12] for i.i.d. Gaussian measurement matrices to the case of Haar matrices. This paper presents a constructive proof for the conditional distribution of a Haar matrix. The proposed conditioning technique is applicable to any MP algorithm for signal recovery from unitarily invariant measurements, such as the AMP, unless the algorithm contains nonlinear processing in the measurement vector \mathbf{y} , e.g. quantization [28]. However, whether the obtained SE becomes simple depends on the MP algorithm and the statistics of \mathbf{A} [39]. Thus, it is an important future work to design a low-complexity MP algorithm such that simple SE equations are obtained for unitarily invariant measurements.

E. Organization

The remainder of this paper is organized as follows: After summarizing the notation used in this paper, Section II presents the definition of unitarily invariant matrices and technical results associated with Haar matrices. In Section III, we introduce assumptions used throughout this paper, and then formulate an EP-based algorithm. The main result is presented in Section IV, and proved in Section V. Several technical results are proved in appendices.

F. Notation

The notation $\mathcal{o}(1)$ denotes a vector of which the Euclidean norm converges almost surely toward zero in the large system limit. For a vector $\mathbf{v} \in \mathbb{C}^N$, we write the n th element of \mathbf{v} as v_n . For a subset $\mathcal{N} \subset \{1, \dots, N\}$, the vector $\mathbf{x}_{\mathcal{N}}$ consists of the elements $\{x_n : n \in \mathcal{N}\}$, while $\mathbf{x}_{\setminus \mathcal{N}}$ is obtained by eliminating $\{x_n : n \in \mathcal{N}\}$ from \mathbf{x} . For a scalar function $f : \mathbb{C} \rightarrow \mathbb{C}$, we introduce a convention in which $f(\mathbf{v})$ denotes the vector obtained by the component-wise application of f to \mathbf{v} , i.e. $[f(\mathbf{v})]_n = f(v_n)$.

For a complex number $z \in \mathbb{C}$ and a matrix $\mathbf{M} \in \mathbb{C}^{M \times N}$, the complex conjugate, transpose, and the conjugate transpose are denoted by z^* , \mathbf{M}^T , and \mathbf{M}^H . We write the (m, n) th element of \mathbf{M} as M_{mn} . When \mathbf{M} is Hermitian, $\lambda_{\min}(\mathbf{M})$ represents the minimum eigenvalue of \mathbf{M} . For $M \geq N$, $\mathcal{U}_{M \times N}$ denotes the space of all possible $M \times N$ matrices with orthonormal columns, while $\mathcal{U}_{M \times N}$ for $M < N$ represents the space of all possible $M \times N$ matrices with orthonormal rows. When $M = N$ holds, $\mathcal{U}_{M \times N}$ is written as \mathcal{U}_N , which is the space of all possible $N \times N$ unitary matrices.

We write the singular-value decomposition (SVD) of \mathbf{M} as

$$\mathbf{M} = \Phi_{\mathbf{M}}(\Sigma_{\mathbf{M}}, \mathbf{O})\Psi_{\mathbf{M}}^H \quad (2)$$

for $M \leq N$, with $\Phi_{\mathbf{M}} \in \mathcal{U}_M$ and $\Psi_{\mathbf{M}} \in \mathcal{U}_N$. Furthermore, $\Sigma_{\mathbf{M}}$ is an $M \times M$ positive semi-definite diagonal matrix. The unitary matrix $\Psi_{\mathbf{M}}$ is partitioned as $\Psi_{\mathbf{M}} = (\Psi_{\mathbf{M}}^{\parallel}, \Psi_{\mathbf{M}}^{\perp})$, in which $\Psi_{\mathbf{M}}^{\parallel} \in \mathcal{U}_{N \times M}$ is composed of the first M columns of $\Psi_{\mathbf{M}}$, while $\Psi_{\mathbf{M}}^{\perp} \in \mathcal{U}_{N \times (N-M)}$ consists of the remaining columns. For $M > N$, we write the SVD of \mathbf{M} as

$$\mathbf{M} = \Phi_{\mathbf{M}} \begin{pmatrix} \Sigma_{\mathbf{M}} \\ \mathbf{O} \end{pmatrix} \Psi_{\mathbf{M}}^H, \quad (3)$$

with $\Phi_{\mathbf{M}} \in \mathcal{U}_M$ and $\Psi_{\mathbf{M}} \in \mathcal{U}_N$. Furthermore, $\Sigma_{\mathbf{M}}$ is an $N \times N$ positive semi-definite diagonal matrix. The unitary matrix $\Phi_{\mathbf{M}} = (\Phi_{\mathbf{M}}^{\parallel}, \Phi_{\mathbf{M}}^{\perp})$ is partitioned in the same manner as for $M \leq N$.

When \mathbf{M} is full rank, the pseudo-inverse of \mathbf{M} is denoted by $\mathbf{M}^{\dagger} = (\mathbf{M}^H \mathbf{M})^{-1} \mathbf{M}^H \in \mathbb{C}^{N \times M}$ for $M > N$. Let $\mathbf{P}_{\mathbf{M}}^{\parallel}$ denote the orthogonal projection matrix onto the space spanned by the columns of \mathbf{M} . We have $\mathbf{P}_{\mathbf{M}}^{\parallel} = \Phi_{\mathbf{M}}^{\parallel}(\Phi_{\mathbf{M}}^{\parallel})^H = \mathbf{M} \mathbf{M}^{\dagger}$. The projection matrix $\mathbf{P}_{\mathbf{M}}^{\perp}$ onto the orthogonal complement is given by $\mathbf{P}_{\mathbf{M}}^{\perp} = \mathbf{I}_M - \mathbf{P}_{\mathbf{M}}^{\parallel}$. For $M \leq N$, we define $\mathbf{M}^{\dagger} = \mathbf{M}^H (\mathbf{M} \mathbf{M}^H)^{-1}$, $\mathbf{P}_{\mathbf{M}}^{\parallel} = \Psi_{\mathbf{M}}^{\parallel}(\Psi_{\mathbf{M}}^{\parallel})^H = \mathbf{M}^{\dagger} \mathbf{M}$, and $\mathbf{P}_{\mathbf{M}}^{\perp} = \mathbf{I}_N - \mathbf{P}_{\mathbf{M}}^{\parallel}$.

The proper complex Gaussian distribution with mean \mathbf{m} and covariance Σ is denoted by $\mathcal{CN}(\mathbf{m}, \Sigma)$. The expectation

and variance of a random variable X is denoted by $\mathbb{E}[X]$ and $\mathbb{V}[X]$, respectively. The notation $X \stackrel{\text{a.s.}}{=} Y$ means that X is almost surely equal to Y . Similarly, $\stackrel{\text{a.s.}}{\rightarrow}$, $\stackrel{\text{a.s.}}{\geq}$, and $\stackrel{\text{a.s.}}{\leq}$ indicate that \rightarrow , \geq , and \leq hold almost surely. The notation $X \sim Y$ means that X follows the same distribution as Y . The notation $X|_Y$ indicates that we focus on the conditional distribution of X given Y .

II. PRELIMINARIES

A. Definitions

The purpose of this section is to present the strong law of large numbers for a Haar matrix. The result corresponds to [12, Lemma 2] for an i.i.d. Gaussian matrix. We first present several definitions.

Definition 1: A unitary random matrix $\mathbf{U} \in \mathcal{U}_n$ is called a Haar matrix if \mathbf{U} is uniformly distributed on \mathcal{U}_n .

An important property of a Haar matrix is bi-unitary invariance [35]—used throughout this paper.

Definition 2: A random matrix \mathbf{M} is said to be bi-unitarily invariant if $\mathbf{M} \sim \mathbf{U} \mathbf{M} \mathbf{V}$ holds for all deterministic unitary matrices \mathbf{U} and \mathbf{V} .

In this paper, the functions $z^* f(x+z)$ and $|x - f(x+z)|^2$ of $x \in \mathbb{C}$ and $z \in \mathbb{C}$ are considered for a Lipschitz-continuous function $f : \mathbb{C} \rightarrow \mathbb{C}$. To characterize these functions, we follow [12] to define pseudo-Lipschitz functions.

Definition 3: For $k \geq 1$, we say that a function $f : \mathbb{C}^n \rightarrow \mathbb{C}$ is pseudo-Lipschitz of order k if there is some Lipschitz constant $L > 0$ such that

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\| (1 + \|\mathbf{x}\|^{k-1} + \|\mathbf{y}\|^{k-1}) \quad (4)$$

holds for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.

Note that any pseudo-Lipschitz function of order 1 is Lipschitz-continuous. A pseudo-Lipschitz function $f(\mathbf{x})$ of order k is $\mathcal{O}(\|\mathbf{x}\|^k)$ as $\|\mathbf{x}\| \rightarrow \infty$.

The following proposition is used for further evaluation of the upper bound (4) throughout this paper.

Proposition 1: For any $k \geq 1$, there is some constant $C > 0$ such that

$$(a+b)^k \leq C(a^k + b^k) \quad (5)$$

holds for all $a \geq 0$ and $b \geq 0$.

Proof: The inequality follows from a general upper bound $\|\cdot\|_1 \leq 2^{1-1/k} \|\cdot\|_k$ on \mathbb{C}^2 . ■

B. Results

We consider an array $\{\mathbf{X}_N \in \mathbb{C}^N\}_{N=1}^{\infty}$ of dependent random variables $\mathbf{X}_N = (X_{1,N}, \dots, X_{N,N})^T$. An array $\{\mathbf{X}_N\}$ allows the distribution of each element $X_{n,N}$ to change as N grows, while the distribution of each element is fixed in a sequence $\mathbf{X} \in \mathbb{C}^N$. We first present the strong law of large numbers for an array $\{\mathbf{X}_N \in \mathbb{C}^N\}_{N=1}^{\infty}$ of dependent random variables.

Theorem 1 (Lyons [38]): Let $\{\mathbf{X}_N\}$ denote an array of complex random variables with finite second moments, and define $S_N = \sum_{n=1}^N X_{n,N}$. If the following assumption holds:

$$\sum_{N=1}^{\infty} \frac{\sqrt{\mathbb{V}[S_N]}}{N^2} < \infty, \quad (6)$$

the strong law of large numbers for $T_N = (S_N - \mathbb{E}[S_N])/N$ holds, i.e. $\lim_{N \rightarrow \infty} T_N \stackrel{\text{a.s.}}{=} 0$.

Proof: Lyons [38, Theorem 6] proved Theorem 1 for a sequence of complex random variables, i.e. $X_{n,N} = X_{n,N'}$ for all $N \neq N'$. However, the proof is applicable to the array case with no changes, by defining $X_{n,N} = 0$ for $n > N$. Thus, Theorem 1 holds. ■

The condition (6) is satisfied when $\mathbb{V}[S_N] = \mathcal{O}(N^\alpha)$ holds for some $\alpha < 2$. In particular, we have $\alpha = 1$ when $\{X_{n,N}\}$ are uncorrelated random variables. We next present the strong law of large numbers associated with a Haar matrix.

Lemma 1: For $t' \in \mathbb{N}$, suppose that $f_n : \mathbb{C}^{t'+1} \rightarrow \mathbb{C}$ denote a pseudo-Lipschitz function of order k with a Lipschitz constant $L_n > 0$. Let $\epsilon_N = (\epsilon_{1,N}, \dots, \epsilon_{N,N})^T \in \mathbb{C}^N$ denote a vector that satisfies

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N L_n |\epsilon_{n,N}|^2 \stackrel{\text{a.s.}}{=} 0, \quad (7)$$

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N L_n |\epsilon_{n,N}|^{2k-2} < \infty. \quad (8)$$

Suppose that $\mathbf{a}_{\tau,N} = (a_{\tau,1,N}, \dots, a_{\tau,N,N})^T \in \mathbb{C}^N$ for $\tau = 0, \dots, t'$ satisfies

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N L_n^i |a_{\tau,n,N}|^{2k-2} < \infty \quad \text{for } i = 1, 2. \quad (9)$$

For $t > 0$, let $\mathbf{E}_N = (e_{1,N}^T, \dots, e_{N,N}^T)^T \in \mathbb{C}^{N \times t}$ denote a matrix that satisfies

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N L_n \|\mathbf{e}_{n,N}\|^{\max\{2, 2k-2\}} < \infty, \quad (10)$$

$$\liminf_{N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} \mathbf{E}_N^H \mathbf{E}_N \right) \stackrel{\text{a.s.}}{>} C \quad (11)$$

for some constant $C > 0$. Suppose that $\{\mathbf{X}_N \in \mathbb{C}^N\}$ is an array of unitarily invariant random variables conditioned on ϵ_N , $\{\mathbf{a}_{\tau,N}\}$, and \mathbf{E}_N , i.e. $\Phi \mathbf{X}_N \sim \mathbf{X}_N$ conditioned on ϵ_N , $\{\mathbf{a}_{\tau,N}\}$, and \mathbf{E}_N for any deterministic unitary matrix $\Phi \in \mathcal{U}_N$. For some $v > 0$, postulate the following:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \|\mathbf{X}_N\|^2 \stackrel{\text{a.s.}}{=} v > 0. \quad (12)$$

Let $z \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ denote a standard complex Gaussian random vector. Then, the following two properties hold:

1) Postulate the following assumptions:

- ϵ_N has finite $(2k-2)$ th moments and vanishing second moments, i.e. $\mathbb{E}[|\epsilon_{n,N}|^2] \rightarrow 0$ as $N \rightarrow \infty$.
- $\mathbf{a}_{\tau,N}$ has finite $(2k-2)$ th moments.
- \mathbf{E}_N has finite $\max\{2, 2k-2\}$ th moments.
- \mathbf{X}_N has finite $(\max\{2, 2k-2\} + \epsilon)$ th moments for some $\epsilon > 0$.

Then, for any $t \geq 0$

$$\begin{aligned} & \lim_{N \rightarrow \infty} \mathbb{E} \left[f_n(a_{n,0,N}, \dots, a_{n,t'-1,N}, \right. \\ & \quad \left. a_{n,t',N} + \epsilon_{n,N} + [\Phi_{\mathbf{E}_N}^\perp \mathbf{X}_{N-t}]_n \right) \\ & = \mathbb{E} \left[f_n(a_{n,0,N}, \dots, a_{n,t'-1,N}, a_{n,t',N} + \sqrt{v}z_n) \right], \end{aligned} \quad (13)$$

where the convention $\Phi_{\mathbf{E}_N}^\perp = \mathbf{I}_N$ is introduced for $t = 0$.
2) If the sequence of Lipschitz constants satisfies

$$\frac{1}{N} \sum_{n=1}^N L_n^2 < \infty, \quad (14)$$

then for $t \geq 0$

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \left\{ f_n(a_{n,0,N}, \dots, a_{n,t'-1,N}, \right. \\ & \quad \left. a_{n,t',N} + \epsilon_{n,N} + [\Phi_{\mathbf{E}_N}^\perp \mathbf{X}_{N-t}]_n \right) \\ & - \mathbb{E}_{z_n} [f_n(a_{n,0,N}, \dots, a_{n,t'-1,N}, a_{n,t',N} + \sqrt{v}z_n)] \Big\} \stackrel{\text{a.s.}}{=} 0. \end{aligned} \quad (15)$$

Proof: See Appendix A. ■

Lemma 1 is used repeatedly to prove the main theorem of this paper. Finally, we prove the following corollary that is used in the derivation of the EP-based algorithm.

Corollary 1: Let $\mathbf{a} \in \mathbb{C}^N$ denote a vector that satisfies $\lim_{N \rightarrow \infty} N^{-1} \|\mathbf{a}\|^2 \stackrel{\text{a.s.}}{=} 1$. Suppose that $\mathbf{D} \in \mathbb{C}^{N \times N}$ is a Hermitian matrix with $\lim_{N \rightarrow \infty} N^{-1} \text{Tr}(\mathbf{D}^i) \stackrel{\text{a.s.}}{=} d_i$ for $i = 1, 2$. Let $\mathbf{V} \in \mathcal{U}_N$ denote a Haar matrix independent of \mathbf{a} and \mathbf{D} . Then,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{a}^H \mathbf{V}^H \mathbf{D} \mathbf{V} \mathbf{a} \stackrel{\text{a.s.}}{=} d_1. \quad (16)$$

Proof: Without loss of generality, we can assume that \mathbf{D} is diagonal since \mathbf{V} is a Haar matrix. For $\mathbf{X}_N = \mathbf{V} \mathbf{a}$, we have

$$\frac{1}{N} \mathbf{a}^H \mathbf{V}^H \mathbf{D} \mathbf{V} \mathbf{a} = \frac{1}{N} \sum_{n=1}^N f_n(X_{n,N}), \quad (17)$$

with $f_n(z) = D_n |z|^2$, in which D_n denotes the n th diagonal element of \mathbf{D} . Since f_n is a pseudo-Lipschitz function of order 2 with the Lipschitz constant $|D_n|$, the assumptions on \mathbf{a} and \mathbf{D} imply that all assumptions in Lemma 1 are satisfied with $v = 1$. Thus, we use Lemma 1 to arrive at

$$\frac{1}{N} \mathbf{a}^H \mathbf{V}^H \mathbf{D} \mathbf{V} \mathbf{a} \stackrel{\text{a.s.}}{=} \frac{1}{N} \sum_{n=1}^N D_n \mathbb{E}[|z_n|^2] + o(1) \xrightarrow{\text{a.s.}} d_1 \quad (18)$$

as $N \rightarrow \infty$, which implies Corollary 1. ■

III. SYSTEM MODEL

A. Assumptions

Assumptions on the measurement model (1) are presented.

Assumption 1: The signal vector \mathbf{x} is composed of zero-mean i.i.d. *non-Gaussian* elements with unit variance and finite $(2 + \epsilon)$ th moments for some $\epsilon > 0$.

From the strong law of large numbers [40], Assumption 1 implies that $N^{-1} \|\mathbf{x}\|^2$ converges almost surely to 1 as $N \rightarrow \infty$. The i.i.d. assumption for \mathbf{x} is implicitly used in the derivation of an EP-based algorithm. We require no additional assumptions for the prior distribution of each element to prove the main theorem, whereas it is practically important to postulate some prior distribution indicating the sparsity of \mathbf{x} .

Definition 4: A Hermitian random matrix \mathbf{M} is said to be unitarily invariant if $\mathbf{M} \sim \mathbf{U} \mathbf{M} \mathbf{U}^H$ holds for any deterministic unitary matrix \mathbf{U} .

Assumption 2: The measurement matrix \mathbf{A} has the following properties:

- $\mathbf{A}^H \mathbf{A}$ is unitarily invariant.
- The empirical eigenvalue distribution of $\mathbf{A} \mathbf{A}^H$ converges almost surely to a deterministic distribution $\rho(\lambda)$ with a compact support in the large system limit.

We write the SVD of \mathbf{A} as

$$\mathbf{A} = \mathbf{U}(\boldsymbol{\Sigma}, \mathbf{O})\mathbf{V}^H, \quad (19)$$

with $\mathbf{U} \in \mathcal{U}_M$ and $\mathbf{V} \in \mathcal{U}_N$. Furthermore, $\boldsymbol{\Sigma}$ is an $M \times M$ positive semi-definite diagonal matrix. From Assumption 2, \mathbf{V} is a Haar matrix and independent of $\mathbf{U}\boldsymbol{\Sigma}$ [35].

Assumption 3: The noise vector \mathbf{w} has finite $(2 + \epsilon)$ th moments for some $\epsilon > 0$. Let $\mathbf{D} \in \mathbb{C}^{M \times M}$ denote any Hermitian matrix such that \mathbf{D} is independent of $\mathbf{U}^H \mathbf{w}$, and that $M^{-1} \text{Tr}(\mathbf{D}^2)$ converges almost surely as $M \rightarrow \infty$. Then,

$$\lim_{M \rightarrow \infty} \frac{1}{M} \left\{ \mathbf{w}^H \mathbf{U} \mathbf{D} \mathbf{U}^H \mathbf{w} - \sigma^2 \text{Tr}(\mathbf{D}) \right\} \stackrel{\text{a.s.}}{=} 0. \quad (20)$$

Assumption 3 implies that σ^2 corresponds to the noise power $\sigma^2 \stackrel{\text{a.s.}}{=} \lim_{M \rightarrow \infty} M^{-1} \|\mathbf{w}\|^2$ per element, by selecting $\mathbf{D} = \mathbf{I}_M$. Assumption 3 is satisfied if \mathbf{w} is unitarily invariant, e.g. $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$, or if \mathbf{U} is a Haar matrix.

B. Expectation Propagation

We start with an MP algorithm proposed in [27]. Let the detector postulate that the noise vector \mathbf{w} in (1) is a circularly symmetric complex Gaussian (CSCG) random vector with covariance $\sigma^2 \mathbf{I}_M$. This postulation needs not be consistent with the true distribution of \mathbf{w} .

As derived in Appendix B, the MP algorithm for this case is based on EP and composed of two modules. In iteration t , a first module—called module A—calculates the *extrinsic* mean $\mathbf{x}_{A \rightarrow B}^t$ and variance $v_{A \rightarrow B}^t$ of the signal vector \mathbf{x} from $\mathbf{x}_{B \rightarrow A}^t$ and $v_{B \rightarrow A}^t$ provided by the other module—called module B.

$$\mathbf{x}_{A \rightarrow B}^t = \mathbf{x}_{B \rightarrow A}^t + \gamma_t \mathbf{W}^t (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A}^t), \quad (21)$$

$$v_{A \rightarrow B}^t = \gamma_t - v_{B \rightarrow A}^t. \quad (22)$$

In the initial iteration $t = 0$, the prior mean $\mathbf{x}_{B \rightarrow A}^0 = \mathbf{0}$ and variance $v_{B \rightarrow A}^0 = N^{-1} \mathbb{E}[\|\mathbf{x}\|^2] = 1$ are used.

In (21), the linear minimum mean-square error (LMMSE) filter $\mathbf{W}^t \in \mathbb{C}^{N \times M}$ is given by

$$\mathbf{W}^t = \mathbf{A}^H \left(\sigma^2 \mathbf{I}_M + v_{B \rightarrow A}^t \mathbf{A} \mathbf{A}^H \right)^{-1}. \quad (23)$$

The normalization coefficient² γ_t in (21) is defined as

$$\frac{1}{\gamma_t} = \lim_{M = \delta N \rightarrow \infty} \frac{1}{N} \text{Tr}(\mathbf{W}^t \mathbf{A}) \stackrel{\text{a.s.}}{=} \frac{1}{\gamma(v_{B \rightarrow A}^t)} \quad (24)$$

due to Assumption 2, with

$$\frac{1}{\gamma(v)} = \int \frac{\delta \lambda}{\sigma^2 + v \lambda} d\rho(\lambda), \quad (25)$$

where $\rho(\lambda)$ denotes the asymptotic eigenvalue distribution of $\mathbf{A} \mathbf{A}^H$ in the large system limit. The coefficient γ_t keeps the orthogonality between estimation errors in the two modules.

² $\gamma_t^{-1} = N^{-1} \text{Tr}(\mathbf{W}^t \mathbf{A})$ may be used in practical situations.

On the other hand, module B computes the minimum mean-square error (MMSE) estimator and the posterior variance of \mathbf{x}

$$\tilde{\eta}_t(\mathbf{x}_{A \rightarrow B}^t) = \mathbb{E}[\mathbf{x} | \mathbf{x}_{A \rightarrow B}^t], \quad (26)$$

$$v_B^{t+1} = \frac{1}{N} \left\{ \mathbb{E}[\|\mathbf{x}\|^2 | \mathbf{x}_{A \rightarrow B}^t] - \|\tilde{\eta}_t(\mathbf{x}_{A \rightarrow B}^t)\|^2 \right\}, \quad (27)$$

given the virtual additive white Gaussian noise (AWGN) observation,

$$\mathbf{x}_{A \rightarrow B}^t = \mathbf{x} + \mathbf{z}^t, \quad \mathbf{z}^t \sim \mathcal{CN}(\mathbf{0}, v_{A \rightarrow B}^t \mathbf{I}_N). \quad (28)$$

If a termination condition is satisfied, module B outputs $\tilde{\eta}_t(\mathbf{x}_{A \rightarrow B}^t)$ as an estimate of \mathbf{x} . Otherwise, module B feeds the extrinsic mean $\mathbf{x}_{B \rightarrow A}^{t+1}$ and variance $v_{B \rightarrow A}^{t+1}$ of \mathbf{x} back to module A, given by

$$\mathbf{x}_{B \rightarrow A}^{t+1} = \eta_t(\mathbf{x}_{A \rightarrow B}^t), \quad (29)$$

$$\frac{1}{v_{B \rightarrow A}^{t+1}} = \frac{1}{v_B^{t+1}} - \frac{1}{v_{A \rightarrow B}^t}, \quad (30)$$

where the extrinsic decision function $\eta_t : \mathbb{C} \rightarrow \mathbb{C}$ is defined as

$$\eta_t(z) = v_{B \rightarrow A}^{t+1} \left(\frac{\tilde{\eta}_t(z)}{v_B^{t+1}} - \frac{z}{v_{A \rightarrow B}^t} \right). \quad (31)$$

Remark 1: The extrinsic decision function (31) is zero if $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ holds. We have postulated Assumption 1 to avoid a constant decision function.

It is not trivial whether the posterior variance (27) is bounded. Therefore, we postulate the following assumption:

Assumption 4: Each posterior variance $\mathbb{E}[|x_n|^2 | x_{n, A \rightarrow B}^t] - |\tilde{\eta}_t(x_{n, A \rightarrow B}^t)|^2$ is almost surely bounded.

Assumption 4 is a necessary condition for utilizing the EP-based algorithm in practical situations. The author believes that Assumption 4 can be proved without additional conditions.

We present important properties of the Bayes-optimal decision function $\tilde{\eta}_t$ in module B. We start with the definition of the Wirtinger derivative of a complex function.

Definition 5 (Wirtinger derivative): For a complex number $z = x + iy$, the Wirtinger derivative is defined as

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right). \quad (32)$$

For a complex function $f : \mathbb{C} \rightarrow \mathbb{C}$, we write $(\partial/\partial z)(\Re[f] + i\Im[f])$ as $\partial f/\partial z$.

Lemma 2 (Ma and Ping [27]): Suppose that $z \sim \mathcal{CN}(0, v_{A \rightarrow B}^t)$ is a CSCG random variable with variance $v_{A \rightarrow B}^t$ and independent of x_n . Then, the decision function $\tilde{\eta}_t$ is Lipschitz-continuous and satisfies

$$\mathbb{E}_z [z^* \tilde{\eta}_t(x_n + z)] = v_{A \rightarrow B}^t \mathbb{E}_z \left[\frac{\partial \tilde{\eta}_t}{\partial z}(x_n + z) \right], \quad (33)$$

$$\mathbb{E} [z^* \tilde{\eta}_t(x_n + z)] = \text{MMSE}(v_{A \rightarrow B}^t) \quad (34)$$

for any n , where $\text{MMSE}(v_{A \rightarrow B}^t)$ denotes the MMSE based on an AWGN observation, given by

$$\text{MMSE}(v_{A \rightarrow B}^t) = \mathbb{E} [|x_n - \tilde{\eta}_t(x_n + z)|^2]. \quad (35)$$

Proof: See Appendix C for the proof based on [27]. ■

Lemma 2 is used to prove the orthogonality between estimation errors in the two modules. The identity (33) is a generalization of Stein's lemma [41] to the complex-valued case.

Remark 2: As considered in [24], [27], we can replace the decision function $\tilde{\eta}_t$ with another suboptimal function. Such a replacement may be important when the true prior distribution of the signal elements is unknown. Nonetheless, for simplicity, we only consider the optimal decision function $\tilde{\eta}_t$. See [24] for a generalization of the decision function.

C. Error Recursion

An error recursion for the EP-based algorithm is formulated to analyze the convergence property. Let $\mathbf{h}_t = \mathbf{x}_{A \rightarrow B}^t - \mathbf{x}$ and $\mathbf{q}_t = \mathbf{x}_{B \rightarrow A}^t - \mathbf{x}$ denote the estimation errors for the extrinsic estimates in modules A and B, respectively. Substituting the system model (1) into the update rule (21) of $\mathbf{x}_{A \rightarrow B}^t$, and using the SVD (19) and the update rule (29) of $\mathbf{x}_{B \rightarrow A}^t$, we obtain the error recursion

$$\mathbf{b}_t = \mathbf{V}^H \mathbf{q}_t, \quad (36)$$

$$\mathbf{m}_t = \mathbf{b}_t - \gamma_t \tilde{\mathbf{W}}_t \{(\boldsymbol{\Sigma}, \mathbf{O})\mathbf{b}_t - \tilde{\mathbf{w}}\}, \quad (37)$$

$$\mathbf{h}_t = \mathbf{V} \mathbf{m}_t, \quad (38)$$

$$\mathbf{q}_{t+1} = \eta_t(\mathbf{x} + \mathbf{h}_t) - \mathbf{x}, \quad (39)$$

with $\tilde{\mathbf{w}} = \mathbf{U}^H \mathbf{w}$. In (37), the linear filter $\tilde{\mathbf{W}}_t$ is given by

$$\tilde{\mathbf{W}}_t = (\boldsymbol{\Sigma}, \mathbf{O})^H (\sigma^2 \mathbf{I}_M + v_{B \rightarrow A}^t \boldsymbol{\Sigma}^2)^{-1}. \quad (40)$$

Furthermore, we define $\eta_{-1}(\cdot) = 0$ to obtain $\mathbf{q}_0 = -\mathbf{x}$.

In analyzing the convergence property, we focus on the distribution of the estimation error \mathbf{h}_t conditioned on the preceding iteration history. Thus, it is useful to represent the error recursion in the matrix form. Define

$$\begin{aligned} \mathbf{Q}_t &= (\mathbf{q}_0, \dots, \mathbf{q}_{t-1}) \in \mathbb{C}^{N \times t}, \\ \mathbf{B}_t &= (\mathbf{b}_0, \dots, \mathbf{b}_{t-1}) \in \mathbb{C}^{N \times t}, \\ \mathbf{M}_t &= (\mathbf{m}_0, \dots, \mathbf{m}_{t-1}) \in \mathbb{C}^{N \times t}, \\ \mathbf{H}_t &= (\mathbf{h}_0, \dots, \mathbf{h}_{t-1}) \in \mathbb{C}^{N \times t}. \end{aligned} \quad (41)$$

The error recursion is represented as

$$\mathbf{V}^H \mathbf{Q}_t = \mathbf{B}_t, \quad (42)$$

$$\mathbf{M}_t = \mathbf{G}_t(\mathbf{B}_t, \tilde{\mathbf{w}}), \quad (43)$$

$$\mathbf{V} \mathbf{M}_t = \mathbf{H}_t, \quad (44)$$

$$\mathbf{Q}_{t+1} = \mathbf{F}_t(\mathbf{H}_t, \mathbf{x}), \quad (45)$$

where the τ th columns of $\mathbf{G}_t(\mathbf{B}_t, \tilde{\mathbf{w}})$ and $\mathbf{F}_t(\mathbf{H}_t, \mathbf{x})$ are equal to the right-hand sides (RHSs) of (37) and (39) for $t = \tau$, respectively.

The random vectors defined in Section III may have elements of which the distributions change as N grows. Thus, the subscript N should have been added in terms of the mathematical notation. Nonetheless, we have omitted the subscript N for notational simplicity.

IV. MAIN RESULT

Ma and Ping [27] conjectured that the following SE equations describe the dynamics of the EP-based algorithm in the large system limit:

$$\bar{v}_{A \rightarrow B}^t = \gamma(\bar{v}_{B \rightarrow A}^t) - \bar{v}_{B \rightarrow A}^t, \quad (46)$$

$$\frac{1}{\bar{v}_{B \rightarrow A}^{t+1}} = \frac{1}{\text{MMSE}(\bar{v}_{A \rightarrow B}^t)} - \frac{1}{\bar{v}_{A \rightarrow B}^t}, \quad (47)$$

with $\bar{v}_{B \rightarrow A}^0 = 1$, in which $\gamma(\cdot)$ and $\text{MMSE}(\cdot)$ are given in (25) and (35), respectively. The following theorem justifies their conjecture.

Theorem 2: Define $\bar{v}_{A \rightarrow B}^t$ and $\bar{v}_{B \rightarrow A}^t$ via the SE equations (46) and (47). Then, the following results hold in the large system limit:

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\mathbf{x}_{A \rightarrow B}^t - \mathbf{x}\|^2 \stackrel{\text{a.s.}}{=} \bar{v}_{A \rightarrow B}^t, \quad (48)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\tilde{\eta}_t(\mathbf{x}_{A \rightarrow B}^t) - \mathbf{x}\|^2 \stackrel{\text{a.s.}}{=} \text{MMSE}(\bar{v}_{A \rightarrow B}^t), \quad (49)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\eta_t(\mathbf{x}_{A \rightarrow B}^t) - \mathbf{x}\|^2 \stackrel{\text{a.s.}}{=} \bar{v}_{B \rightarrow A}^{t+1}. \quad (50)$$

The update rules (22) and (30) in the EP-based algorithm have the same representation as that in the SE equations (46) and (47). This implies that the EP-based algorithm predicts the exact dynamics of the extrinsic variances in the large system limit. The FPs of the SE equations were proved in [27] to correspond to those of an asymptotic energy function that describes the Bayes-optimal performance [30]–[32]. Thus, the Bayes-optimal performance is achievable when the SE equations have a unique FP, or equivalently when the compression rate δ is larger than the BP threshold.

The following theorem justifies the SE equations (46) and (47) in terms of individual MSEs.

Theorem 3: Define $\bar{v}_{A \rightarrow B}^t$ and $\bar{v}_{B \rightarrow A}^t$ via the SE equations (46) and (47). Then, for any n

$$\lim_{M=\delta N \rightarrow \infty} \mathbb{E}[\|\tilde{\eta}_t(x_{n,A \rightarrow B}^t) - x_n\|^2] = \text{MMSE}(\bar{v}_{A \rightarrow B}^t), \quad (51)$$

$$\lim_{M=\delta N \rightarrow \infty} \mathbb{E}[\|\eta_t(x_{n,A \rightarrow B}^t) - x_n\|^2] = \bar{v}_{B \rightarrow A}^{t+1}. \quad (52)$$

Remark 3: For simplicity, the individual MSE for the extrinsic estimate in module A is not analyzed in this paper. Furthermore, we have assumed the i.i.d. property of the elements of the signal vector \mathbf{x} . However, our proof strategy can be applied to justifying that the individual MSE $\mathbb{E}[\|x_{n,A \rightarrow B}^t - x_n\|^2]$ for module A converges to $\bar{v}_{A \rightarrow B}^t$ in the large system limit. Furthermore, the assumption on \mathbf{x} can be relaxed to the case of independent but non-identically distributed signals.

We shall introduce several notations to present a general theorem, of which corollaries are Theorems 2 and 3. The random variables in the error recursions (42)–(45) are divided into three groups: \mathcal{V} , $\Theta = \{\boldsymbol{\Sigma}, \tilde{\mathbf{w}}, \mathbf{x}\}$, and

$$\begin{aligned} \mathcal{X}_{t,t'} &= \left\{ \mathbf{Q}_{t+1}, \mathbf{B}_{t'}, \mathbf{M}_{t'}, \mathbf{H}_t \mid \mathbf{B}_{t'}^H \mathbf{M}_{t'} = \mathbf{Q}_{t'}^H \mathbf{H}_t, \right. \\ &\quad \left. \mathbf{M}_{t'} = \mathbf{G}_{t'}(\mathbf{B}_{t'}, \tilde{\mathbf{w}}), \mathbf{Q}_{t+1} = \mathbf{F}_t(\mathbf{H}_t, \mathbf{x}) \right\}, \end{aligned} \quad (53)$$

TABLE I
NOTATIONAL CONVENTIONS FOR $t = 0$.

$\mathcal{X}_{0,0} = \{\mathbf{Q}_1\}, \mathcal{X}_{0,1} = \{\mathbf{Q}_1, \mathbf{B}_1, \mathbf{M}_1 \mathbf{M}_1 = \mathbf{G}_1(\mathbf{B}_1, \tilde{\mathbf{w}})\},$ $\mathbf{Q}_0 = \mathbf{O}, \mathbf{B}_0 = \mathbf{O}, \mathbf{M}_0 = \mathbf{O}, \mathbf{H}_0 = \mathbf{O}, \mathbf{M}_0^\dagger = \mathbf{O}, \mathbf{Q}_0^\dagger = \mathbf{O},$ $\boldsymbol{\alpha}_0 = \mathbf{0}, \boldsymbol{\beta}_0 = \mathbf{0}, \boldsymbol{\Phi}_0^\perp = \mathbf{I}, \boldsymbol{\Phi}_{(M,\mathbf{O})}^\perp = \boldsymbol{\Phi}_M^\perp \text{ for any } M.$

for $t' = t$ or $t' = t + 1$, while we define $\mathcal{X}_{0,0} = \{\mathbf{Q}_1\}$ and $\mathcal{X}_{0,1} = \{\mathbf{Q}_1, \mathbf{B}_1, \mathbf{M}_1 | \mathbf{M}_1 = \mathbf{G}_1(\mathbf{B}_1, \tilde{\mathbf{w}})\}$. See Table I for the notational conventions used in this paper.

The set Θ is fixed throughout this paper. Thus, conditioning on Θ is omitted. The set $\mathcal{X}_{t,t}$ describes the history of all preceding iterations just before updating (36), while $\mathcal{X}_{t,t+1}$ represents the history just before updating (38). Note that the condition $\mathbf{B}_{t'}^\mathbf{H} \mathbf{M}_t = \mathbf{Q}_{t'}^\mathbf{H} \mathbf{H}_t$ is a constraint imposing $\mathbf{V} \in \mathcal{U}_N$, and follows from (42) and (44). In order to investigate the dynamics of the error recursions, the distribution of the Haar matrix \mathbf{V} conditioned on $\mathcal{X}_{t,t'}$ is analyzed.

Let $\mathbf{m}_t^\perp = \mathbf{P}_{\mathbf{M}_t}^\perp \mathbf{m}_t$. Since $\mathbf{m}_t^\parallel = \mathbf{m}_t - \mathbf{m}_t^\perp$ is in the space spanned by the columns of \mathbf{M}_t , we have $(\mathbf{m}_t^\parallel)^\mathbf{H} \mathbf{m}_t^\perp = 0$. Furthermore, \mathbf{m}_t^\parallel is represented as $\mathbf{m}_t^\parallel = \mathbf{M}_t \boldsymbol{\alpha}_t$, with $\boldsymbol{\alpha}_t = \mathbf{M}_t^\dagger \mathbf{m}_t \in \mathbb{C}^t$. Similarly, we define $\mathbf{q}_{t'}^\perp = \mathbf{P}_{\mathbf{Q}_{t'}}^\perp \mathbf{q}_{t'}$ and $\mathbf{q}_{t'}^\parallel = \mathbf{q}_{t'} - \mathbf{q}_{t'}^\perp = \mathbf{Q}_{t'} \boldsymbol{\beta}_{t'}$, with $\boldsymbol{\beta}_{t'} = \mathbf{Q}_{t'}^\dagger \mathbf{q}_{t'} \in \mathbb{C}^{t'}$.

For notational convenience, we define the conventions listed in Table I, which imply $\mathbf{P}_0^\perp = \mathbf{I}$, $\mathbf{m}_0^\perp = \mathbf{m}_0$, and $\mathbf{q}_0^\perp = \mathbf{q}_0$.

Theorem 4: Let $\mathbf{D} = \text{diag}\{D_1, \dots, D_N\}$ denote any $N \times N$ real diagonal matrix that is a function of $\boldsymbol{\Sigma}$. For some $\epsilon > 0$, suppose that $N^{-1} \text{Tr}(\mathbf{D}^k)$ converges almost surely for all $k \in [0, 4 + \epsilon]$ as $N \rightarrow \infty$. Then, the following properties for module A hold for each iteration $\tau = 0, 1, \dots$

(A1) Define

$$\tilde{\mathbf{b}}_\tau = \mathbf{B}_\tau \boldsymbol{\beta}_\tau + \mathbf{M}_\tau \mathbf{o}(1) + \mathbf{B}_\tau \mathbf{o}(1) + \boldsymbol{\Phi}_{(\mathbf{B}_\tau, \mathbf{M}_\tau)}^\perp \mathbf{z}_\tau, \quad (54)$$

with

$$\mathbf{z}_\tau = \tilde{\mathbf{V}}^\mathbf{H} (\boldsymbol{\Phi}_{(\mathbf{Q}_\tau, \mathbf{H}_\tau)}^\perp)^\mathbf{H} \mathbf{q}_\tau, \quad (55)$$

where $\tilde{\mathbf{V}} \in \mathcal{U}_{N-2t}$ is a Haar matrix and independent of Θ and $\mathcal{X}_{\tau,\tau}$. Then, we have

$$\mathbf{b}_\tau |_{\Theta, \mathcal{X}_{\tau,\tau}} \sim \tilde{\mathbf{b}}_\tau \quad (56)$$

conditioned on Θ and $\mathcal{X}_{\tau,\tau}$ in the large system limit, with

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \left\{ \|\mathbf{z}_\tau\|^2 - \|\mathbf{q}_\tau^\perp\|^2 \right\} \stackrel{\text{a.s.}}{=} 0. \quad (57)$$

(A2) For all $\tau' \leq \tau$,

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{b}_{\tau'}^\mathbf{H} \mathbf{D} \tilde{\mathbf{W}}_\tau \tilde{\mathbf{w}} \stackrel{\text{a.s.}}{=} 0, \quad (58)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \left\{ \mathbf{b}_{\tau'}^\mathbf{H} \mathbf{D} \mathbf{b}_\tau - \frac{\text{Tr}(\mathbf{D})}{N} \mathbf{q}_{\tau'}^\mathbf{H} \mathbf{q}_\tau \right\} \stackrel{\text{a.s.}}{=} 0, \quad (59)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{b}_{\tau'}^\mathbf{H} \mathbf{m}_\tau \stackrel{\text{a.s.}}{=} 0, \quad (60)$$

where $\tilde{\mathbf{W}}_\tau$ is given by (40).

(A3) Define $\bar{v}_{\mathbf{A} \rightarrow \mathbf{B}}^\tau$ in the SE equations (46) and (47), Then,

$$\lim_{M=\delta N \rightarrow \infty} v_{\mathbf{A} \rightarrow \mathbf{B}}^\tau \stackrel{\text{a.s.}}{=} \bar{v}_{\mathbf{A} \rightarrow \mathbf{B}}^\tau, \quad (61)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\mathbf{m}_\tau\|^2 \stackrel{\text{a.s.}}{=} \bar{v}_{\mathbf{A} \rightarrow \mathbf{B}}^\tau. \quad (62)$$

(A4) For some $\epsilon > 0$ and $C > 0$,

$$\lim_{M=\delta N \rightarrow \infty} \mathbb{E} [|\mathbf{m}_{\tau,n}|^{2+\epsilon}] < \infty, \quad (63)$$

$$\limsup_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{m}_\tau^\mathbf{H} \mathbf{D} \mathbf{m}_\tau \stackrel{\text{a.s.}}{<} \infty, \quad (64)$$

$$\liminf_{M=\delta N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} \mathbf{M}_{\tau+1}^\mathbf{H} \mathbf{M}_{\tau+1} \right) \stackrel{\text{a.s.}}{>} C. \quad (65)$$

The following properties hold for module B:

(B1) Define

$$\tilde{\mathbf{h}}_\tau = \mathbf{H}_\tau \boldsymbol{\alpha}_\tau + \mathbf{Q}_{\tau+1} \mathbf{o}(1) + \mathbf{H}_\tau \mathbf{o}(1) + \boldsymbol{\Phi}_{(\mathbf{Q}_{\tau+1}, \mathbf{H}_\tau)}^\perp \tilde{\mathbf{z}}_\tau, \quad (66)$$

with

$$\tilde{\mathbf{z}}_\tau = \tilde{\mathbf{V}} (\boldsymbol{\Phi}_{(\mathbf{B}_{\tau+1}, \mathbf{M}_\tau)}^\perp)^\mathbf{H} \mathbf{m}_\tau, \quad (67)$$

where $\tilde{\mathbf{V}} \in \mathcal{U}_{N-(2t+1)}$ is a Haar matrix and independent of Θ and $\mathcal{X}_{\tau,\tau+1}$. Then, we have

$$\mathbf{h}_\tau |_{\Theta, \mathcal{X}_{\tau,\tau+1}} \sim \tilde{\mathbf{h}}_\tau \quad (68)$$

conditioned on Θ and $\mathcal{X}_{\tau,\tau+1}$ in the large system limit, with

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \left\{ \|\tilde{\mathbf{z}}_\tau\|^2 - \|\mathbf{m}_\tau^\perp\|^2 \right\} \stackrel{\text{a.s.}}{=} 0. \quad (69)$$

(B2) For all $\tau' \leq \tau$,

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbf{h}_{\tau'}^\mathbf{H} \mathbf{q}_{\tau'+1} \stackrel{\text{a.s.}}{=} 0. \quad (70)$$

(B3) Define $\bar{v}_{\mathbf{A} \rightarrow \mathbf{B}}^\tau$ and $\bar{v}_{\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1}$ in the SE equations (46) and (47), Then,

$$\lim_{M=\delta N \rightarrow \infty} v_{\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1} \stackrel{\text{a.s.}}{=} \text{MMSE}(\bar{v}_{\mathbf{A} \rightarrow \mathbf{B}}^\tau), \quad (71)$$

$$\lim_{M=\delta N \rightarrow \infty} v_{\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1} \stackrel{\text{a.s.}}{=} \bar{v}_{\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1}, \quad (72)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\tilde{\eta}_t(\mathbf{x} + \mathbf{h}_t) - \mathbf{x}\|^2 \stackrel{\text{a.s.}}{=} \text{MMSE}(\bar{v}_{\mathbf{A} \rightarrow \mathbf{B}}^\tau), \quad (73)$$

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \|\mathbf{q}_{\tau+1}\|^2 \stackrel{\text{a.s.}}{=} \bar{v}_{\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1}. \quad (74)$$

(B4) For some $\epsilon > 0$ and $C > 0$,

$$\lim_{M=\delta N \rightarrow \infty} \mathbb{E} [|\mathbf{q}_{\tau+1,n}|^{2+\epsilon}] < \infty, \quad (75)$$

$$\liminf_{M=\delta N \rightarrow \infty} \lambda_{\min} \left(\frac{1}{N} \mathbf{Q}_{\tau+2}^\mathbf{H} \mathbf{Q}_{\tau+2} \right) \stackrel{\text{a.s.}}{>} C. \quad (76)$$

(B5) Define $\bar{v}_{\mathbf{A} \rightarrow \mathbf{B}}^\tau$ and $\bar{v}_{\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1}$ in the SE equations (46) and (47), Then,

$$\lim_{M=\delta N \rightarrow \infty} \mathbb{E} [|\tilde{\eta}_\tau(x_n + h_{\tau,n}) - x_n|^2] \stackrel{\text{a.s.}}{=} \text{MMSE}(\bar{v}_{\mathbf{A} \rightarrow \mathbf{B}}^\tau), \quad (77)$$

$$\lim_{M=\delta N \rightarrow \infty} \mathbb{E} [|\mathbf{q}_{\tau+1,n}|^2] \stackrel{\text{a.s.}}{=} \bar{v}_{\mathbf{B} \rightarrow \mathbf{A}}^{\tau+1}. \quad (78)$$

Proof: See Section V. ■

Ma and Ping [27, Assumption 1] postulated that $\tilde{\mathbf{z}}_\tau$ in (66) has independent CSCG elements. The assumption is too strong to be justified. In fact, the references [42], [43] imply that the assumption is not correct, while the assumption holds only

for finite subsets of the elements of $\tilde{\mathbf{z}}_\tau$. However, the weaker property (B1) is sufficient to prove Theorem 4.

Proof of Theorem 2: The property (48) follows from the definition (38) of \mathbf{h}_t and (62). Furthermore, (49) and (50) are due to (73) and (74), respectively. ■

Proof of Theorem 3: Theorem 3 follows from (77) and (78). ■

V. PROOF OF THEOREM 4

A. Technical Lemma

We need to evaluate the two distributions $p(\mathbf{m}_t, \mathbf{b}_t | \Theta, \mathcal{X}_{t,t})$ and $p(\mathbf{q}_{t+1}, \mathbf{h}_t | \Theta, \mathcal{X}_{t,t+1})$. The former distribution represents the error recursions (36) and (37) conditioned on the history of all preceding iterations, while the latter describes the error recursions (38) and (39). We follow the proof strategy in [12] to evaluate the two distributions via the conditional distribution $p(\mathbf{V} | \Theta, \mathcal{X}_{t,t'})$ for $t' = t$ or $t' = t + 1$. See Section I-B for the main idea in analyzing the conditional distributions.

The following lemma provides a useful representation of $\mathbf{V} \in \mathcal{U}_N$ conditioned on Θ and $\mathcal{X}_{t,t'}$, and corresponds to [12, Lemma 10]. See Section I-F for the notation.

Lemma 3: Suppose that $\mathbf{Y} \in \mathbb{C}^{N \times t}$ is full-rank for $0 < t < N$, and consider the noiseless and compressed observation $\mathbf{X} \in \mathbb{C}^{N \times t}$ of \mathbf{V} given by

$$\mathbf{X} = \mathbf{V}^H \mathbf{Y}. \quad (79)$$

Then, the conditional distribution of the Haar matrix \mathbf{V} given \mathbf{X} and \mathbf{Y} satisfies

$$\mathbf{V} |_{\mathbf{X}, \mathbf{Y}} \sim \mathbf{Y}(\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{X}^H + \Phi_{\mathbf{Y}}^\perp \tilde{\mathbf{V}} (\Phi_{\mathbf{X}}^\perp)^H, \quad (80)$$

where $\tilde{\mathbf{V}} \in \mathcal{U}_{N-t}$ is a Haar matrix independent of \mathbf{X} and \mathbf{Y} .

Proof: See Appendix D. ■

Proposition 2: Let $\mathbf{a} \in \mathbb{C}^t$ and $\mathbf{M} = (\mathbf{m}_0, \dots, \mathbf{m}_{t-1}) \in \mathbb{C}^{N \times t}$. If $N^{-1} \|\mathbf{m}_\tau\|^2$ is bounded for any τ as $N \rightarrow \infty$, then

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \mathbf{a}^H \mathbf{M}^H \mathbf{M} \mathbf{a} < \infty. \quad (81)$$

Proof: Proposition 2 follows from $N^{-1} \|\mathbf{m}_\tau\|^2 < \infty$ and $\|\mathbf{M} \mathbf{a}\|^2 \leq \|\mathbf{M}\|^2 \|\mathbf{a}\|^2$. ■

We are ready to prove Theorem 4. The proof is by induction. We first prove the properties of modules A and B for $\tau = 0$. Then, the properties are proved for $\tau = t$ under the induction hypotheses for all $\tau < t$.

B. Module A for $\tau = 0$

Property (A1) for $\tau = 0$: Property (A1) for $\tau = 0$ is trivial from the definition (36) of \mathbf{b}_0 , because of the notational convention. ■

Eq. (58)–(60) for $\tau = 0$: We first prove (58) and (59) for $\tau = 0$. Let $\mathbf{X}_N = \mathbf{b}_0$ and $f_n(z) = z^* [\mathbf{D} \tilde{\mathbf{W}}_0 \tilde{\mathbf{w}}]_n$ to have the representation

$$\frac{1}{N} \mathbf{b}_0^H \mathbf{D} \tilde{\mathbf{W}}_0 \tilde{\mathbf{w}} = \frac{1}{N} \sum_{n=1}^N f_n(b_{0,n}). \quad (82)$$

From Property (A1) for $\tau = 0$, \mathbf{X}_N is unitarily invariant. The definition (36) of \mathbf{b}_0 , $\mathbf{q}_0 = -\mathbf{x}$, and Assumption 1 imply the

condition (12) with $v = 0$. Since f_n is Lipschitz-continuous with the Lipschitz constant $L_n = |[\mathbf{D} \tilde{\mathbf{W}}_0 \tilde{\mathbf{w}}]_n|$, we need to prove the condition (14) to use Lemma 1. Using the definition $\tilde{\mathbf{w}} = \mathbf{U}^H \mathbf{w}$ and Assumption 3 yields

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N L_n^2 &= \frac{1}{N} \tilde{\mathbf{w}}^H \tilde{\mathbf{W}}_0^H \mathbf{D}^2 \tilde{\mathbf{W}}_0 \tilde{\mathbf{w}} \\ &\stackrel{\text{a.s.}}{=} \frac{\sigma^2}{N} \text{Tr} \left(\mathbf{D}^2 \tilde{\mathbf{W}}_0 \tilde{\mathbf{W}}_0^H \right) + o(1) \\ &\leq \sigma^2 \left\{ \frac{\text{Tr}(\mathbf{D}^4)}{N} \right\}^{1/2} \left\{ \frac{\text{Tr}\{(\tilde{\mathbf{W}}_0 \tilde{\mathbf{W}}_0^H)^2\}}{N} \right\}^{1/2} + o(1) \\ &\stackrel{\text{a.s.}}{<} \infty \end{aligned} \quad (83)$$

in the large system limit, where the first inequality follows from the Cauchy-Schwarz inequality, and where the boundedness is due to the definition of \mathbf{D} , the definition (40) of $\tilde{\mathbf{W}}_0$, and Assumption 2. Thus, we can use Lemma 1 to obtain (58) for $\tau = 0$. Similarly, we use Lemma 1 for $f_n(z) = D_n |z|^2$ to have (59) for $\tau = 0$.

We next prove (60) for $\tau = 0$. Let $k \in [0, 4 + \epsilon)$. We use Hölder inequality for any $p \in (1, (4 + \epsilon)/k)$ to obtain

$$\begin{aligned} \frac{1}{N} \text{Tr} \left\{ \left(\mathbf{D} \tilde{\mathbf{W}}_0(\Sigma, \mathbf{O}) \right)^k \right\} \\ \leq \frac{1}{N} \left\{ \text{Tr}(\mathbf{D}^{kp}) \right\}^{1/p} \left\{ \text{Tr} \left[\left(\tilde{\mathbf{W}}_0(\Sigma, \mathbf{O}) \right)^{kq} \right] \right\}^{1/q} \\ \stackrel{\text{a.s.}}{<} \infty \end{aligned} \quad (84)$$

as $N \rightarrow \infty$, with $q^{-1} = 1 - 1/p$, where the boundedness is obtained by repeating the proof of the boundedness in (83). Thus, we use (58) and (59) for $\tau = 0$ to have

$$\begin{aligned} \frac{\gamma_0}{N} \mathbf{b}_0^H \mathbf{D} \tilde{\mathbf{W}}_0 \{(\Sigma, \mathbf{O}) \mathbf{b}_0 - \tilde{\mathbf{w}}\} \\ \stackrel{\text{a.s.}}{=} \gamma_0 \frac{\text{Tr}(\mathbf{D} \tilde{\mathbf{W}}_0(\Sigma, \mathbf{O}))}{N} \frac{1}{N} \mathbf{q}_0^H \mathbf{q}_0 + o(1). \end{aligned} \quad (85)$$

In particular, for $\mathbf{D} = \mathbf{I}_N$ we use the definition (24) of γ_0 to obtain

$$\frac{\gamma_0}{N} \text{Tr} \left\{ \tilde{\mathbf{W}}_0(\Sigma, \mathbf{O}) \right\} \stackrel{\text{a.s.}}{=} 1 + o(1). \quad (86)$$

Applying (86) to (85), we find

$$\frac{\gamma_0}{N} \mathbf{b}_0^H \tilde{\mathbf{W}}_0 \{(\Sigma, \mathbf{O}) \mathbf{b}_0 - \tilde{\mathbf{w}}\} \stackrel{\text{a.s.}}{=} \frac{1}{N} \mathbf{q}_0^H \mathbf{q}_0 + o(1). \quad (87)$$

From the definition (37) of \mathbf{m}_0 , (59) with $\mathbf{D} = \mathbf{I}_N$ for $\tau = 0$, and (87), we arrive at (60) for $\tau = 0$. ■

Eqs. (61)–(65) for $\tau = 0$: The almost sure convergence (61) for $\tau = 0$ follows from the update rule (22) of $v_{A \rightarrow B}^0$, the definition (24) of γ_0 , the SE (46) for module A, and $v_{B \rightarrow A}^0 = \bar{v}_{B \rightarrow A}^0 = 1$.

Let us prove (64) for $\tau = 0$, before proving (62). Using the definition (37) of \mathbf{m}_0 , (85), and Assumption 3, as well as (59) for $\tau = 0$, we have

$$\begin{aligned} \frac{\mathbf{m}_0^H \mathbf{D} \mathbf{m}_0}{N} \\ \stackrel{\text{a.s.}}{=} \frac{\text{Tr}(\mathbf{D})}{N} \frac{\mathbf{q}_0^H \mathbf{q}_0}{N} - 2\gamma_0 \frac{\text{Tr}(\mathbf{D} \tilde{\mathbf{W}}_0(\Sigma, \mathbf{O}))}{N} \frac{\mathbf{q}_0^H \mathbf{q}_0}{N} \\ + \frac{\gamma_0^2 \text{Tr}(\tilde{\mathbf{D}})}{N} \frac{\mathbf{q}_0^H \mathbf{q}_0}{N} + \frac{\sigma^2 \gamma_0^2}{N} \text{Tr} \left(\tilde{\mathbf{W}}_0^H \mathbf{D} \tilde{\mathbf{W}}_0 \right) + o(1) \end{aligned} \quad (88)$$

in the large system limit, with

$$\tilde{\mathbf{D}} = \begin{pmatrix} \Sigma \\ \mathbf{O} \end{pmatrix} \tilde{\mathbf{W}}_0^H \mathbf{D} \tilde{\mathbf{W}}_0(\Sigma, \mathbf{O}). \quad (89)$$

It is straightforward to confirm the boundedness of (88). Thus, (64) holds for $\tau = 0$.

In particular, for $\mathbf{D} = \mathbf{I}_N$ we have

$$\text{Tr}(\tilde{\mathbf{D}}) = \text{Tr} \left\{ (\sigma^2 \mathbf{I}_M + v_{\text{B} \rightarrow \text{A}}^0 \Sigma^2)^{-2} \Sigma^4 \right\}, \quad (90)$$

$$\text{Tr} \left(\tilde{\mathbf{W}}_0^H \tilde{\mathbf{W}}_0 \right) = \text{Tr} \left\{ (\sigma^2 \mathbf{I}_M + v_{\text{B} \rightarrow \text{A}}^0 \Sigma^2)^{-2} \Sigma^2 \right\}. \quad (91)$$

Applying these results, (86), and $N^{-1} \|\mathbf{q}_0\|^2 = N^{-1} \|\mathbf{x}\|^2 \xrightarrow{\text{a.s.}} v_{\text{B} \rightarrow \text{A}}^0 = 1$ to (88), we obtain (62) for $\tau = 0$.

To prove the moment property (63) for $\tau = 0$, we observe that \mathbf{b}_0 has finite $(2 + \epsilon)$ th moments, by using the definition (36) of \mathbf{b}_0 , $\mathbf{q}_0 = -\mathbf{x}$, and Assumption 1. Thus, the moment property (63) for $\tau = 0$ follows from the definition (37) of \mathbf{m}_0 , the definition (40) of $\tilde{\mathbf{W}}_0$, Assumption 2, and Assumption 3.

Finally, (65) for $\tau = 0$ is equivalent to $N^{-1} \|\mathbf{m}_0\|^2 \xrightarrow{\text{a.s.}} \bar{v}_{\text{A} \rightarrow \text{B}}^0 > 0$, which follows from (62) for $\tau = 0$. ■

C. Module B for $\tau = 0$

Property (B1) for $\tau = 0$: We prove (68) for $\tau = 0$. Applying Lemma 3 with $\mathbf{X} = \mathbf{b}_0$ and $\mathbf{Y} = \mathbf{q}_0$ to the definition (38) of \mathbf{h}_0 yields

$$\mathbf{h}_0 \sim \frac{\mathbf{b}_0^H \mathbf{m}_0}{\|\mathbf{q}_0\|^2} \mathbf{q}_0 + \Phi_{\mathbf{q}_0}^\perp \tilde{\mathbf{V}} (\Phi_{\mathbf{b}_0}^\perp)^H \mathbf{m}_0 \quad (92)$$

conditioned on Θ and $\mathcal{X}_{0,1}$, in which $\tilde{\mathbf{V}} \in \mathcal{U}_{N-1}$ is a Haar matrix and independent of Θ and $\mathcal{X}_{0,1}$. From $\mathbf{q}_0 = -\mathbf{x}$, Assumption 1, and (60) for $\tau = 0$, we find

$$\mathbf{h}_0 \sim \mathbf{q}_0 o(1) + \Phi_{\mathbf{q}_0}^\perp \tilde{\mathbf{V}} (\Phi_{\mathbf{b}_0}^\perp)^H \mathbf{m}_0 \quad (93)$$

in the large system limit, which implies (68) for $\tau = 0$, because of the notational convention.

In order to complete the proof, we shall prove (69) for $\tau = 0$. Define

$$\nu_0 = \frac{1}{N} \mathbf{m}_0^H \mathbf{P}_{\mathbf{b}_0}^\perp \mathbf{m}_0. \quad (94)$$

Applying $\mathbf{P}_{\mathbf{b}_0}^\perp = \mathbf{I}_N - \|\mathbf{b}_0\|^{-2} \mathbf{b}_0 \mathbf{b}_0^H$ to (94), and using (59) and (60) for $\tau = 0$, we have

$$\nu_0 \stackrel{\text{a.s.}}{=} \frac{1}{N} \mathbf{m}_0^H \mathbf{m}_0 + o(1) \stackrel{\text{a.s.}}{=} \bar{v}_{\text{A} \rightarrow \text{B}}^0 + o(1) \quad (95)$$

in the large system limit, where the last equality follows from (62) for $\tau = 0$. In particular, we have the convention $\mathbf{m}_0^\perp = \mathbf{m}_0$ to find (69) for $\tau = 0$. Thus, Property (B1) holds for $\tau = 0$. ■

Let $\mathbf{X}_N = \tilde{\mathbf{z}}_0$ given in (67), $\mathbf{a}_{0,N} = \mathbf{x}$, $\epsilon_N = \mathbf{q}_0 o(1)$, and $\mathbf{E}_N = \mathbf{q}_0$. For $k = 1$ or $k = 2$, we prove that all conditions in Lemma 1 with $v = \bar{v}_{\text{A} \rightarrow \text{B}}^0$ are satisfied for any pseudo-Lipschitz function $f_n : \mathbb{C}^2 \rightarrow \mathbb{C}$ of order k with an n -independent Lipschitz constant $L > 0$. Thus,

$$\mathbb{E} [f_n(x_n, h_{0,n})] \stackrel{\text{a.s.}}{=} \mathbb{E} [f_n(x_n, \tilde{z}_{0,n})] + o(1), \quad (96)$$

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N f_n(x_n, h_{0,n}) &\stackrel{\text{a.s.}}{=} \frac{1}{N} \sum_{n=1}^N \mathbb{E}_{\tilde{z}_{0,n}} [f_n(x_n, \tilde{z}_{0,n})] + o(1) \\ &\stackrel{\text{a.s.}}{=} \frac{1}{N} \sum_{n=1}^N \mathbb{E} [f_n(x_n, \tilde{z}_{0,n})] + o(1) \end{aligned} \quad (97)$$

in the large system limit, with $\tilde{\mathbf{z}}_0 \sim \mathcal{CN}(\mathbf{0}, \bar{v}_{\text{A} \rightarrow \text{B}}^0 \mathbf{I}_N)$, where the latter equality follows from Assumption 1.

The conditions (7), (8), (9), and (10) follow from $\mathbf{q}_0 = -\mathbf{x}$, Assumption 1, and Theorem 1. The condition (11) is due to $N^{-1} \|\mathbf{q}_0\|^2 \xrightarrow{\text{a.s.}} 1$. The condition (12) with $v = \bar{v}_{\text{A} \rightarrow \text{B}}^0$ follows from (62) and (69) for $\tau = 0$, as well as the convention $\mathbf{m}_0^\perp = \mathbf{m}_0$. The moment conditions of \mathbf{a}_N , ϵ_N , and \mathbf{E}_N are due to $\mathbf{q}_0 = -\mathbf{x}$ and Assumption 1. The moment condition of \mathbf{X} follows from the definition (67) of $\tilde{\mathbf{z}}_0$ and (63) for $\tau = 0$. Thus, all conditions in Lemma 1 are satisfied.

Eqs. (71) and (72) for $\tau = 0$: We first prove (71) for $\tau = 0$. From the definition (27) of the posterior variance v_{B}^1 , we have

$$v_{\text{B}}^1 = \frac{1}{N} \sum_{n=1}^N f_n(x_n, h_{0,n}), \quad (98)$$

with $f_n(x, z) = \mathbb{V}[x_n | x_{n,\text{A} \rightarrow \text{B}}^0 = x + z]$ defined via the virtual AWGN observation (28). From Assumption 4, the posterior variance $\mathbb{V}[x_n | x_{n,\text{A} \rightarrow \text{B}}^0]$ is bounded, so that $f_n(x, z)$ is a Lipschitz-continuous function with a Lipschitz constant $L > 0$. We use (97) to arrive at

$$v_{\text{B}}^1 \stackrel{\text{a.s.}}{=} \text{MMSE}(\bar{v}_{\text{A} \rightarrow \text{B}}^0) + o(1) \quad (99)$$

in the large system limit, where we have used the fact that the expectation of the posterior variance is equal to the MMSE (35). Thus, (71) holds for $\tau = 0$.

We next prove (72) for $\tau = 0$. From (61) and (71) for $\tau = 0$, we observe that $v_{\text{B} \rightarrow \text{A}}^1$ given in (30) converges almost surely to $\bar{v}_{\text{B} \rightarrow \text{A}}^1$ given in (47) in the large system limit. Thus, (72) holds for $\tau = 0$. ■

Eq. (70) for $\tau = 0$: The Lipschitz-continuity of $\tilde{\eta}_0$ proved in Lemma 2 implies that $f_n(x_n, z) = z^* \tilde{\eta}_0(x_n + z)$ is a pseudo-Lipschitz function of order 2 with an n -independent Lipschitz constant $L > 0$. From (97), we obtain

$$\frac{1}{N} \mathbf{h}_0^H \tilde{\eta}_0(\mathbf{x} + \mathbf{h}_0) \stackrel{\text{a.s.}}{=} \mathbb{E} [\tilde{z}_{0,n}^* \tilde{\eta}_0(x_n + \tilde{z}_{0,n})] + o(1) \quad (100)$$

in the large system limit. Using Lemma 2 yields

$$\frac{1}{N} \mathbf{h}_0^H \tilde{\eta}_0(\mathbf{x} + \mathbf{h}_0) \stackrel{\text{a.s.}}{=} \text{MMSE}(\bar{v}_{\text{A} \rightarrow \text{B}}^0) + o(1) \quad (101)$$

in the large system limit. Similarly, we obtain

$$\frac{1}{N} \mathbf{h}_0^H \mathbf{x} \stackrel{\text{a.s.}}{\rightarrow} 0 \quad (102)$$

in the large system limit.

We use the definition (39) of \mathbf{q}_1 , the definition (31) of η_0 , (72) for $\tau = 0$, and (102) to obtain

$$\begin{aligned} \frac{1}{N} \mathbf{h}_0^H \mathbf{q}_1 &\stackrel{\text{a.s.}}{=} \bar{v}_{\text{B} \rightarrow \text{A}}^{-1} \left(\frac{\mathbf{h}_0^H \tilde{\eta}_0(\mathbf{x} + \mathbf{h}_0)}{N \text{MMSE}(\bar{v}_{\text{A} \rightarrow \text{B}}^0)} - \frac{\|\mathbf{h}_0\|^2}{N \bar{v}_{\text{A} \rightarrow \text{B}}^0} \right) + o(1) \\ &\stackrel{\text{a.s.}}{=} o(1) \end{aligned} \quad (103)$$

in the large system limit, where the last equality follows from the definition (38) of \mathbf{h}_0 , (62) for $\tau = 0$, and (101). Thus, (70) holds for $\tau = 0$. ■

Eqs. (73) and (74) for $\tau = 0$: We first prove (73) for $\tau = 0$. By repeating the proof of (70) for $\tau = 0$, we find

$$\frac{1}{N} \|\tilde{\eta}_0(\mathbf{x} + \mathbf{h}_0) - \mathbf{x}\|^2 \stackrel{\text{a.s.}}{=} \mathbb{E} [|\tilde{\eta}_0(x_n + \tilde{z}_{0,n}) - x_n|^2] + o(1) \quad (104)$$

in the large system limit. Since the variance of $\tilde{z}_{0,n}$ is equal to $\bar{v}_{A \rightarrow B}^0$, from the definition (35) of the MMSE function, we arrive at (73) for $\tau = 0$.

We next prove (74) for $\tau = 0$. Using the definition (39) of \mathbf{q}_1 and the definition (31) of η_0 , and the definition (47) of $\bar{v}_{B \rightarrow A}^1$, as well as (61), (71), and (72) for $\tau = 0$, we have

$$\mathbf{q}_1 = \frac{\bar{v}_{B \rightarrow A}^1 \{\tilde{\eta}_0(\mathbf{x} + \mathbf{h}_0) - \mathbf{x}\}}{\text{MMSE}(\bar{v}_{A \rightarrow B}^t)} - \frac{\bar{v}_{B \rightarrow A}^1}{\bar{v}_{A \rightarrow B}^0} \mathbf{h}_0. \quad (105)$$

Applying (70) and (73) for $\tau = 0$, as well as (101), (102), and $N^{-1} \|\mathbf{h}_0\|^2 \stackrel{\text{a.s.}}{\rightarrow} \bar{v}_{A \rightarrow B}^0$ obtained from the definition (38) of \mathbf{h}_0 and (62) for $\tau = 0$, we have

$$\frac{1}{N} \|\mathbf{q}_1\|^2 \stackrel{\text{a.s.}}{=} \frac{(\bar{v}_{B \rightarrow A}^1)^2}{\text{MMSE}(\bar{v}_{A \rightarrow B}^0)} - \frac{(\bar{v}_{B \rightarrow A}^1)^2}{\bar{v}_{A \rightarrow B}^0} + o(1) \stackrel{\text{a.s.}}{\rightarrow} \bar{v}_{B \rightarrow A}^1 \quad (106)$$

in the large system limit, where the last equality follows from the definition (47) of $\bar{v}_{B \rightarrow A}^1$. Thus, (74) holds for $\tau = 0$. ■

Eq. (75) for $\tau = 0$: From the definition (39) of \mathbf{q}_1 and Proposition 1, we have

$$\mathbb{E}[|\mathbf{q}_{1,n}|^{2+\epsilon}] < C (\mathbb{E}[|\eta_0(x_n + h_{0,n})|^{2+\epsilon}] + \mathbb{E}[|x_n|^{2+\epsilon}]) \quad (107)$$

for some constant $C > 0$. Since Assumption 1 implies the boundedness of the second term, it is sufficient to prove that $\eta_0(x_n + h_{0,n})$ has a finite $(2 + \epsilon)$ th moment for some $\epsilon > 0$.

From Lemma 2 $\tilde{\eta}_0$ is Lipschitz-continuous, so that η_0 given by (31) is so. Thus, we use Proposition 1 to have

$$\mathbb{E}[|\eta_0(x_n + h_{0,n})|^{2+\epsilon}] \leq L (1 + \mathbb{E}[|x_n|^{2+\epsilon}] + \mathbb{E}[|h_{0,n}|^{2+\epsilon}]), \quad (108)$$

for some $L > 0$. The boundedness of $\mathbb{E}[|h_{0,n}|^{2+\epsilon}]$ follows from the definition (38) of \mathbf{h}_0 and (63) for $\tau = 0$. Thus, (75) holds for $\tau = 0$. ■

Eq. (76) for $\tau = 0$: If $\liminf_{M=\delta N \rightarrow \infty} N^{-1} \|\mathbf{q}_1^\perp\|^2$ converges almost surely to a strictly positive constant, (76) holds for $\tau = 0$ [12, Lemmas 8 and 9]. Using (74) for $\tau = 0$, we have

$$\frac{\|\mathbf{q}_1^\perp\|^2}{N} = \frac{\mathbf{q}_1^H \mathbf{P}_{\mathbf{q}_0}^\perp \mathbf{q}_1}{N} \stackrel{\text{a.s.}}{=} \frac{\mathbb{E}[\|\mathbf{q}_1\|^2]}{N} - \left| \frac{\sqrt{N} (\Phi_{\mathbf{q}_0}^\parallel)^H \mathbf{q}_1}{N} \right|^2 + o(1), \quad (109)$$

where \mathbf{q}_1 in the first term is given by $\mathbf{q}_1 = \eta_0(\mathbf{x} + \tilde{\mathbf{z}}_0) - \mathbf{x}$.

Let $f_n(x_n, z) = \sqrt{N} [\Phi_{\mathbf{q}_0}^\parallel]_n^* \{\eta(x_n + z) - x_n\}$. The function f_n is a Lipschitz-continuous function with the Lipschitz constant $L_n = L \sqrt{N} |[\Phi_{\mathbf{q}_0}^\parallel]_n|$ for some $L > 0$. The normalization $\|\Phi_{\mathbf{q}_0}^\parallel\|^2 = 1$ implies $N^{-1} \sum_{n=1}^N L_n^2 = L$, so that we can use (97) to obtain

$$\frac{\sqrt{N} (\Phi_{\mathbf{q}_0}^\parallel)^H \mathbf{q}_1}{N} \stackrel{\text{a.s.}}{=} \frac{\mathbb{E}\{(\Phi_{\mathbf{q}_0}^\parallel)^H \mathbb{E}_{\tilde{\mathbf{z}}_0}[\mathbf{q}_1]\}}{\sqrt{N}} + o(1). \quad (110)$$

Using the Cauchy-Schwarz inequality yields

$$\left| \mathbb{E} \left[(\Phi_{\mathbf{q}_0}^\parallel)^H \mathbb{E}_{\tilde{\mathbf{z}}_0}[\mathbf{q}_1] \right] \right| \leq \mathbb{E} \{ \|\mathbb{E}_{\tilde{\mathbf{z}}_0}[\mathbf{q}_1]\| \} \leq \mathbb{E} [\|\mathbf{q}_1\|], \quad (111)$$

where the latter inequality follows from Jensen's inequality. Thus, we obtain

$$\frac{\|\mathbf{q}_1^\perp\|^2}{N} \stackrel{\text{a.s.}}{\geq} \frac{1}{N} \left\{ \mathbb{E}[\|\mathbf{q}_1\|^2] - (\mathbb{E}[\|\mathbf{q}_1\|])^2 \right\} + o(1) \quad (112)$$

which is strictly positive in the large system limit. Thus, (76) holds for $\tau = 0$. ■

Eqs. (77) and (78) for $\tau = 0$: From (73) and (74) for $\tau = 0$, we may conclude (77) and (78) for $\tau = 0$, since \mathbf{x} and \mathbf{h}_0 have identically distributed elements in the large system limit. Nonetheless, we present a generic proof applicable to the non-identically-distributed case.

We only prove (77) for $\tau = 0$, since (78) can be proved in the same manner. Lemma 2 implies that $|\tilde{\eta}_0(x_n + h_{0,n}) - x_n|^2$ is a pseudo-Lipschitz function of order 2. We use (96) to have

$$\mathbb{E}[|\tilde{\eta}_0(x_n + h_{0,n}) - x_n|^2] \rightarrow \text{MMSE}(\bar{v}_{A \rightarrow B}^0) \quad (113)$$

in the large system limit. ■

We have proved that Theorem 4 holds for $\tau = 0$. Next, we assume that Theorem 4 is correct for all $\tau < t$, and prove that Theorem 4 holds for $\tau = t$.

D. Module A by Induction

Property (A1) for $\tau = t$: We prove (56) for $\tau = t$. Let $\mathbf{Y} = (\mathbf{Q}_t, \mathbf{H}_t)$ and $\mathbf{X} = (\mathbf{B}_t, \mathbf{M}_t)$ in Lemma 3. The induction hypotheses (65) and (76) $\tau < t$ imply that \mathbf{M}_t and \mathbf{Q}_t are full rank. From the definition (44) of \mathbf{H}_t and the induction hypothesis (70) for $\tau < t$, we find that \mathbf{Y} is full rank. Using the definition (36) of \mathbf{b}_t and Lemma 3 yields

$$\mathbf{b}_t \sim (\mathbf{B}_t, \mathbf{M}_t) (\mathbf{Q}_t, \mathbf{H}_t)^\dagger \mathbf{q}_t + \Phi_{(\mathbf{B}_t, \mathbf{M}_t)}^\perp \mathbf{z}_t \quad (114)$$

conditioned on Θ and $\mathcal{X}_{t,t}$, with \mathbf{z}_t defined in (55).

We evaluate the first term on the RHS of (114). Using the induction hypothesis (70) for $\tau < t$ yields

$$\begin{aligned} (\mathbf{Q}_t, \mathbf{H}_t)^\dagger &= \frac{1}{N} \begin{pmatrix} N^{-1} \mathbf{Q}_t^H \mathbf{Q}_t & N^{-1} \mathbf{Q}_t^H \mathbf{H}_t \\ N^{-1} \mathbf{H}_t^H \mathbf{Q}_t & N^{-1} \mathbf{H}_t^H \mathbf{H}_t \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Q}_t^H \\ \mathbf{H}_t^H \end{pmatrix} \\ &\stackrel{\text{a.s.}}{=} \begin{pmatrix} \mathbf{Q}_t^\dagger + o(N^{-1}) \mathbf{H}_t^H \\ \mathbf{H}_t^\dagger + o(N^{-1}) \mathbf{Q}_t^H \end{pmatrix}. \end{aligned} \quad (115)$$

Substituting (115) into the first term, and using the same induction hypothesis again, we obtain

$$(\mathbf{B}_t, \mathbf{M}_t) (\mathbf{Q}_t, \mathbf{H}_t)^\dagger \mathbf{q}_t \stackrel{\text{a.s.}}{=} \mathbf{B}_t \beta_t + \mathbf{B}_t \mathbf{o}(1) + \mathbf{M}_t \mathbf{o}(1), \quad (116)$$

which implies (56) for $\tau = t$.

We next prove (57) for $\tau = t$. Repeating the derivation of (116) with (115) yields

$$\begin{aligned} \frac{1}{N} \|\mathbf{z}_t\|^2 &= \frac{1}{N} \mathbf{q}_t^H \mathbf{P}_{(\mathbf{Q}_t, \mathbf{H}_t)}^\perp \mathbf{q}_t \\ &\stackrel{\text{a.s.}}{=} \frac{\mathbf{q}_t^H}{N} \left\{ \mathbf{P}_{\mathbf{Q}_t}^\perp - \mathbf{P}_{\mathbf{H}_t}^\parallel + o(1) \frac{\mathbf{Q}_t \mathbf{H}_t^H}{N} + o(1) \frac{\mathbf{H}_t \mathbf{Q}_t^H}{N} \right\} \mathbf{q}_t \\ &\stackrel{\text{a.s.}}{=} \frac{1}{N} \|\mathbf{q}_t^\perp\|^2 + o(1), \end{aligned} \quad (117)$$

where the last equality follows from the induction hypothesis (70) for $\tau < t$. Thus, (57) holds for $\tau = t$. ■

Let $\mathbf{X}_N = \mathbf{z}_t$, and $\epsilon_N = \mathbf{B}_t \mathbf{o}(1) + \mathbf{M}_t \mathbf{o}(1)$, $\mathbf{a}_{1,N} = \mathbf{B}_t \boldsymbol{\beta}_t$, and $\mathbf{E}_N = (\mathbf{B}_t, \mathbf{M}_t)$ in Lemma 1. We prove that, for $k = 1$ or $k = 2$, all conditions in Lemma 1 with $v = \mu_t = \lim_{M=\delta N \rightarrow \infty} N^{-1} \|\mathbf{q}_t^\perp\|^2$ are satisfied for any pseudo-Lipschitz function f_n of order k with a Lipschitz-constant L_n , in which $\lim_{N \rightarrow \infty} N^{-1} \sum_{n=1}^N L_n^2 < \infty$ holds for $k = 1$ and in which $\lim_{N \rightarrow \infty} N^{-1} \sum_{n=1}^N L_n^4 < \infty$ holds for $k = 2$, so that

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N f_n(b_{t,n}) \\ \stackrel{\text{a.s.}}{=} & \frac{1}{N} \sum_{n=1}^N \mathbb{E}[f_n([\mathbf{B}_t \boldsymbol{\beta}_t]_n + z_{t,n}) | \Theta, \mathcal{X}_{t,t}] + o(1) \end{aligned} \quad (118)$$

conditioned on Θ and $\mathcal{X}_{t,t}$ in the large system limit, with $\mathbf{z}_t \sim \mathcal{CN}(\mathbf{0}, \mu_t \mathbf{I}_N)$.

The conditions (7) and (10) follow from Proposition 2, the induction hypotheses (59), (64), and (74) for $\tau < t$. For $k = 1$, the conditions (8) and (9) are trivial. For $k = 2$, the condition (8) is due to (7). The condition (9) follows from Proposition 2, the induction hypotheses (59), (74) for $\tau < t$, as well as from the boundedness of $\|\boldsymbol{\beta}_t\|$,

$$\begin{aligned} \|\boldsymbol{\beta}_t\|^2 &= \frac{\mathbf{q}_t^H \mathbf{Q}_t}{N} \left(\frac{1}{N} \mathbf{Q}_t^H \mathbf{Q}_t \right)^{-2} \frac{\mathbf{Q}_t^H \mathbf{q}_t}{N} \\ &\stackrel{\text{a.s.}}{<} C \left\| \frac{1}{N} \mathbf{Q}_t^H \mathbf{q}_t \right\|^2 \leq \frac{C}{N} \sum_{\tau=0}^{t-1} \|\mathbf{q}_\tau\|^2 \frac{\|\mathbf{q}_t\|^2}{N} \stackrel{\text{a.s.}}{<} \infty \end{aligned} \quad (119)$$

for some constant $C > 0$, where the first two inequalities follow from the induction hypothesis (76) for $\tau = t - 2$ and from the Cauchy-Schwarz inequality, respectively, and where the boundedness is due to the induction hypothesis (74) for $\tau < t$.

The condition (11) follows from the induction hypotheses (60), (65), and (76) for $\tau < t$, as well as the definition (42) of \mathbf{B}_t . The condition (12) with $v = \mu_t$ follows from (57) for $\tau = t$. We have proved all assumptions in Lemma 1. Thus, (118) holds.

Eqs. (58)–(60) for $\tau = t$: We first prove (58) for $\tau = t$. Define the Lipschitz-continuous function $f_n(z) = z^* [\mathbf{D} \tilde{\mathbf{W}}_t \tilde{\mathbf{w}}]_n$. We note that $\tilde{\mathbf{W}}_t$ given in (40) is independent of $v_{B \rightarrow A}^t$ in the large system limit, because of the induction hypothesis (72) for $\tau = t - 1$. Repeating the proof of (83), we find that $N^{-1} \|\mathbf{D} \tilde{\mathbf{W}}_t \tilde{\mathbf{w}}\|^2$ is almost surely bounded as $N \rightarrow \infty$. Thus, we can use (118) to obtain

$$\frac{1}{N} \mathbf{b}_\tau^H \mathbf{D} \tilde{\mathbf{W}}_t \tilde{\mathbf{w}} \stackrel{\text{a.s.}}{=} \frac{1}{N} \boldsymbol{\beta}_\tau^H \mathbf{B}_\tau^H \mathbf{D} \tilde{\mathbf{W}}_t \tilde{\mathbf{w}} + o(1) \xrightarrow{\text{a.s.}} 0, \quad (120)$$

where the last convergence follows from the induction hypothesis (58) for all $\tau < t$. Thus, (58) holds for $\tau = t$.

We next prove (59) for $\tau = t$. From (118) for $f_n(z) = D_n |z|^2$, we have

$$\frac{1}{N} \mathbf{b}_\tau^H \mathbf{D} \mathbf{b}_t \stackrel{\text{a.s.}}{=} \frac{1}{N} \mathbb{E} \left[\mathbf{b}_\tau^H \mathbf{D} (\mathbf{B}_t \boldsymbol{\beta}_t + \mathbf{z}_t) \middle| \Theta, \mathcal{X}_{t,t} \right] + o(1) \quad (121)$$

in the large system limit for all $\tau \leq t$, where \mathbf{b}_τ is replaced by $\mathbf{B}_t \boldsymbol{\beta}_t + \mathbf{z}_t$ for $\tau = t$. For $\tau < t$, we have

$$\frac{1}{N} \mathbf{b}_\tau^H \mathbf{D} \mathbf{b}_t \stackrel{\text{a.s.}}{=} \frac{1}{N} \mathbf{b}_\tau^H \mathbf{D} \mathbf{B}_t \boldsymbol{\beta}_t + o(1). \quad (122)$$

Using the induction hypothesis (59) for $\tau < t$, $\mathbf{q}_t^\parallel = \mathbf{Q}_t \boldsymbol{\beta}_t$, and $\mathbf{q}_{\tau'}^H \mathbf{q}_t^\perp = 0$ yields (59) for $\tau = t$ and $\tau' < t$.

For $\tau = t$, we obtain

$$\frac{1}{N} \mathbf{b}_t^H \mathbf{D} \mathbf{b}_t \stackrel{\text{a.s.}}{=} \frac{1}{N} \boldsymbol{\beta}_t^H \mathbf{B}_t^H \mathbf{D} \mathbf{B}_t \boldsymbol{\beta}_t + \frac{\mu_t}{N} \text{Tr}(\mathbf{D}) + o(1) \quad (123)$$

in the large system limit. The induction hypothesis (59) for $\tau < t$ implies that the first term converges almost surely to $\lim_{M=\delta N \rightarrow \infty} N^{-1} \|\mathbf{q}_t^\parallel\|^2 N^{-1} \text{Tr}(\mathbf{D})$. Thus, (59) holds for $\tau = \tau' = t$.

Finally, we prove (60) for $\tau = t$. Repeating the proof of (84) yields the boundedness of $N^{-1} \text{Tr}\{(\mathbf{D} \tilde{\mathbf{W}}_t(\boldsymbol{\Sigma}, \mathbf{O}))^k\}$ for $k \in [0, 4 + \epsilon)$. Thus, we can use (58) and (59) to find

$$\begin{aligned} & \frac{\gamma_t}{N} \mathbf{b}_\tau^H \mathbf{D} \tilde{\mathbf{W}}_t \{(\boldsymbol{\Sigma}, \mathbf{O}) \mathbf{b}_t - \tilde{\mathbf{w}}\} \\ & \stackrel{\text{a.s.}}{=} \gamma_t \frac{\text{Tr}\{\mathbf{D} \tilde{\mathbf{W}}_t(\boldsymbol{\Sigma}, \mathbf{O})\}}{N} \frac{1}{N} \mathbf{q}_\tau^H \mathbf{q}_t + o(1). \end{aligned} \quad (124)$$

In particular, for $\mathbf{D} = \mathbf{I}_N$ we find that the almost sure convergence (60) for $\tau = t$ follows from the definition (24) of γ_t , the definition (40) of $\tilde{\mathbf{W}}_t$, and Assumption 2, as well as the boundedness of $N^{-1} \mathbf{q}_\tau^H \mathbf{q}_t$, obtained from the Cauchy-Schwarz inequality and the induction hypothesis (74) for $\tau < t$. ■

Eqs. (61)–(65) for $\tau = t$: The almost sure convergence (61) for $\tau = t$ follows from the definition (22) of $v_{A \rightarrow B}^t$, the definition (24) of γ_t , the definition (46) of $v_{A \rightarrow B}^t$, and the induction hypothesis (72) for $\tau = t - 1$.

The properties (62) and (64) for $\tau = t$ are obtained by repeating the proofs of (62) and (64) for $\tau = 0$. The moment property (63) for $\tau = t$ follows from the definition (37) of \mathbf{m}_t , the definition (40) of $\tilde{\mathbf{W}}_t$, Assumption 2, and Assumption 3, since we have already proved the boundedness of the $(2 + \epsilon)$ th moments of \mathbf{b}_t .

Finally, we prove (65) for $\tau = t$. The induction hypothesis (65) for $\tau < t$ implies that (65) holds for $\tau = t$ if $\liminf_{M=\delta N \rightarrow \infty} N^{-1} \|\mathbf{m}_t^\perp\|^2$ converges almost surely to a strictly positive constant. We use the definition (37) of \mathbf{m}_t , (62), and (118) for $f_n(z) = [\sqrt{N}(\boldsymbol{\Phi}_{M_t}^\parallel)^\text{H} \{ \mathbf{I}_N - \gamma_t \tilde{\mathbf{W}}_t(\boldsymbol{\Sigma}, \mathbf{O}) \}]_{\tau,n} z$ for $\tau < t$ to obtain

$$\frac{\|\mathbf{m}_t^\perp\|^2}{N} \stackrel{\text{a.s.}}{=} \frac{\mathbb{E}_{\mathbf{z}_t}[\|\mathbf{m}_t\|^2]}{N} - \left\| \mathbb{E}_{\mathbf{z}_t} \left[(\boldsymbol{\Phi}_{M_t}^\parallel)^\text{H} \frac{\mathbf{m}_t}{\sqrt{N}} \right] \right\|^2 + o(1). \quad (125)$$

By repeating the proof of (76) for $\tau = 0$, we arrive at

$$\frac{\|\mathbf{m}_t^\perp\|^2}{N} \stackrel{\text{a.s.}}{\geq} \frac{1}{N} (\mathbb{E}_{\mathbf{z}_t}[\|\mathbf{m}_t\|^2] - \|\mathbb{E}_{\mathbf{z}_t}[\mathbf{m}_t]\|^2) + o(1), \quad (126)$$

which is strictly positive in the large system limit. Thus, (65) holds for $\tau = t$. ■

E. Module B by Induction

Property (B1) for $\tau = t$: Let us prove (68) for $\tau = t$. Using (38) and Lemma 3 with $\mathbf{Y} = (\mathbf{Q}_{t+1}, \mathbf{H}_t)$ and $\mathbf{X} = (\mathbf{B}_{t+1}, \mathbf{M}_t)$ yields

$$\mathbf{h}_t \sim (\mathbf{Q}_{t+1}, \mathbf{H}_t) (\mathbf{B}_{t+1}, \mathbf{M}_t)^\dagger \mathbf{m}_t + \boldsymbol{\Phi}_{(\mathbf{Q}_{t+1}, \mathbf{H}_t)}^\perp \tilde{\mathbf{z}}_t \quad (127)$$

conditioned on Θ and $\mathcal{X}_{t,t+1}$, with $\tilde{\mathbf{z}}_t$ given in (67), where we have used the identity $\mathbf{X}^H \mathbf{X} = \mathbf{Y}^H \mathbf{Y}$. Repeating the proof of Property (A1) for $\tau = t$, we arrive at Property (B1) for $\tau = t$. ■

Let $\mathbf{X}_N = \tilde{\mathbf{z}}_t$ given in (67), $\mathbf{a}_{0,N} = \mathbf{x}$, $\mathbf{a}_{\tau+1,N} = \mathbf{h}_\tau$ for $\tau < t$, $\mathbf{a}_{t+1,N} = \mathbf{H}_t \boldsymbol{\alpha}_t$, $\boldsymbol{\epsilon}_N = \mathbf{Q}_{t+1} \mathbf{o}(1) + \mathbf{H}_t \mathbf{o}(1)$, and $\mathbf{E}_N = (\mathbf{Q}_{t+1}, \mathbf{H}_t)$. For $k = 1$ or $k = 2$, let $f_n : \mathbb{C}^{t+2} \rightarrow \mathbb{C}$ denote a pseudo-Lipschitz function of order k with an n -independent Lipschitz constant $L > 0$. We shall prove

$$\begin{aligned} & \mathbb{E}[f_n(x_n, h_{0,n}, \dots, h_{t,n})] \\ \stackrel{\text{a.s.}}{=} & \mathbb{E}[g_n(x_n, h_{0,n}, \dots, h_{t-1,n})] + o(1), \end{aligned} \quad (128)$$

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N f_n(x_n, h_{0,n}, \dots, h_{t,n}) \\ \stackrel{\text{a.s.}}{=} & \frac{1}{N} \sum_{n=1}^N g_n(x_n, h_{0,n}, \dots, h_{t-1,n}) + o(1) \end{aligned} \quad (129)$$

conditioned on Θ and $\mathcal{X}_{t,t+1}$ in the large system limit, with

$$\begin{aligned} & g_n(x_n, h_{0,n}, \dots, h_{t-1,n}) \\ = & \mathbb{E}_{\tilde{\mathbf{z}}_{t,n}} [f_n(x_n, h_{0,n}, \dots, h_{t-1,n}, [\mathbf{H}_t \boldsymbol{\alpha}_t]_n + \tilde{z}_{t,n})], \end{aligned} \quad (130)$$

where $\tilde{\mathbf{z}}_t \sim \mathcal{CN}(\mathbf{0}, \nu_t \mathbf{I}_N)$ is a CSCG vector with $\nu_t = \lim_{M=\delta N \rightarrow \infty} N^{-1} \|\mathbf{m}_t^\perp\|^2$.

It is sufficient to confirm that all conditions in Lemma 1 hold. The conditions (7) and (10) follow from the definition (38) of \mathbf{h}_t , the induction hypotheses (62), (74) for $\tau < t$, and Proposition 2. The conditions (8) and (9) are trivial for $k = 1$. For $k = 2$, the condition (8) is due to (7). The condition (9) follows from Assumption 1 when $\tau = 0$, the definition (38) of \mathbf{h}_τ and the induction hypothesis (62) when $\tau = 1, \dots, t-1$, and from the boundedness of $\|\boldsymbol{\alpha}_t\|^2$ when $\tau = t$, obtained by repeating the proof of (119) with the induction hypotheses (62) and (65) for $\tau < t$. The condition (11) is due to the induction hypotheses (65), (70), and (76) for $\tau < t-1$, as well as the definition (44) of \mathbf{H}_t . The condition (12) follows from (69) for $\tau = t$.

The moment conditions of $\boldsymbol{\epsilon}_N$, $\mathbf{a}_{\tau,N}$, and \mathbf{E}_N follow from Assumption 1, the induction hypothesis (75) for $\tau < t$, and the boundedness of the $(2 + \epsilon)$ th moments of \mathbf{h}_τ for $\tau < t$, of which the last is due to the definition (38) of \mathbf{h}_τ and the induction hypothesis (63) for $\tau < t$. The moment condition of \mathbf{X}_N is due to (63) for $\tau = t$ and the definition (67) of $\tilde{\mathbf{z}}_t$. Thus, all conditions in Lemma 1 hold.

Define \mathbf{h}_τ^G recursively as

$$\mathbf{h}_\tau^G = \mathbf{H}_\tau^G \boldsymbol{\alpha}_\tau + \tilde{\mathbf{z}}_\tau, \quad (131)$$

with $\mathbf{H}_\tau^G = (\mathbf{h}_0^G, \dots, \mathbf{h}_{\tau-1}^G)$, where $\{\tilde{\mathbf{z}}_\tau \sim \mathcal{CN}(\mathbf{0}, \nu_t \mathbf{I}_N)\}$ are independent CSCG vectors. By definition, \mathbf{h}_τ^G conditioned on $\{\boldsymbol{\alpha}_\tau\}$ and $\{\nu_\tau\}$ is a CSCG vector. Comparing the definition (66) of \mathbf{h}_τ and the definition (131) of \mathbf{h}_τ^G , from the definition (38) of \mathbf{h}_t and (62) for $\tau = t$ we find $N^{-1} \mathbb{E}[\|\mathbf{h}_t^G\|^2] \rightarrow \bar{v}_{A \rightarrow B}^t$ in the large system limit.

It is straightforward to confirm that the function (130) is pseudo-Lipschitz of order k with an n -independent Lipschitz

constant. Thus, we can repeat the argument in (128) and (129) to arrive at

$$\begin{aligned} & \mathbb{E}[f_n(x_n, h_{0,n}, \dots, h_{t,n})] \\ \stackrel{\text{a.s.}}{=} & \mathbb{E}[f_n(x_n, h_{0,n}^G, \dots, h_{t-1,n}^G)] + o(1), \end{aligned} \quad (132)$$

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N f_n(x_n, h_{0,n}, \dots, h_{t,n}) \\ \stackrel{\text{a.s.}}{=} & \frac{1}{N} \sum_{n=1}^N \mathbb{E}[f_n(x_n, h_{0,n}^G, \dots, h_{t-1,n}^G)] + o(1). \end{aligned} \quad (133)$$

Eqs. (71) and (72) for $\tau = t$: Repeating the proofs of (71) and (72) for $\tau = 0$ with (133), we arrive at (71) and (72) for $\tau = t$. ■

Eq. (70) for $\tau = t$: For $\tau < t$, we use the definition (36) of \mathbf{b}_t and the definition (38) of \mathbf{h}_t to obtain

$$\frac{1}{N} \mathbf{h}_t^H \mathbf{q}_{\tau+1} = \frac{1}{N} \mathbf{m}_t^H \mathbf{b}_{\tau+1} \xrightarrow{\text{a.s.}} o(1) \quad (134)$$

in the large system limit, where the last convergence follows from (60) for $\tau = t$ and $\tau' = \tau + 1 \leq t$.

For $\tau = t$, we use (133) for the pseudo-Lipschitz function $f_n(x_n, h_{t,n}) = h_{t,n}^* \{\eta_t(x_n + h_{t,n}) - x_n\}$ of order 2 to have

$$\frac{1}{N} \mathbf{h}_t^H \mathbf{q}_{t+1} \stackrel{\text{a.s.}}{=} \frac{1}{N} \mathbb{E}[(\mathbf{h}_t^G)^H \eta_t(\mathbf{x} + \mathbf{h}_t^G)] + o(1) \quad (135)$$

in the large system limit. Since \mathbf{h}_t^G has independent CSCG elements with variance $\bar{v}_{A \rightarrow B}^t$, we repeat the proof of (70) for $\tau = 0$ to obtain (70) for $\tau = \tau' = t$. ■

Eqs. (73)–(75) for $\tau = t$: Repeat the proofs of (73), (74), and (75) for $\tau = 0$ with (133). ■

Eq. (76) for $\tau = t$: Repeat the proof of (65) for $\tau = t$ with (129). ■

Eqs. (77)–(78) for $\tau = t$: Repeat the proofs of (77)–(78) for $\tau = 0$ with (132). ■

APPENDIX A PROOF OF LEMMA 1

A. Technical Results

Consider $t = 0$. Since $\mathbf{X}_N \in \mathbb{C}^N$ is unitarily invariant, we use the SVD of \mathbf{X}_N to obtain $\mathbf{X}_N = \Phi_{\mathbf{X}_N}^\parallel \|\mathbf{X}_N\|$, in which $\Phi_{\mathbf{X}_N}^\parallel \in \mathcal{U}_{N \times 1}$ is Haar-distributed and independent of the singular value $\|\mathbf{X}_N\|$ [35]. Furthermore, $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is unitarily invariant, so that its SVD is given by $\mathbf{u} = \Phi_{\mathbf{u}}^\parallel \|\mathbf{u}\|$, in which $\Phi_{\mathbf{u}}^\parallel \in \mathcal{U}_{N \times 1}$ is Haar-distributed and independent of $\|\mathbf{u}\|$. Since $\Phi_{\mathbf{X}_N}^\parallel \sim \Phi_{\mathbf{u}}^\parallel$ holds, we have the following representation:

$$\boldsymbol{\epsilon}_N + \mathbf{X}_N \sim \boldsymbol{\epsilon}_N + \frac{\|\mathbf{X}_N\|}{\|\mathbf{u}\|} \mathbf{u}. \quad (136)$$

Let $\mathcal{N} = \{1, \dots, t\}$ for $t > 0$. We repeat the same argument to obtain

$$\begin{aligned} \boldsymbol{\epsilon}_N + \Phi_{\mathbf{E}_N}^\perp \mathbf{X}_{N-t} & \sim \boldsymbol{\epsilon}_N + \frac{\|\mathbf{X}_{N-t}\|}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} \Phi_{\mathbf{E}_N}^\perp \mathbf{u}_{\setminus \mathcal{N}} \\ & = \tilde{\boldsymbol{\epsilon}}_N + \frac{\|\mathbf{X}_{N-t}\|}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} \mathbf{z}, \end{aligned} \quad (137)$$

with $\mathbf{z} = \Phi_{\mathbf{E}_N} \mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$, $\tilde{\epsilon}_N = \epsilon_N - \mathbf{E}_N \delta_N$, and

$$\delta_N = \frac{\|\mathbf{X}_{N-t}\|}{\|\mathbf{u}_{\setminus N}\|} (\mathbf{E}_N^H \mathbf{E}_N)^{-1} \mathbf{E}_N^H \Phi_{\mathbf{E}_N} \mathbf{u}_N. \quad (138)$$

Introducing the convention $\mathbf{u}_{\setminus N} = \mathbf{u}$, $\tilde{\epsilon}_N = \epsilon_N$, and $\Phi_{\mathbf{E}_N} = \mathbf{I}_N$ for $t = 0$, we arrive at the unified representation (137) for $t \geq 0$.

We first prove that $\|\delta_N\|^2$ given in (138) converges almost surely to zero as $N \rightarrow \infty$. From the assumption (12) and $(N-t)^{-1} \|\mathbf{u}_{\setminus N}\|^2 \xrightarrow{\text{a.s.}} 1$, we have

$$\|\delta_N\|^2 \stackrel{\text{a.s.}}{=} v \mathbf{u}_N^H (\Phi_{\mathbf{E}_N}^H)^H \mathbf{E}_N (\mathbf{E}_N^H \mathbf{E}_N)^{-2} \mathbf{E}_N^H \Phi_{\mathbf{E}_N} \mathbf{u}_N + o(1). \quad (139)$$

Using $\mathbf{E}_N = \Phi_{\mathbf{E}_N} \Sigma_{\mathbf{E}_N} \Psi_{\mathbf{E}_N}^H$ and the assumption (11) yields

$$\|\delta_N\|^2 \stackrel{\text{a.s.}}{=} \frac{v}{N} \mathbf{u}_N^H \left(\frac{1}{N} \Sigma_{\mathbf{E}_N}^2 \right)^{-1} \mathbf{u}_N + o(1) \stackrel{\text{a.s.}}{<} \frac{C}{N} \|\mathbf{u}_N\|^2 \quad (140)$$

for some constant $C > 0$. For any $a > 0$, we utilize Chebyshev's inequality to obtain

$$\sum_{N=t+1}^{\infty} \Pr \left(\frac{\|\mathbf{u}_N\|^2}{N} > a \right) \leq \frac{\mathbb{E}[\|\mathbf{u}_N\|^4]}{a^2} \sum_{N=t+1}^{\infty} \frac{1}{N^2} < \infty. \quad (141)$$

Thus, the Borel-Cantelli lemma implies that $\|\delta_N\|^2$ converges almost surely to zero as $N \rightarrow \infty$.

Before proving Lemma 1, we prove several technical results.

Proposition 3: Let $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. For any $k \geq 0$,

$$\mathbb{E} [\|\mathbf{z}\|^{-k}] \leq \frac{1 + o(1)}{N^{k/2}} \quad \text{as } N \rightarrow \infty. \quad (142)$$

Proof: By definition, $2\|\mathbf{z}\|^2$ follows the chi-square distribution with $2N$ degrees of freedom. Let $\Gamma(x)$ denote the gamma function. For $N > k/2$, we use the probability density function of the chi-square distribution to have

$$\begin{aligned} \mathbb{E} \left[\frac{1}{\|\mathbf{z}\|^k} \right] &= 2^{k/2} \int_0^{\infty} \frac{1}{x^{k/2}} \frac{x^{N-1} e^{-x/2}}{2^N \Gamma(N)} dx \\ &= \int_0^{\infty} \frac{x^{N-k/2-1} e^{-x}}{\Gamma(N)} dx = \frac{\Gamma(N-k/2)}{\Gamma(N)}, \end{aligned} \quad (143)$$

where the last equality follows from the definition of the gamma function. Using $\Gamma(x+1) = x\Gamma(x)$ and Gautschi's inequality $\Gamma(x+s)/\Gamma(x) \leq x^s$ for all $x > 0$ and $s \in [0, 1]$, we have

$$\begin{aligned} \mathbb{E} \left[\frac{1}{\|\mathbf{z}\|^k} \right] &= \frac{\Gamma(N-k/2)}{(N-1) \cdots (N-\lceil k/2 \rceil) \Gamma(N-\lceil k/2 \rceil)} \\ &\leq \frac{1}{N^{k/2}} \frac{N^{\lceil k/2 \rceil}}{\prod_{i=1}^{\lceil k/2 \rceil} (N-i)} \end{aligned} \quad (144)$$

for $N > \lceil k/2 \rceil$. Since the latter factor tends to 1 as $N \rightarrow \infty$, Proposition 3 holds. ■

Proposition 4: Let $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. For any $k \geq 0$,

$$\mathbb{E} [|N - \|\mathbf{z}\|^2|^k] = \mathcal{O}(N^{k/2}) \quad \text{as } N \rightarrow \infty. \quad (145)$$

Proof: Let $Z_N = N^{-1/2} \sum_{n=1}^N (|z_n|^2 - 1)$. By definition, we have

$$\frac{1}{N^{k/2}} \mathbb{E} [|N - \|\mathbf{z}\|^2|^k] = \mathbb{E} [|Z_N|^k]. \quad (146)$$

The central limit theorem implies that Z_N converges in distribution to a zero-mean Gaussian random variable Z as $N \rightarrow \infty$. Furthermore, the sequence $\{|Z_N|^k\}$ is uniformly integrable [44] since the $(k+1)$ th moment of Z_N is bounded. Thus, we arrive at

$$\lim_{N \rightarrow \infty} \frac{1}{N^{k/2}} \mathbb{E} [|N - \|\mathbf{z}\|^2|^k] = \mathbb{E} [|Z|^k] < \infty, \quad (147)$$

which implies Proposition 4. ■

Proposition 5: Let $v_N = \|\mathbf{X}_N\|^2/N$, and postulate (12) and the moment assumption on \mathbf{X}_N in Lemma 1. For some any $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} \mathbb{E} [|\sqrt{v_N} - \sqrt{v}|^\rho] = 0 \quad (148)$$

for any $\rho \in [0, \max\{2, 2k-2\} + \epsilon)$.

Proof: From (12), $\sqrt{v_N}$ converges almost surely to \sqrt{v} as $N \rightarrow \infty$. Furthermore, $(N^{-1} \|\mathbf{X}_N\|)^\rho$ is uniformly integrable for all $\rho \in [0, \max\{2, 2k-2\} + \epsilon']$ with any $\epsilon' \in (0, \epsilon)$, because of the moment assumption on \mathbf{X}_N . Thus, Proposition 5 holds. ■

Note that Proposition 5 implies the convergence of the ρ th moment

$$\lim_{N \rightarrow \infty} \mathbb{E} [v_N^{\rho/2}] = v^{\rho/2}. \quad (149)$$

B. Discussion

From the almost sure convergence $\|\delta_N\|^2 \xrightarrow{\text{a.s.}} 0$, as well as $\|\mathbf{X}_N\|^2/\|\mathbf{u}\|^2 \xrightarrow{\text{a.s.}} v$, Rangan *et al.* [36, Proof of Lemma 5] concluded Lemma 1. However, what they have proved should be regarded as not the almost sure convergence but as the convergence in probability.

For simplicity, we assume $t = 0$, $f_n(z) = z$, and $\epsilon_N = \mathbf{0}$. Furthermore, let $S_N = N^{-1} \sum_{n=1}^N X_{n,N}$ and $\tilde{S}_N = (\|\mathbf{X}_N\|/\|\mathbf{u}\|) N^{-1} \sum_{n=1}^N u_n$. From (136), for any $\epsilon > 0$ and $\epsilon' > 0$ we have

$$\begin{aligned} \Pr(|S_N| > \epsilon) &= \Pr(\mathcal{E}_{N,\epsilon'}) \Pr(|\tilde{S}_N| > \epsilon | \mathcal{E}_{N,\epsilon'}) \\ &\quad + \Pr(\mathcal{E}_{N,\epsilon'}^c) \Pr(|\tilde{S}_N| > \epsilon | \mathcal{E}_{N,\epsilon'}^c), \end{aligned} \quad (150)$$

with

$$\mathcal{E}_{N,\epsilon'} = \left\{ \left| \frac{\|\mathbf{X}_N\|^2}{\|\mathbf{u}\|^2} - v \right| \leq \epsilon' \right\}. \quad (151)$$

The almost sure convergence $\|\mathbf{X}_N\|^2/\|\mathbf{u}\|^2 \xrightarrow{\text{a.s.}} v$ implies that the second term tends to zero as $N \rightarrow \infty$. Using Chebyshev's inequality for the first term yields

$$\Pr(\mathcal{E}_{N,\epsilon'}) \Pr(|\tilde{S}_N| > \epsilon | \mathcal{E}_{N,\epsilon'}) < \frac{\epsilon' + v}{N\epsilon^2} \rightarrow 0. \quad (152)$$

Thus, we arrive at the convergence in probability $\Pr(|S_N| > \epsilon) \rightarrow 0$ as $N \rightarrow \infty$.

However, it is not straightforward to prove the almost sure convergence. To construct a simple counterexample, suppose that $p_{N,\epsilon} = \Pr(|\tilde{S}_N| > \epsilon)$ is $\mathcal{O}(N^{-1})$. Then, we find

$$\sum_{N=1}^{\infty} \Pr(|S_N| > \epsilon) = \sum_{N=1}^{\infty} p_{N,\epsilon} = \infty. \quad (153)$$

While we do not introduce any statistical properties of $\{\mathbf{X}_N\}$ with respect to N , we assume the independence of $\{S_N\}$

to construct a counterexample. Then, from the second Borel-Cantelli lemma we can conclude that S_N does not converge almost surely to zero. This counterexample implies that we need information about the convergence speed of $p_{N,\epsilon}$ to establish the almost sure convergence in (15). Instead of evaluating the actual convergence speed of $p_{N,\epsilon}$, we use Theorem 1 to prove the almost sure convergence directly.

C. Proof of (13)

Since ϵ_N has vanishing second moments and finite $(2k - 2)$ th moments and since \mathbf{E}_N has finite $\max\{2, 2k - 2\}$ th moments, the almost sure convergence $\|\delta_N\|^2 \xrightarrow{\text{a.s.}} 0$ implies that $\tilde{\epsilon}_N = \epsilon_N - \mathbf{E}_N \mathbf{o}(1)$ has vanishing second moments and finite $(2k - 2)$ th moments. Furthermore, we only prove the case $t' = 1$ with $a_{n,1,N} = 0$ since an extension of the proof to the general case is straightforward. For notational simplicity, we write \mathbf{X}_{N-t} and $a_{n,0,N}$ as \mathbf{X} and $a_{n,N}$.

Let

$$Y_{n,N}^1 = f_n \left(a_{n,N}, \epsilon_{n,N} + [\Phi_{\mathbf{E}_N}^\perp \mathbf{X}]_n \right) - f_n \left(a_{n,N}, \tilde{\epsilon}_{n,N} + v_{N-t}^{1/2} z_n \right), \quad (154)$$

$$Y_{n,N}^2 = f_n \left(a_{n,N}, \tilde{\epsilon}_{n,N} + v_{N-t}^{1/2} z_n \right) - f_n \left(a_{n,N}, v_{N-t}^{1/2} z_n \right), \quad (155)$$

$$Y_{n,N}^3 = f_n \left(a_{n,N}, v_{N-t}^{1/2} z_n \right) - f_n \left(a_{n,N}, v_{N-t}^{1/2} z_n \right), \quad (156)$$

with $v_{N-t} = \|\mathbf{X}_{N-t}\|^2/N$. It is sufficient to prove

$$\mathbb{E} [|Y_{n,N}^1|] = \mathcal{O} \left(\frac{L_n}{\sqrt{N}} \right), \quad (157)$$

$$\mathbb{E} [|Y_{n,N}^2|] = o(L_n), \quad (158)$$

$$\mathbb{E} [|Y_{n,N}^3|] = o(L_n) \quad (159)$$

as $N \rightarrow \infty$.

Let $\mathcal{E} = \{\|\mathbf{X}\|, a_N, \epsilon_N, \mathbf{E}_N\}$. We first evaluate the conditional expectation $\mathbb{E}[|Y_{n,N}^1| | \mathcal{E}]$ to prove (157). Using the representation (137), the pseudo-Lipschitz property of f_n , and Proposition 1 yields

$$\begin{aligned} & \mathbb{E} [|Y_{n,N}^1| | \mathcal{E}] \\ & \leq L_n \mathbb{E} \left[\left| \frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} - \frac{1}{\sqrt{N-t}} \right| \|\mathbf{X}\| |z_n| \left\{ 1 + |a_{n,N}|^{k-1} \right. \right. \\ & \quad \left. \left. + |\tilde{\epsilon}_{n,N}|^{k-1} + \frac{\|\mathbf{X}\|^{k-1} |z_n|^{k-1}}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^{k-1}} + v_{N-t}^{\frac{k-1}{2}} |z_n|^{k-1} \right\} \right] | \mathcal{E} \end{aligned} \quad (160)$$

for some $L_n > 0$. Using the following upper bound:

$$\begin{aligned} \left| \frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} - \frac{1}{\sqrt{N-t}} \right| &= \frac{|N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2|}{\|\mathbf{u}_{\setminus \mathcal{N}}\| \sqrt{N-t} (\sqrt{N-t} + \|\mathbf{u}_{\setminus \mathcal{N}}\|)} \\ &< \frac{|N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2|}{N-t} \frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|}, \end{aligned} \quad (161)$$

we have

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} - \frac{1}{\sqrt{N-t}} \right| \|\mathbf{X}\| |z_n| |a_{n,N}|^{k-1} \right] | \mathcal{E} \\ & < v_{N-t}^{1/2} |a_{n,N}|^{k-1} \mathbb{E} \left[\frac{|N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2|}{\sqrt{N-t}} \frac{|z_n|}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} \right] | \mathcal{E} \end{aligned} \quad (162)$$

To evaluate the conditional expectation, we use the Cauchy-Schwarz inequality repeatedly to obtain

$$\begin{aligned} & \mathbb{E} \left[\frac{|N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2|}{\sqrt{N-t}} \frac{|z_n|}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} \right] | \mathcal{E} \\ & \leq \left\{ \mathbb{E} \left[\frac{|N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2|^2}{N-t} \right] \right\}^{1/2} \left\{ \mathbb{E} \left[\frac{|z_n|^2}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^2} \right] | \mathcal{E} \right\}^{1/2} \\ & \leq C \left\{ \mathbb{E} \left[\frac{|N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2|^2}{N-t} \right] \right\}^{1/2} \left\{ \mathbb{E} \left[\frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^4} \right] \right\}^{1/4} \\ & = \mathcal{O}(N^{-1/2}) \end{aligned} \quad (163)$$

for some $C > 0$, where the last follows from Propositions 3 and 4.

We repeat the same argument in evaluating the remaining terms in (160). We only present evaluation of the fourth term, since the other terms can be bounded in the same manner. Applying the upper bound (161) and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} - \frac{1}{\sqrt{N-t}} \right| \frac{\|\mathbf{X}\|^k |z_n|^k}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^{k-1}} \right] | \mathcal{E} \\ & < \frac{v_{N-t}^{k/2}}{\sqrt{N-t}} \mathbb{E} \left[\frac{|N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2|}{\sqrt{N-t}} \frac{(N-t)^{k/2} |z_n|^k}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^k} \right] | \mathcal{E} \\ & < \frac{C v_{N-t}^{k/2}}{\sqrt{N-t}} \left\{ \mathbb{E} \left[\frac{|N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2|^2}{N-t} \right] \right\}^{\frac{1}{2}} \left\{ \mathbb{E} \left[\frac{(N-t)^{2k}}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^{4k}} \right] \right\}^{\frac{1}{4}} \\ & \stackrel{\text{a.s.}}{=} \mathcal{O}(v_{N-t}^{k/2} N^{-1/2}), \end{aligned} \quad (164)$$

for some $C > 0$, where the last follows from Propositions 3 and 4. Evaluating the remaining terms on the RHS of (160) in the same manner, we arrive at

$$\begin{aligned} \mathbb{E} [|Y_{n,N}^1| | \mathcal{E}] & \stackrel{\text{a.s.}}{=} \mathcal{O} \left\{ \frac{L_n v_N^{1/2}}{\sqrt{N}} (1 + |a_{n,N}|^{k-1} \right. \\ & \quad \left. + |\tilde{\epsilon}_{n,N}|^{k-1} + v_N^{(k-1)/2}) \right\}. \end{aligned} \quad (165)$$

Using the Cauchy-Schwarz inequality to evaluate the expectation over \mathcal{E} , we obtain

$$\begin{aligned} \mathbb{E} [|Y_{n,N}^1|] & = \mathcal{O} \left(\frac{L_n}{\sqrt{N}} \left\{ \mathbb{E}[v_N^{1/2}] + (\mathbb{E}[v_N] \mathbb{E}[|a_{n,N}|^{2k-2}])^{1/2} \right. \right. \\ & \quad \left. \left. + (\mathbb{E}[v_N] \mathbb{E}[|\tilde{\epsilon}_{n,N}|^{2k-2}])^{1/2} + \mathbb{E}[v_N^{k/2}] \right\} \right), \end{aligned} \quad (166)$$

which reduces to (157), because of Proposition 5 and the moment properties of $a_{n,N}$ and $\tilde{\epsilon}_{n,N}$.

We next prove (158). Using the definition (155) of $Y_{n,N}^2$, the pseudo-Lipschitz property of f_n , and Proposition 1 yields

$$\begin{aligned} \frac{\mathbb{E}[|Y_{n,N}^2| | \mathcal{E}]}{L_n} & \leq |\tilde{\epsilon}_{n,N}| (1 + |a_{n,N}|^{k-1} + |\tilde{\epsilon}_{n,N}|^{k-1}) \\ & \quad + v_{N-t}^{(k-1)/2} |\tilde{\epsilon}_{n,N}| \mathbb{E} [|z_n|^{k-1}] \end{aligned} \quad (167)$$

for some $L_n > 0$. Using the Cauchy-Schwarz inequality and Proposition 2, we have

$$\begin{aligned} \frac{\mathbb{E}[|Y_{n,N}^2|]}{L_n} & \stackrel{\text{a.s.}}{\leq} C (\mathbb{E} [|\tilde{\epsilon}_{n,N}|^2])^{1/2} \\ & \quad \cdot (\mathbb{E} [1 + |a_{n,N}|^{2k-2} + |\tilde{\epsilon}_{n,N}|^{2k-2}])^{1/2} \\ & \quad + C (\mathbb{E}[v_{N-t}^{k-1}] \mathbb{E} [|\tilde{\epsilon}_{n,N}|^2])^{1/2} \end{aligned} \quad (168)$$

for some $C > 0$. Since $|\tilde{\epsilon}_{n,N}|$ has a vanishing second moment and finite $(2k - 2)$ th moment and since $|a_{n,N}|$ has a finite $(2k - 2)$ th moment, we use Proposition 5 to arrive at (158).

Finally, we prove (159). Using the definition (156) of $Y_{n,N}^3$, the pseudo-Lipschitz property of f_n , and Proposition 1 yields

$$|Y_{n,N}^3| \leq L_n \left| v_{N-t}^{1/2} - \sqrt{v} \right| |z_n| \left\{ 1 + |a_{n,N}|^{k-1} + v_{N-t}^{(k-1)/2} |z_n|^{k-1} + v^{(k-1)/2} |z_n|^{k-1} \right\}. \quad (169)$$

Evaluating the conditional expectation yields

$$\mathbb{E} [|Y_{n,N}^3| | \mathcal{E}] \stackrel{\text{a.s.}}{\leq} C L_n \left| v_{N-t}^{1/2} - \sqrt{v} \right| \cdot (1 + |a_{n,N}|^{k-1} + v_{N-t}^{(k-1)/2}) \quad (170)$$

for some $C > 0$. Using the Cauchy-Schwarz inequality to evaluate the expectation of the second term, we have

$$\begin{aligned} & \left(\mathbb{E} \left[\left| v_{N-t}^{1/2} - \sqrt{v} \right| |a_{n,N}|^{k-1} \right] \right)^2 \\ & \leq \mathbb{E} \left[\left| v_{N-t}^{1/2} - \sqrt{v} \right|^2 \right] \mathbb{E} [|a_{n,N}|^{2k-2}] \rightarrow 0, \end{aligned} \quad (171)$$

where the convergence follows from Proposition 5 and the moment assumption of $a_{n,N}$. For the last term, we let $\epsilon' \in (0, \epsilon/k)$ to have

$$\begin{aligned} & \mathbb{E} \left[\left\{ \left| v_{N-t}^{1/2} - \sqrt{v} \right| v_{N-t}^{(k-1)/2} \right\}^{1+\epsilon'} \right] \\ & \leq \mathbb{E} \left[v_{N-t}^{k(1+\epsilon')/2} \right] + v^{\frac{1+\epsilon'}{2}} \mathbb{E} \left[v_{N-t}^{(k-1)(1+\epsilon')/2} \right] < \infty, \end{aligned} \quad (172)$$

where the boundedness follows from Proposition 5. In other words, the last term on the upper bound (170) is uniformly integrable over $\|\mathbf{X}\|$. Thus, we use the assumption (12) to arrive at (159).

D. Proof of (15)

Since ϵ_N satisfies the assumptions (7) and (8), and since \mathbf{E}_N satisfies the assumption (10), the almost sure convergence $\|\delta_N\|^2 \stackrel{\text{a.s.}}{\rightarrow} 0$ implies that $\tilde{\epsilon}_N = \epsilon_N - \mathbf{E}_N \mathbf{o}(1)$ satisfies (7) and (8) with ϵ_N replaced by $\tilde{\epsilon}_N$. Furthermore, we only prove the case $t' = 1$ with $a_{n,1,N} = 0$ since an extension of the proof to the general case is straightforward. For notational simplicity, we write \mathbf{X}_{N-t} and $a_{n,0,N}$ as \mathbf{X} and $a_{n,N}$, and omit the tilde on $\tilde{\epsilon}_{n,N}$.

From the definitions (154) and (155) of $Y_{n,N}^1$ and $Y_{n,N}^2$, we need to prove

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N Y_{n,N}^1 \stackrel{\text{a.s.}}{=} 0, \quad (173)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N Y_{n,N}^2 \stackrel{\text{a.s.}}{=} 0, \quad (174)$$

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \left\{ f_n \left(a_{n,N}, v_{N-t}^{1/2} z_n \right) \right. \\ & \quad \left. - \mathbb{E}_{z_n} \left[f_n \left(a_{n,N}, \sqrt{v} z_n \right) \right] \right\} \stackrel{\text{a.s.}}{=} 0. \end{aligned} \quad (175)$$

Let us prove the first convergence (173). From the representation (137) and Theorem 1, it is sufficient to prove that $Y_{n,N}^1$ given in (154) satisfies

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n,n'=1}^N \mathbb{E} [|Y_{n,N}^1 Y_{n',N}^1| | \mathcal{E}] \stackrel{\text{a.s.}}{<} \infty, \quad (176)$$

with $\mathcal{E} = \{\|\mathbf{X}\|, \mathbf{a}_N, \epsilon_N, \mathbf{E}_N\}$.

Repeating the derivation of (160), we have

$$\begin{aligned} & \frac{\mathbb{E} [|Y_{n,N}^1 Y_{n',N}^1| | \mathcal{E}]}{L_n L_{n'}} \leq \mathbb{E} \left[\left| \frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} - \frac{1}{\sqrt{N-t}} \right|^2 \right. \\ & \cdot \|\mathbf{X}\|^2 |z_n| |z_{n'}| \left\{ 1 + |a_{n,N}|^{k-1} + |\epsilon_{n,N}|^{k-1} + v_{N-t}^{\frac{k-1}{2}} |z_n|^{k-1} \right. \\ & \left. \left. + \frac{\|\mathbf{X}\|^{k-1} |z_n|^{k-1}}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^{k-1}} \right\} \left\{ 1 + |a_{n',N}|^{k-1} + |\epsilon_{n',N}|^{k-1} \right. \right. \\ & \left. \left. + \frac{\|\mathbf{X}\|^{k-1} |z_{n'}|^{k-1}}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^{k-1}} + v_{N-t}^{(k-1)/2} |z_{n'}|^{k-1} \right\} | \mathcal{E} \right]. \end{aligned} \quad (177)$$

Let

$$A_{n,n'} = L_n L_{n'} \left| \frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} - \frac{1}{\sqrt{N-t}} \right|^2 \frac{\|\mathbf{X}\|^{2k} |z_n|^k |z_{n'}|^k}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^{2k-2}}, \quad (178)$$

$$\begin{aligned} & B_{n,n'} = L_n L_{n'} \left| \frac{1}{\|\mathbf{u}_{\setminus \mathcal{N}}\|} - \frac{1}{\sqrt{N-t}} \right|^2 \|\mathbf{X}\|^2 |z_n| |z_{n'}| \\ & \cdot (|a_{n,N}|^{k-1} + |\epsilon_{n,N}|^{k-1}) (|a_{n',N}|^{k-1} + |\epsilon_{n',N}|^{k-1}). \end{aligned} \quad (179)$$

We only evaluate the conditional expectation of $A_{n,n'}$ and $B_{n,n'}$, since the other terms can be evaluated in the same manner. Using the upper bound (161) yields

$$\begin{aligned} & \mathbb{E} [A_{n,n'} | \mathcal{E}] < L_n L_{n'} v_{N-t}^k \mathbb{E} \left[\frac{[N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2]^2}{(N-t)^2} \right. \\ & \quad \left. \cdot (N-t)^k \frac{|z_n|^k |z_{n'}|^k}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^{2k}} | \mathcal{E} \right]. \end{aligned} \quad (180)$$

Repeating the proof of (163), we find that the last factor is $\mathcal{O}(N^{-1})$. Thus, we obtain

$$\frac{1}{N} \sum_{n,n'=1}^N \mathbb{E} [A_{n,n'} | \mathcal{E}] \stackrel{\text{a.s.}}{=} \mathcal{O} \left\{ v_N^k \left(\frac{1}{N} \sum_{n=1}^N L_n \right)^2 \right\} \stackrel{\text{a.s.}}{=} \mathcal{O}(1), \quad (181)$$

because of the assumptions (12) and (14).

Similarly, we use the upper bound (161) to have

$$\begin{aligned} & \mathbb{E} [B_{n,n'} | \mathcal{E}] \\ & < L_n (|a_{n,N}|^{k-1} + |\epsilon_{n,N}|^{k-1}) L_{n'} (|a_{n',N}|^{k-1} + |\epsilon_{n',N}|^{k-1}) \\ & \cdot v_{N-t} \mathbb{E} \left[\frac{[N-t - \|\mathbf{u}_{\setminus \mathcal{N}}\|^2]^2}{N-t} \frac{|z_n| |z_{n'}|}{\|\mathbf{u}_{\setminus \mathcal{N}}\|^2} | \mathcal{E} \right]. \end{aligned} \quad (182)$$

We repeat the proof of (163) to find that the last factor is $\mathcal{O}(N^{-1})$. Thus, we arrive at

$$\begin{aligned} & \frac{1}{N} \sum_{n,n'=1}^N \mathbb{E}[B_{n,n'}|\mathcal{E}] \\ & \stackrel{\text{a.s.}}{=} \mathcal{O} \left\{ v_N \left(\frac{1}{N} \sum_{n=1}^N L_n (|a_{n,N}|^{k-1} + |\epsilon_{n,N}|^{k-1}) \right)^2 \right\} \\ & \stackrel{\text{a.s.}}{=} \mathcal{O} \left\{ v_N \left(\frac{1}{N} \sum_{n=1}^N L_n (|a_{n,N}|^{2k-2} + |\epsilon_{n,N}|^{2k-2}) \right)^2 \right\} \\ & = \mathcal{O}(1), \end{aligned} \quad (183)$$

where the second equality follows from the Cauchy-Schwarz inequality and the assumption (14), and where the last is due to the assumptions (8), (9), and (12). Evaluating the conditional expectation of the other terms in (177) in the same manner, we arrive at (176). Thus, (173) holds.

We next prove the second convergence (174). Repeating the proof of (167) yields

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N |Y_{n,N}^2| & \leq \frac{1}{N} \sum_{n=1}^N L_n |\epsilon_{n,N}| \left\{ 1 + |a_{n,N}|^{k-1} \right. \\ & \quad \left. + |\epsilon_{n,N}|^{k-1} + v_{N-t}^{(k-1)/2} |z_n|^{k-1} \right\}. \end{aligned} \quad (184)$$

Using the Cauchy-Schwarz inequality for the second term yields

$$\begin{aligned} & \left(\frac{1}{N} \sum_{n=1}^N L_n |\epsilon_{n,N}| |a_{n,N}|^{k-1} \right)^2 \\ & \leq \frac{1}{N} \sum_{n=1}^N L_n |\epsilon_{n,N}|^2 \frac{1}{N} \sum_{n=1}^N L_n |a_{n,N}|^{2k-2} \stackrel{\text{a.s.}}{\rightarrow} 0 \end{aligned} \quad (185)$$

as $N \rightarrow \infty$, because of the assumptions (7) and (9). Similarly, we find that the third term converges almost surely to zero as $N \rightarrow \infty$.

We use the Cauchy-Schwarz inequality for the last term on the upper bound (184) to obtain

$$\begin{aligned} \frac{v_{N-t}^{(k-1)/2}}{N} \sum_{n=1}^N L_n |\epsilon_{n,N}| |z_n|^{k-1} & \leq \left(\frac{1}{N} \sum_{n=1}^N L_n |\epsilon_{n,N}|^2 \right)^{1/2} \\ & \cdot v_{N-t}^{\frac{k-1}{2}} \left(\frac{1}{N} \sum_{n=1}^N L_n |z_n|^{2k-2} \right)^{\frac{1}{2}}. \end{aligned} \quad (186)$$

The assumptions (7) and (12) imply that the first and second factors converge almost surely to zero and $v_{N-t}^{(k-1)/2}$ as $N \rightarrow \infty$, respectively. Furthermore, from the assumption (14) we use Theorem 1 to find

$$\frac{1}{N} \sum_{n=1}^N L_n |z_n|^{2k-2} \stackrel{\text{a.s.}}{=} \frac{\mathbb{E}[|z_1|^{2k-2}]}{N} \sum_{n=1}^N L_n + o(1) < \infty. \quad (187)$$

Thus, the last term on the upper bound (184) converges almost surely to zero as $N \rightarrow \infty$. Since the almost sure convergence of the remaining terms to zero can be proved in the same manner, we arrive at (174).

Finally, we prove the last convergence (175). We observe that $\{f_n(a_{n,N}, v_{N-t}^{1/2} z_n)\}$ are conditionally independent given \mathcal{E} . Furthermore, we use the pseudo-Lipschitz property of f_n to obtain

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \mathbb{V} \left[f_n \left(a_{n,N}, v_{N-t}^{1/2} z_n \right) \middle| \mathcal{E} \right] \\ & \leq \frac{v_{N-t}}{N} \sum_{n=1}^N L_n^2 \mathbb{E}_{z_n, Z} [|z_n - Z|^2 (1 + |a_{n,N}|^{2k-2} \\ & \quad + v_{N-t}^{k-1} |z_n|^{2k-2} + v_{N-t}^{k-1} |Z|^{2k-2}) | \mathcal{E}] \stackrel{\text{a.s.}}{<} \infty, \end{aligned} \quad (188)$$

where Z is a standard complex Gaussian random variable and independent of z_n , and where the boundedness follows from the assumptions (9), (12), and (14). Thus, we can use Theorem 1 to find

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \left\{ f_n \left(a_{n,N}, v_{N-t}^{1/2} z_n \right) \right. \\ & \quad \left. - \mathbb{E} \left[f_n \left(a_{n,N}, v_{N-t}^{1/2} z_n \right) \middle| \mathcal{E} \right] \right\} \stackrel{\text{a.s.}}{=} 0. \end{aligned} \quad (189)$$

To obtain (175), from the definition (156) of $Y_{n,N}^3$ we need to prove $N^{-1} \sum_{n=1}^N \mathbb{E}[Y_{n,N}^3 | \mathcal{E}] \stackrel{\text{a.s.}}{\rightarrow} 0$ as $N \rightarrow \infty$. Using (170) yields

$$\begin{aligned} & \left| \frac{1}{N} \sum_{n=1}^N \mathbb{E} [Y_{n,N}^3 | \mathcal{E}] \right| \stackrel{\text{a.s.}}{\leq} C \left| v_{N-t}^{1/2} - \sqrt{v} \right| \\ & \quad \cdot \frac{1}{N} \sum_{n=1}^N L_n (1 + |a_{n,N}|^{k-1} + v_{N-t}^{(k-1)/2}), \end{aligned} \quad (190)$$

which converges almost surely to zero as $N \rightarrow \infty$, because of the assumptions (9), (12), and (14). Thus, (175) holds.

APPENDIX B DERIVATION OF MESSAGE-PASSING

EP [20], [29] provides a framework for deriving MP algorithms that calculate the marginal posterior distribution $p(x_n | \mathbf{y}, \mathbf{A}) = \int p(\mathbf{x} | \mathbf{y}, \mathbf{A}) d\mathbf{x}_{\setminus n}$, in which $\mathbf{x}_{\setminus n}$ is the vector obtained by eliminating x_n from \mathbf{x} . We consider the large system limit to derive an MP algorithm, which coincides with the algorithm derived in a heuristic manner [27].

We approximate the marginal posterior distribution $p(x_n | \mathbf{y}, \mathbf{A})$ by a tractable probability density function (pdf) $q_A(x_n) = \int q_A(\mathbf{x}) d\mathbf{x}_{\setminus n}$, given by

$$q_A(\mathbf{x}) \propto p(\mathbf{y} | \mathbf{A}, \mathbf{x}) \prod_{n=1}^N q_{B \rightarrow A}(x_n). \quad (191)$$

In (191), the notation $f(\mathbf{x}) \propto g(\mathbf{x})$ means that there is a positive constant C such that $f(\mathbf{x}) = Cg(\mathbf{x})$ holds. Furthermore, $q_{B \rightarrow A}(x_n)$ is a conjugate prior to the likelihood $p(\mathbf{y} | \mathbf{A}, \mathbf{x})$. When the noise vector \mathbf{w} in (1) is regarded as a CSCG random vector with covariance $\sigma^2 \mathbf{I}_M$, the conjugate prior $q_{B \rightarrow A}(x_n)$ is proper complex Gaussian,

$$q_{B \rightarrow A}(x_n) \propto \exp \left(-\frac{|x_n - x_{n,B \rightarrow A}|^2}{v_{B \rightarrow A}} \right), \quad (192)$$

where $x_{n,B \rightarrow A}$ and $v_{B \rightarrow A}$ are the mean and variance of $q_{B \rightarrow A}(x_n)$, respectively. In order to derive the MP algorithm proposed in [27], we have selected the identical variance $v_{B \rightarrow A}$ for all n , while Céspedes *et al.* [20] selected different values for different n to improve the performance for finite-sized systems.

We first evaluate the marginal pdf $q_A(x_n)$ in the large system limit, defined via (191). Since the conjugate prior (192) has been selected, the joint pdf $q_A(\mathbf{x})$ is also Gaussian.

$$q_A(\mathbf{x}) \propto \exp\left\{-\left(\mathbf{x} - \mathbf{x}_A\right)^H \mathbf{V}_A^{-1} \left(\mathbf{x} - \mathbf{x}_A\right)\right\}, \quad (193)$$

where the mean and covariance are given by

$$\mathbf{x}_A = \mathbf{x}_{B \rightarrow A} + \frac{1}{\sigma^2} \mathbf{V}_A \mathbf{A}^H (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A}), \quad (194)$$

$$\mathbf{V}_A = \left(\frac{1}{v_{B \rightarrow A}} \mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{A}^H \mathbf{A} \right)^{-1}, \quad (195)$$

respectively. Using the matrix inversion lemma, it is possible to show that (194) and (195) reduce to

$$\mathbf{x}_A = \mathbf{x}_{B \rightarrow A} + v_{B \rightarrow A} \mathbf{A}^H \Xi^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A}), \quad (196)$$

$$[\mathbf{V}_A]_{n,n} = v_{B \rightarrow A} - \mathbf{a}_n^H \Xi^{-1} \mathbf{a}_n v_{B \rightarrow A}^2, \quad (197)$$

respectively, with

$$\Xi = \sigma^2 \mathbf{I}_M + v_{B \rightarrow A} \mathbf{A} \mathbf{A}^H. \quad (198)$$

We shall prove that $\mathbf{a}_n^H \Xi^{-1} \mathbf{a}_n$ converges almost surely to $\gamma(v_{B \rightarrow A})^{-1}$ in the large system limit for all n , in which $\gamma(v_{B \rightarrow A})$ is given by (25). Applying the SVD (19) to $\mathbf{a}_n^H \Xi^{-1} \mathbf{a}_n$, defined via (198), we have

$$\mathbf{a}_n^H \Xi^{-1} \mathbf{a}_n = \mathbf{e}_n^H \mathbf{V} \mathbf{D} \mathbf{V}^H \mathbf{e}_n, \quad (199)$$

with

$$\mathbf{D} = \begin{pmatrix} \Sigma \\ \mathbf{O} \end{pmatrix} (\sigma^2 \mathbf{I}_M + v_{B \rightarrow A} \Sigma^2)^{-1} (\Sigma, \mathbf{O}). \quad (200)$$

In (199), \mathbf{e}_n denotes the n th column of \mathbf{I}_N . Thus, Corollary 1 and Assumption 2 imply that $\mathbf{a}_n^H \Xi^{-1} \mathbf{a}_n$ converges almost surely to $\gamma(v_{B \rightarrow A})^{-1}$ in the large system limit.

This observation indicates that for any n the diagonal element (197) converges almost surely to

$$v_A = v_{B \rightarrow A} - \gamma^{-1}(v_{B \rightarrow A}) v_{B \rightarrow A}^2 \quad (201)$$

in the large system limit. Thus, the marginal pdf $q_A(x_n) = \int q_A(\mathbf{x}) d\mathbf{x}_{\setminus n}$ is the proper complex Gaussian pdf with mean $x_{n,A} = [\mathbf{x}_A]_n$ and variance v_A , i.e.

$$q_A(x_n) \propto \exp\left(-\frac{|x_n - x_{n,A}|^2}{v_A}\right). \quad (202)$$

In order to present a crucial step in EP, we define the extrinsic pdf of x_n as

$$q_{A \rightarrow B}(x_n) \propto \frac{q_A(x_n)}{q_{B \rightarrow A}(x_n)}. \quad (203)$$

Let $x_{n,B}$ and $v_{n,B}$ denote the mean and variance of x_n with respect to the pdf $p_B(x_n) \propto q_{A \rightarrow B}(x_n) p(x_n)$. The crucial step

in EP is to update the message $q_{B \rightarrow A}(x_n)$ so as to satisfy the moment matching conditions [29],

$$\mathbb{E}_{q_B}[x_n] = x_{n,B}, \quad (204)$$

$$\mathbb{V}_{q_B}[x_n] = \frac{1}{N} \sum_{n=1}^N v_{n,B} \equiv v_B, \quad (205)$$

where the expectations are taken with respect to

$$q_B(x_n) \propto q_{A \rightarrow B}(x_n) q_{B \rightarrow A}^{\text{new}}(x_n). \quad (206)$$

In (206), the updated pdf $q_{B \rightarrow A}^{\text{new}}(x_n)$ is given by

$$q_{B \rightarrow A}^{\text{new}}(x_n) \propto \exp\left(-\frac{|x_n - x_{n,B \rightarrow A}^{\text{new}}|^2}{v_{B \rightarrow A}^{\text{new}}}\right). \quad (207)$$

We first derive module A. Using (192) and (202), we find that the extrinsic pdf (203) reduces to

$$q_{A \rightarrow B}(x_n) \propto \exp\left(-\frac{|x_n - x_{n,A \rightarrow B}|^2}{v_{A \rightarrow B}}\right), \quad (208)$$

with

$$x_{n,A \rightarrow B} = v_{A \rightarrow B} \left(\frac{x_{n,A}}{v_A} - \frac{x_{n,B \rightarrow A}}{v_{B \rightarrow A}} \right), \quad (209)$$

$$\frac{1}{v_{A \rightarrow B}} = \frac{1}{v_A} - \frac{1}{v_{B \rightarrow A}}. \quad (210)$$

Substituting (209) into (210) yields

$$v_{A \rightarrow B} = \gamma(v_{B \rightarrow A}) - v_{B \rightarrow A}, \quad (211)$$

which results in the update rule (22). Similarly, Applying (196), (201), (210), and (211) to (209), we arrive at

$$\mathbf{x}_{A \rightarrow B} = \mathbf{x}_{B \rightarrow A} + \gamma(v_{B \rightarrow A}) \mathbf{A}^H \Xi^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A}), \quad (212)$$

which implies the update rule (21).

We next evaluate the moment matching conditions (204) and (205) to derive module B. Substituting (207) and (208) into (206) yields

$$q_B(x_n) \propto \exp\left(-\frac{|x_n - \tilde{x}_{n,B}|^2}{\tilde{v}_B}\right), \quad (213)$$

with

$$\tilde{x}_{n,B} = \tilde{v}_B \left(\frac{x_{n,A \rightarrow B}}{v_{A \rightarrow B}} + \frac{x_{n,B \rightarrow A}^{\text{new}}}{v_{B \rightarrow A}^{\text{new}}} \right), \quad (214)$$

$$\frac{1}{\tilde{v}_B} = \frac{1}{v_{A \rightarrow B}} + \frac{1}{v_{B \rightarrow A}^{\text{new}}}. \quad (215)$$

Using the moment matching conditions (204) and (205), we arrive at the update rules (29) and (30) in module B,

$$\mathbf{x}_{B \rightarrow A}^{\text{new}} = v_{B \rightarrow A}^{\text{new}} \left(\frac{\mathbf{x}_B}{v_B} - \frac{\mathbf{x}_{A \rightarrow B}}{v_{A \rightarrow B}} \right), \quad (216)$$

$$\frac{1}{v_{B \rightarrow A}^{\text{new}}} = \frac{1}{v_B} - \frac{1}{v_{A \rightarrow B}}. \quad (217)$$

APPENDIX C
 PROOF OF LEMMA 2

We utilize the following technical lemma:

Lemma 4: We define the cumulant generating function $\chi_t : \mathbb{C} \rightarrow \mathbb{R}$ of the posterior distribution of x_n as

$$\chi_t(z) = \frac{v_{A \rightarrow B}^t}{2} \ln \mathbb{E}_{x_n} \left[\exp \left(-\frac{|z - x_n|^2}{v_{A \rightarrow B}^t} \right) \right] + \frac{|z|^2}{2}. \quad (218)$$

Then, χ_t is twice continuously differentiable with respect to $\Re[z]$ and $\Im[z]$, and satisfies

$$\frac{\partial \chi_t}{\partial \Re[z]} = \Re[\tilde{\eta}_t(z)], \quad \frac{\partial \chi_t}{\partial \Im[z]} = \Im[\tilde{\eta}_t(z)], \quad (219)$$

$$\frac{v_{A \rightarrow B}^t}{2} \frac{\partial^2 \chi_t}{\partial \Re[z]^2} = \mathbb{E} \left[(\Re[x_n] - \Re[\tilde{\eta}_t(z)])^2 | z \right], \quad (220)$$

$$\frac{v_{A \rightarrow B}^t}{2} \frac{\partial^2 \chi_t}{\partial \Im[z]^2} = \mathbb{E} \left[(\Im[x_n] - \Im[\tilde{\eta}_t(z)])^2 | z \right], \quad (221)$$

$$\frac{v_{A \rightarrow B}^t}{2} \frac{\partial^2 \chi_t}{\partial \Re[z] \partial \Im[z]} = \mathbb{E} \left[\Re[x_n] \Im[x_n] | z \right] - \Re[\tilde{\eta}_t(z)] \Im[\tilde{\eta}_t(z)]. \quad (222)$$

Proof: The former statement follows from Assumption 1 and the dominated convergence theorem. The latter statement is obtained by calculating the derivatives of χ_t directly. ■

We first prove the Lipschitz-continuity of $\tilde{\eta}_t$. We need to prove that all first-order derivatives of $\Re[\tilde{\eta}_t]$ and $\Im[\tilde{\eta}_t]$ are bounded. From Assumption 4 and Lemma 4, it is sufficient to confirm that (222) is almost surely bounded. Using the Cauchy-Schwarz inequality yields

$$\begin{aligned} & \left(\mathbb{E} \left[\Re[x_n] \Im[x_n] | z \right] - \Re[\tilde{\eta}_t(z)] \Im[\tilde{\eta}_t(z)] \right)^2 \\ &= \left(\mathbb{E} \left[(\Re[x_n] - \Re[\tilde{\eta}_t(z)]) (\Im[x_n] - \Im[\tilde{\eta}_t(z)]) | z \right] \right)^2 \\ &\leq \mathbb{E} \left[(\Re[x_n] - \Re[\tilde{\eta}_t(z)])^2 | z \right] \mathbb{E} \left[(\Im[x_n] - \Im[\tilde{\eta}_t(z)])^2 | z \right], \end{aligned} \quad (223)$$

which is almost surely bounded, because of Assumption 4. Thus, $\tilde{\eta}_t$ is Lipschitz-continuous.

We next prove (33) and (34). For notational convenience, we write $\tilde{\eta}_t(x_n + z)$ as $\tilde{\eta}$. By definition, we have

$$z^* \tilde{\eta} = \Re[z] \Re[\tilde{\eta}] + \Im[z] \Im[\tilde{\eta}] + i(\Re[z] \Im[\tilde{\eta}] - \Im[z] \Re[\tilde{\eta}]). \quad (224)$$

Since $\Re[z]$ and $\Im[z]$ are independent Gaussian random variables with zero-mean and variance $v_{A \rightarrow B}^t/2$, using Stein's lemma [41] yields

$$\begin{aligned} \mathbb{E}_z [z^* \tilde{\eta}] &= \frac{v_{A \rightarrow B}^t}{2} \mathbb{E}_z \left[\frac{\partial \Re[\tilde{\eta}]}{\partial \Re[z]} + \frac{\partial \Im[\tilde{\eta}]}{\partial \Im[z]} \right] \\ &\quad + \frac{i v_{A \rightarrow B}^t}{2} \mathbb{E}_z \left[\frac{\partial \Im[\tilde{\eta}]}{\partial \Re[z]} - \frac{\partial \Re[\tilde{\eta}]}{\partial \Im[z]} \right] \\ &= v_{A \rightarrow B}^t \mathbb{E}_z \left[\frac{\partial}{\partial z} (\Re[\tilde{\eta}] + i \Im[\tilde{\eta}]) \right], \end{aligned} \quad (225)$$

where $\partial/\partial z$ denotes the Wirtinger derivative (32). This implies that (33) holds. Furthermore, applying Lemma 4 to the former expression in (225), we obtain

$$\mathbb{E}_z [z^* \tilde{\eta}] = \mathbb{E}_z \left[|x_n - \tilde{\eta}_t(x_n + z)|^2 \right]. \quad (226)$$

Taking the expectation of both sides over x_n , we arrive at Lemma 2.

 APPENDIX D
 PROOF OF LEMMA 3

For $\hat{\mathbf{V}} = \mathbf{V} \Phi_{\mathbf{X}} \in \mathcal{U}_N$, we first prove the identity

$$\hat{\mathbf{V}} = \left(\Phi_{\mathbf{Y}}^{\parallel}, \Phi_{\mathbf{Y}}^{\perp} \tilde{\mathbf{V}} \right), \quad (227)$$

with some unitary matrix $\tilde{\mathbf{V}} \in \mathcal{U}_{N-t}$.

Since \mathbf{V} is unitary, using the constraint (79) yields $\mathbf{X}^H \mathbf{X} = \mathbf{Y}^H \mathbf{Y}$. This implies that \mathbf{X} and \mathbf{Y} have identical singular values and right-singular vectors, i.e. $\mathbf{X} = \Phi_{\mathbf{X}}(\Sigma_{\mathbf{X}}, \mathbf{O})^T \Psi_{\mathbf{X}}^H$ and $\mathbf{Y} = \Phi_{\mathbf{Y}}(\Sigma_{\mathbf{Y}}, \mathbf{O})^T \Psi_{\mathbf{Y}}^H$ with $\Sigma_{\mathbf{X}} = \Sigma_{\mathbf{Y}}$ and $\Psi_{\mathbf{X}} = \Psi_{\mathbf{Y}}$. Since $\Sigma_{\mathbf{X}} = \Sigma_{\mathbf{Y}}$ is assumed to be invertible, applying these SVDs to the constraint (79) yields

$$\Phi_{\mathbf{Y}}^{\parallel} = \mathbf{V} \Phi_{\mathbf{X}} \begin{pmatrix} \mathbf{I}_t \\ \mathbf{O}_{N \times (N-t)} \end{pmatrix}. \quad (228)$$

Consider the partition $\hat{\mathbf{V}} = (\hat{\mathbf{V}}_0, \hat{\mathbf{V}}_1)$, with $\hat{\mathbf{V}}_0 \in \mathbb{C}^{N \times t}$ and $\hat{\mathbf{V}}_1 \in \mathbb{C}^{N \times (N-t)}$. From (228) we have $\hat{\mathbf{V}}_0 = \Phi_{\mathbf{Y}}^{\parallel}$. Thus, the orthogonality between the columns of $\hat{\mathbf{V}}_0$ and $\hat{\mathbf{V}}_1$ implies the structure (227) with some matrix $\tilde{\mathbf{V}} \in \mathbb{C}^{(N-t) \times (N-t)}$. Furthermore, from the orthonormality between the columns of $\hat{\mathbf{V}}_1$ we find that $\tilde{\mathbf{V}}$ is a unitary matrix. Thus, (227) is correct.

We next prove that (227) is equivalent to the RHS of (80). Substituting (227) into $\mathbf{V} = \hat{\mathbf{V}} \Phi_{\mathbf{X}}^H$ yields

$$\mathbf{V} = \Phi_{\mathbf{Y}}^{\parallel} (\Phi_{\mathbf{X}}^{\parallel})^H + \Phi_{\mathbf{Y}}^{\perp} \tilde{\mathbf{V}} (\Phi_{\mathbf{X}}^{\perp})^H. \quad (229)$$

It is straightforward to confirm that the first term on the RHS of (80) reduces to $\Phi_{\mathbf{Y}}^{\parallel} (\Phi_{\mathbf{X}}^{\parallel})^H$, by using the SVDs of \mathbf{X} and \mathbf{Y} with $\Sigma_{\mathbf{X}} = \Sigma_{\mathbf{Y}}$ and $\Psi_{\mathbf{X}} = \Psi_{\mathbf{Y}}$.

To complete the proof of Lemma 3, we prove that $\tilde{\mathbf{V}} \in \mathcal{U}_{N-t}$ is a Haar matrix independent of \mathbf{X} and \mathbf{Y} . Since the Haar matrix \mathbf{V} is bi-unitarily invariant, we have $\mathbf{V} \Phi_{\mathbf{X}} \sim \mathbf{V}$. Thus, without loss of generality, (228) allows us to assume $\mathbf{X} = (\mathbf{I}_t, \mathbf{O})^T$ in the constraint (79). Under this assumption, conditioning on \mathbf{X} and \mathbf{Y} is equivalent to conditioning the first t columns \mathbf{V}_0 of \mathbf{V} .

Consider the following structure:

$$\mathbf{V} = \left(\mathbf{V}_0, \Phi_{\mathbf{V}_0}^{\perp} \tilde{\mathbf{V}} \right). \quad (230)$$

We prove that \mathbf{V} is Haar-distributed if and only if $\tilde{\mathbf{V}}$ is a Haar matrix and independent of \mathbf{V}_0 . Since \mathbf{X} and \mathbf{Y} depend on \mathbf{V} only through \mathbf{V}_0 , we arrive at Lemma 3.

For any deterministic unitary matrix $\Phi \in \mathcal{U}_N$, it is known that the left-invariance $\Phi \mathbf{V} \sim \mathbf{V}$ induces the Haar measure on the unitary group of dimension N satisfying $\mathbf{V} \sim \mathbf{V}^H$, so that we have the right-invariance $\mathbf{V} \Psi \sim \mathbf{V}^H \Psi = (\Psi^H \mathbf{V})^H \sim \mathbf{V}^H \sim \mathbf{V}$ for any deterministic $\Psi \in \mathcal{U}_N$. Thus, we only consider the left-invariance $\Phi \mathbf{V} \sim \mathbf{V}$.

There is some unitary matrix $\mathbf{U}_{\mathbf{V}_0} \in \mathcal{U}_{N-t}$ such that $\Phi \Phi_{\mathbf{V}_0}^{\perp} = \Phi_{\mathbf{V}_0}^{\perp} \mathbf{U}_{\mathbf{V}_0}$ holds, because of

$$\Phi \Phi_{\mathbf{V}_0}^{\perp} (\Phi \Phi_{\mathbf{V}_0}^{\perp})^H = \Phi (\mathbf{I}_N - \mathbf{V}_0 \mathbf{V}_0^H) \Phi^H = \mathbf{P}_{\Phi_{\mathbf{V}_0}^{\perp}}. \quad (231)$$

This implies that (230) satisfies

$$\Phi \mathbf{V} = \left(\Phi \mathbf{V}_0, \Phi_{\Phi_{\mathbf{V}_0}^{\perp}} \mathbf{U}_{\mathbf{V}_0} \tilde{\mathbf{V}} \right), \quad (232)$$

which indicates that $\Phi V \sim V$ holds if and only if $(\Phi V_0, U_{V_0} \tilde{V}) \sim (V_0, \tilde{V})$ is satisfied. Since V_0 is Haar-distributed, $\Phi V \sim V$ holds if and only if \tilde{V} is a Haar matrix independent of V_0 . Thus, Lemma 3 holds.

REFERENCES

- [1] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," in *Proc. 2017 IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 501–505.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [3] E. J. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [4] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.
- [5] D. J. Thouless, P. W. Anderson, and R. G. Palmer, "Solution of 'solvable model of a spin glass'," *Philos. Mag.*, vol. 35, no. 3, pp. 593–601, 1977.
- [6] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A: Math. Gen.*, vol. 36, no. 43, pp. 11 111–11 121, Oct. 2003.
- [7] K. Takeuchi, T. Tanaka, and T. Kawabata, "Performance improvement of iterative multiuser detection for large sparsely-spread CDMA systems by spatial coupling," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1768–1794, Apr. 2015.
- [8] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [9] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact," in *Proc. 2016 IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 665–669.
- [10] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," in *Proc. 54th Annual Allerton Conf. Commun. Control & Computing*, Urbana-Champaign, IL, USA, Sep. 2016, pp. 625–632.
- [11] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
- [12] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
- [13] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.
- [14] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, pp. 021 005–1–18, May 2012.
- [15] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.
- [16] F. Caltagirone, L. Zdeborová, and F. Krzakala, "On convergence of approximate message passing," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014, pp. 1812–1816.
- [17] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014, pp. 236–240.
- [18] M. Opper and O. Winther, "Expectation consistent approximate inference," *J. Mach. Learn. Res.*, vol. 6, pp. 2177–2204, Dec. 2005.
- [19] —, "Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling," *Phys. Rev. E*, vol. 64, no. 5, pp. 056 131–1–14, 2001.
- [20] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug. 2014.
- [21] Y. Kabashima and M. Vehkaperä, "Signal recovery using expectation consistent approximation for linear observations," in *Proc. 2014 IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014, pp. 226–230.
- [22] A. K. Fletcher, M. Sahræe-Ardakan, S. Rangan, and P. Schniter, "Expectation consistent approximate inference: Generalizations and convergence," in *Proc. 2016 IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 190–194.
- [23] B. Çakmak, O. Winther, and B. H. Fleury, "S-AMP: Approximate message passing for general matrix ensembles," in *Proc. 2014 IEEE Inf. Theory Workshop*, Hobart, Australia, Nov. 2014, pp. 192–196.
- [24] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *Proc. 2017 IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 1588–1592.
- [25] X. Yuan and J. Ma, "Iterative equalization for MIMO systems: Algorithm design and evolution analysis," in *Proc. 2013 IEEE Wirel. Commun. Netw. Conf.*, Shanghai, China, Apr. 2013, pp. 3974–3979.
- [26] J. Ma, X. Yuan, and L. Ping, "Turbo compressed sensing with partial DFT sensing matrix," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 158–161, Feb. 2015.
- [27] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, Jan. 2017.
- [28] T. Liu, C.-K. Wen, S. Jin, and X. You, "Generalized turbo signal recovery for nonlinear measurements and orthogonal sensing matrices," in *Proc. 2016 IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 2883–2887.
- [29] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell.*, Seattle, WA, USA, Aug. 2001, pp. 362–369.
- [30] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," *Europhys. Lett.*, vol. 76, no. 6, pp. 1193–1199, 2006.
- [31] A. M. Tulino, G. Caire, S. Verdú, and S. Shamai (Shitz), "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, Jul. 2013.
- [32] J. Barbier, N. Macris, A. Maillard, and F. Krzakala, "The mutual information in random linear estimation beyond i.i.d. matrices," in *Proc. 2018 IEEE Int. Symp. Inf. Theory*, Vail, CO, USA, Jun. 2018, pp. 1390–1394.
- [33] E. Bolthausen, "An iterative construction of solutions of the TAP equations for the Sherrington-Kirkpatrick model," *Commun. Math. Phys.*, vol. 325, no. 1, pp. 333–366, Jan. 2014.
- [34] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. Amer. Math. Soc., 2000.
- [35] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA USA: Now Publishers Inc., 2004.
- [36] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," [Online]. Available: <http://arxiv.org/abs/1610.03082>.
- [37] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," [Online]. Available: <http://arxiv.org/abs/1701.05284>.
- [38] R. Lyons, "Strong laws of large numbers for weakly correlated random variables," *Michigan Math. J.*, vol. 35, no. 3, pp. 353–359, 1988.
- [39] K. Takeuchi, "A unified framework of state evolution for message-passing algorithms," accepted for presentation in *2019 IEEE Int. Symp. Inf. Theory*, 2019, [Online]. Available: <http://arxiv.org/abs/1901.03041>.
- [40] N. Etemadi, "An elementary proof of the strong law of large numbers," *Z. Wahrsch. Verw. Gebiete*, vol. 55, no. 1, pp. 119–122, Feb. 1981.
- [41] C. Stein, "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," in *6th Berkeley Symp. Math. Statist. Prob.*, vol. 2, 1972, pp. 583–602.
- [42] E. Meckes, "Linear functions on the classical matrix groups," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5355–5366, Oct. 2008.
- [43] S. Chatterjee and E. Meckes, "Multivariate normal approximation using exchangeable pairs," *ALEA Latin. Amer. J. Prob. Math. Stat.*, vol. 4, pp. 257–283, 2008.
- [44] P. Billingsley, *Probability and Measure*, 3rd ed. John Wiley & Sons, 1995.