Decentralized Generalized Approximate Message-Passing for Tree-Structured Networks

Keigo Takeuchi, *Member, IEEE*

***Abstract*—Decentralized generalized approximate message-passing (GAMP) is proposed for compressed sensing from distributed generalized linear measurements in a tree-structured network. Consensus propagation is used to realize average consensus required in GAMP via local communications between adjacent nodes. Decentralized GAMP is applicable to all tree-structured networks that do not necessarily have central nodes connected to all other nodes. State evolution is used to analyze the asymptotic dynamics of decentralized GAMP for zero-mean independent and identically distributed Gaussian sensing matrices. The state evolution recursion for decentralized GAMP is proved to have the same fixed points as that for centralized GAMP when homogeneous measurements with an identical dimension in all nodes are considered. Furthermore, existing long-memory proof strategy is used to prove that the state evolution recursion for decentralized GAMP with the Bayes-optimal denoisers converges to a fixed point. These results imply that the state evolution recursion for decentralized GAMP with the Bayes-optimal denoisers converges to the Bayes-optimal fixed point for the homogeneous measurements when the fixed point is unique. Numerical results for decentralized GAMP are presented in the cases of linear measurements and clipping. As examples of tree-structured networks, a one-dimensional chain and a tree with no central nodes are considered.**

***Index Terms*—Compressed sensing, generalized approximate message-passing, decentralized algorithms, consensus propagation, tree-structured networks, state evolution.**

I. INTRODUCTION

A. Background

APPROXIMATE message-passing (AMP) [1] is a powerful iterative algorithm for signal recovery from linear measurements [2], [3]. In particular, AMP using the Bayes-optimal denoiser—called Bayes-optimal AMP—is regarded as an asymptotically exact approximation of loopy belief propagation [4]. Applications of AMP contain compressive imaging [5], [6], radar [7], sparse superposition codes [8], [9], and low-rank matrix estimation [10], [11].

State evolution [12]–[14], motivated by [15], allows us to analyze the asymptotic dynamics of AMP rigorously when the sensing matrix has independent and identically distributed (i.i.d.) zero-mean sub-Gaussian elements. The asymptotic dynamics of AMP is characterized with a discrete-time dynamical system—called state evolution recursion. When the state evolution recursion has a unique fixed point, Bayes-optimal AMP was proved in [12], [13] to achieve the Bayes-optimal performance [16], [17] asymptotically.

Generalized AMP (GAMP) [18] is a generalization of AMP to the case of generalized linear measurements, which allow us to treat general noise beyond the additive noise in the linear measurements. GAMP expands applications of AMP to one-bit compressed sensing [19], [20], phase retrieval [21], [22], and peak-to-average power ratio (PAPR) reduction [23], [24]. Like AMP, the asymptotic dynamics of GAMP was analyzed via state evolution [25]. When the state evolution recursion of GAMP has a unique fixed point, Bayes-optimal GAMP was proved in [26] to achieve the theoretically optimal performance in terms of the minimum mean-square error (MMSE).

Distributed algorithms are more desirable than centralized algorithms that run on a single node exploiting the full information about the sensing matrix and all measurements. Distributed algorithms are separated into two types of algorithms.

A first type contains distributed algorithms that run on a central node and multiple remote nodes. Each remote node only uses local measurements to compute a local estimate, which is aggregated in the central node. The central node combines the local estimates to obtain a global estimate, which is fed back to the remote nodes. The iteration between the central and remote nodes is repeated until the algorithm reaches a final result. Iterative thresholding algorithms for compressed sensing, such as iterative shrinkage thresholding algorithm (ISTA) [27], fast iterative shrinkage thresholding algorithm (FISTA) [28], and iterative hard thresholding (IHT) [29], can be implemented as this type of distributed algorithms. See [30] for distributed IHT.

In the other type of distributed algorithms—called decentralized algorithms in this paper, algorithms run on multiple nodes in an ad hoc network with no central nodes. Decentralized protocols for average consensus [31], [32] are utilized to reach the same result as the corresponding centralized algorithm only by sharing processing results with adjacent nodes locally. As this type of distributed algorithms for compressed sensing, distributed least absolute shrinkage and selection operator (LASSO) [33], distributed spectrum sensing [34], distributed basis pursuit [35], and distributed alternating direction method of multipliers (ADMM) [36], [37] were proposed.

AMP was extended to distributed AMP [38]–[43] exploiting a central node. More precisely, distributed AMP in [38]–[40] utilizes feedback from the central node to refine messages in

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

The author was in part supported by the Grant-in-Aid for Scientific Research (B) (JSPS KAKENHI Grant Numbers 21H01326), Japan. The material in this paper was submitted in part to 2024 IEEE International Conference on Acoustics, Speech and Signal Processing.

each remote node, like distributed IHT [30], while distributed AMP in [41], [42] exploits no feedback from the central node. Hayakawa *et al*. [43] proposed decentralized AMP (D-AMP) for tree-structured networks with no central nodes via consensus propagation [44]. However, D-AMP was shown in numerical simulations [43] to have poor performance compared to that of centralized AMP [1], unless approximately perfect consensus is achieved before denoising in each AMP iteration. This convergence property of D-AMP is different from that of conventional decentralized algorithms [33]–[37] that realize average consensus and signal estimation simultaneously.

B. Contributions

This paper proposes decentralized GAMP (D-GAMP) for compressed sensing in tree-structured networks with no central nodes. Each node only utilizes local measurements to compute the GAMP iteration. Messages in each node are shared with adjacent nodes at every fixed time interval via consensus propagation [44]. D-GAMP repeats the local GAMP iteration and consensus propagation until the algorithm converges.

D-GAMP realizes average consensus and signal estimation simultaneously via consensus propagation between adjacent nodes, like conventional decentralized algorithms [33]–[37]. As a result, D-GAMP can reduce the total number of iterations for consensus propagation.

D-GAMP allows different nodes to use a different number of inner GAMP iterations. This flexibility is useful to reduce latency when different nodes have different processing capability. Waiting for processing in the other nodes can be a cause of latency. To circumvent this waiting issue, D-GAMP shares messages between adjacent nodes at every fixed time interval, rather than after a common number of GAMP iterations in all nodes.

The convergence property of Bayes-optimal D-GAMP is analyzed with a long-memory proof strategy [45]–[47], which is a general strategy for proving the convergence of state evolution recursion. Rigorous state evolution requires evaluation of the covariance matrix between estimation errors in all previous iterations. In the proof strategy, the covariance matrix is utilized to prove the convergence of state evolution recursion with respect to the mean-square errors (MSEs), i.e. its diagonal elements. When the Bayes-optimal denoisers are used in terms of MMSE, the covariance matrix has a special structure that implies the convergence for the sequence of its diagonal elements. See [48]–[53] for the original purpose of long-memory message-passing, i.e. complexity reduction.

The main contributions of this paper are as follows: A first contribution (Theorems 1, 2, and 6) is rigorous state evolution of D-GAMP. This paper proposes and analyzes a general error model that contains the error model of D-GAMP and is applicable to general ad hoc networks without tree structure. While D-GAMP assumes tree-structured networks in consensus propagation, as considered in [43], the state evolution result can be utilized to design another sophisticated protocols for average consensus in general ad hoc networks.

From a technical point of view, state evolution for GAMP [18], [25] is generalized to D-GAMP. The proof strategy in this paper is essentially different from in [18], [25]: Rangan [18] considered vector-valued AMP to analyze the asymptotic dynamics of GAMP. GAMP for rectangular sensing matrices was analyzed via state evolution of GAMP for symmetric sensing matrices in [25]. This paper establishes state evolution of D-GAMP for rectangular matrices directly by defining the general error model appropriately. In this sense, the definition of the general error model is a key contribution in the state evolution analysis.

A second contribution (Theorems 3 and 5) is the convergence analysis of D-GAMP for tree-structured networks. This paper proves that state evolution recursion for D-GAMP has the same fixed point as that of the corresponding centralized GAMP [18] when all nodes have homogeneous measurements with an identical dimension. On the basis of the long-memory proof strategy [46], the state evolution recursion for D-GAMP is proved to converge toward a fixed point when the Bayes-optimal inner and outer denoisers are used in terms of MMSE. These results imply that the state evolution for Bayes-optimal D-GAMP converges to the Bayes-optimal fixed point [26] for the homogeneous measurements when the Bayes-optimal fixed point is unique.

The last contribution is numerical results for D-GAMP. As examples of tree-structured networks, a one-dimensional chain and a tree with no central nodes are considered. For the linear measurements, D-GAMP is numerically shown to reduce the total number of inner iterations for consensus propagation compared to conventional D-AMP [43]. Furthermore, D-GAMP is shown to converge toward the performance of the corresponding centralized GAMP [18] for finite-sized measurements with clipping when homogeneous measurements with an identical dimension in all nodes are considered.

Part of these contributions were presented in [54].

C. Organization

The remainder of this paper is organized as follows: After summarizing the notation used in this paper, Section II formulates signal reconstruction from generalized linear measurements distributed in an ad hoc network without central nodes. The network is modeled as a directed and connected graph in graph theory. In particular, this paper focuses on a tree-structured network, i.e. an undirected and connected graph without cycles.

D-GAMP based on consensus propagation [44] is proposed in Section III. It is regarded as a generalization of D-AMP [43] to the generalized linear measurements. The proposed D-GAMP is more flexible in terms of the iteration schedule than D-AMP [43].

Section IV presents the main results of this paper while the proofs of theorems are summarized in Appendices. The asymptotic dynamics of D-GAMP is analyzed via state evolution. When a tree-structured network is assumed to justify use of consensus propagation, the long-memory proof strategy [46] is utilized to prove that state evolution recursion for Bayes-optimal D-GAMP converges toward a fixed point. In particular, the fixed point is equal to the Bayes-optimal fixed point [26] when the fixed point is unique.

Section V presents numerical results for D-GAMP. A one-dimensional chain and a tree with 8 nodes are considered as examples of tree-structured networks. Section VI concludes this paper.

D. Notation

Throughout this paper, the transpose and determinant of a matrix $\boldsymbol{M}$ are denoted by $\boldsymbol{M}^{\mathrm{T}}$ and $\det \boldsymbol{M}$, respectively. The notation $\boldsymbol{O}$ represents an all-zero matrix. The Kronecker delta is denoted by $\delta_{i,j}$. For $\{a_i\}_{i=1}^{n}$, the notation $\mathrm{diag}\{a_1, \ldots, a_n\}$ represents the diagonal matrix having the ith diagonal element a_i. The norm $\|\cdot\|$ means the Euclidean norm. The notation $\boldsymbol{o}(1)$ denotes a vector of which the Euclidean norm converges almost surely toward zero.

For a vector $\boldsymbol{v}_{\mathcal{I}}$ with a set of indices $\mathcal{I}$, the nth element $[\boldsymbol{v}_{\mathcal{I}}]_n$ of $\boldsymbol{v}_{\mathcal{I}}$ is written as $v_{n,\mathcal{I}}$. Similarly, the tth column of a matrix $\boldsymbol{M}_{\mathcal{I}}$ is represented as $\boldsymbol{m}_{t,\mathcal{I}}$.

The notation $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The almost sure convergence and equivalence are denoted by $\overset{\text{a.s.}}{\to}$ and $\overset{\text{a.s.}}{=}$, respectively.

For a scalar function $f : \mathbb{R} \to \mathbb{R}$ and a vector $\boldsymbol{x} \in \mathbb{R}^n$, the notation $f(\boldsymbol{x})$ means the element-wise application of f to $\boldsymbol{x}$, i.e. $[f(\boldsymbol{x})]_i = f(x_i)$. The arithmetic mean of $\boldsymbol{x} \in \mathbb{R}^n$ is written as $\langle \boldsymbol{x} \rangle = n^{-1} \sum_{i=1}^{n} x_i$. For a multi-variate function $f : \mathbb{R}^t \to \mathbb{R}$, the notation ∂_i represents the partial derivative of f with respect to the ith variable.

The notation $\boldsymbol{M}^{\dagger} = (\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})^{-1}\boldsymbol{M}^{\mathrm{T}}$ represents the pseudo-inverse of a full-rank matrix $\boldsymbol{M} \in \mathbb{R}^{m \times n}$ satisfying $m \geq n$. The matrix $\boldsymbol{P}_{M}^{\parallel} = \boldsymbol{M}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})^{-1}\boldsymbol{M}^{\mathrm{T}}$ is the projection onto the space spanned by the columns of $\boldsymbol{M}$ while $\boldsymbol{P}_{M}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{M}^{\parallel}$ is the projection onto the corresponding orthogonal complement.

II. Measurement Model

This paper considers the reconstruction of an unknown signal vector $\boldsymbol{x} \in \mathbb{R}^N$ from measurements in an ad hoc network with L nodes. While D-GAMP postulates tree-structured networks in consensus propagation [44], state evolution analysis is performed for general ad hoc networks. Thus, this section presents the definition of a general ad hoc network.

The ad hoc network is modeled as a directed and connected graph $\mathfrak{G} = (\mathcal{L}, \mathcal{E})$ with the set of nodes $\mathcal{L} = \{1, \ldots, L\}$, the set of edges $\mathcal{E} \subset \mathcal{L} \otimes \mathcal{L}$, and no self-loops. When the pair $(l_1, l_2) \in \mathcal{E}$ exists, there is an edge connected from node l_1 to node l_2. Since the graph has no self-loops, $(l, l) \notin \mathcal{E}$ holds for all $l \in \mathcal{L}$. The incoming neighborhood $\mathcal{N}[l] = \{l' \in \mathcal{L} : (l', l) \in \mathcal{E}\} \subset \mathcal{L}$ of node l represents the set of nodes that have incoming edges to node l while the outgoing neighborhood $\tilde{\mathcal{N}}[l] = \{l' \in \mathcal{L} : (l, l') \in \mathcal{E}\}$ is the set of nodes that have outgoing edges from node l. Since the graph has no self-loops, we have $l \notin \mathcal{N}[l]$ and $l \notin \tilde{\mathcal{N}}[l]$ for all $l \in \mathcal{L}$. A central node $l \in \mathcal{L}$ is defined as a node that is connected to all other nodes, i.e. $\mathcal{N}[l] = \tilde{\mathcal{N}}[l] = \mathcal{L} \backslash \{l\}$. Throughout this paper, the existence of central nodes is not assumed.

Node l acquires an $M[l]$-dimensional measurement vector $\boldsymbol{y}[l] \in \mathbb{R}^{M[l]}$, modeled as the generalized linear measurements

$$\boldsymbol{y}[l] = g[l](\boldsymbol{z}[l], \boldsymbol{w}[l]), \quad \boldsymbol{z}[l] = \boldsymbol{A}[l]\boldsymbol{x}. \tag{1}$$

In (1), $\boldsymbol{w}[l] \in \mathbb{R}^{M[l]}$ and $\boldsymbol{A}[l] \in \mathbb{R}^{M[l] \times N}$ denote an unknown noise vector and a sensing matrix in node l, respectively. The signal vector $\boldsymbol{x}$ is measured via the linear mapping $\boldsymbol{z}[l]$. The measurement vector $\boldsymbol{y}[l]$ is an element-wise application of a function $g[l] : \mathbb{R}^2 \to \mathbb{R}$ to the two vectors $\boldsymbol{z}[l]$ and $\boldsymbol{w}[l]$. In particular, $g[l](z, w) = z + w$ corresponds to conventional linear measurements.

The goal of this paper is to reconstruct the signal vector $\boldsymbol{x}$ under the following assumptions:

- Node l only has the information about the local measurement vector $\boldsymbol{y}[l]$ and sensing matrix $\boldsymbol{A}[l]$, as well as the measurement model (1).
- Node l can send messages to the outgoing neighborhood $\tilde{\mathcal{N}}[l]$ and receive them from the incoming neighborhood $\mathcal{N}[l]$.
- The communication link between nodes is error-free and latency-free.

The first two assumptions are practical assumptions for decentralized algorithms in ad hoc networks. The last assumption may be reasonable for reliable wired networks or future wireless networks.

III. Distributed GAMP

A. Overview

The proposed D-GAMP algorithm consists of two parts: GAMP iteration [18] in each node and consensus propagation [44] between nodes. GAMP in node $l \in \mathcal{L}$ is composed of two modules, called outer and inner modules[1]. The outer module utilizes the measurement vector $\boldsymbol{y}[l]$ to compute an estimator of $\boldsymbol{z}[l]$ in (1) while the inner module uses prior information on the signal vector $\boldsymbol{x}$ to compute an estimator of $\boldsymbol{x}$.

For consensus propagation all nodes share messages with their adjacent nodes. Each node utilizes the messages sent from its adjacent nodes to update the current message. Message transmission for consensus propagation is repeated J times. While consensus propagation requires tree-structured networks, D-GAMP will be applicable to general ad hoc networks when consensus propagation is replaced with another sophisticated protocol for average consensus. However, research in this direction is beyond the scope of this paper.

The outer module, consensus propagation, and the inner module are executed in this order. After that, node l repeats $T[l] - 1$ iterations between the outer and inner modules in a fixed time interval, without performing consensus propagation. Different $T[l]$ is used for different l since the nodes might not have identical processing capability. In other words, $T[l]$ corresponds to the number of iterations which node l can repeat in the fixed time interval. After the fixed time interval, the outer module, consensus propagation, and the inner module are executed again. Then, each node repeats $T[l] - 1$ GAMP

[1]In GAMP [18], the $M[l]$-dimensional measurement space was referred to an output space while the N-dimensional signal space was called an input space. In this paper, they are referred to as outer and inner spaces, respectively. The terminology "inner module" does not mean that the inner module is physically located inside the outer module.

iterations to refine the estimation of the signal vector. Such iteration rounds are repeated until the consensus is achieved.

Let $T = \max_{l\in\mathcal{L}} T[l]$ denote the maximum number of GAMP iterations among all nodes. For notational convenience, we define a message $a_t[l]$ in iteration t after every $T[l]$ GAMP iterations for node l as that in the corresponding $T[l]$th iteration, i.e. $a_t[l] = a_{T[l]-1}[l]$ for $t \in \{T[l], \ldots, T-1\}$. This notation allows us to use the common number of GAMP iterations T in all nodes. When the total number of GAMP iterations t is counted, consensus propagation is performed in iteration $t = iT$ for all non-negative integers i.

To represent messages in the iteration where consensus propagation is performed, the following underline notation is used throughout this paper:

Definition 1: For integers $t, T \in \mathbb{N}$, and variables $\{a_\tau \in \mathbb{R}\}$, the notation $\underline{a}_t$ is defined as $\underline{a}_t = a_{iT}$ with $i = \lfloor t/T \rfloor$. For vectors $\{\boldsymbol{v}_\tau\}$ the notation $\underline{\boldsymbol{v}}_t = \boldsymbol{v}_{iT}$ is defined in the same manner while the notation $\underline{\boldsymbol{M}}_t = [\underline{\boldsymbol{m}}_0, \ldots, \underline{\boldsymbol{m}}_{t-1}]$ is used for a matrix $\boldsymbol{M}_t = [\boldsymbol{m}_0, \ldots, \boldsymbol{m}_{t-1}]$.

B. Outer Module

For iteration $t \in \{0, 1, \ldots\}$, suppose that the outer module has an estimator $\tilde{\boldsymbol{z}}_t[l] \in \mathbb{R}^{M[l]}$ of $\boldsymbol{z}[l]$, an estimator $v_t[l] > 0$ of $M^{-1}[l]\|\tilde{\boldsymbol{z}}_t[l] - \boldsymbol{z}[l]\|^2$, and an estimator $\hat{\boldsymbol{x}}_t[l] \in \mathbb{R}^N$ of $\boldsymbol{x}$ sent from the inner module, as well as the measurement vector $\boldsymbol{y}[l]$. As initial conditions, $\tilde{\boldsymbol{z}}_0[l] = \boldsymbol{0}$, $v_0[l] = (LM[l])^{-1}\mathbb{E}[\|\boldsymbol{x}\|^2]$, and $\hat{\boldsymbol{x}}_0[l] = \boldsymbol{0}$ are used for all $l \in \mathcal{L}$.

The outer module computes an estimator $\hat{\boldsymbol{z}}_t[l] \in \mathbb{R}^{M[l]}$ of $\boldsymbol{z}[l]$ and a message $\boldsymbol{x}_t[l] \in \mathbb{R}^N$ as follows:

$$\hat{\boldsymbol{z}}_t[l] = f_{\mathrm{out}}[l](\tilde{\boldsymbol{z}}_t[l], \boldsymbol{y}[l]; v_t[l]), \tag{2}$$

$$\boldsymbol{x}_t[l] = \frac{1}{L}\hat{\boldsymbol{x}}_t[l] - \frac{1}{\xi_{\mathrm{out},t}[l]}\boldsymbol{A}^{\mathrm{T}}[l]\hat{\boldsymbol{z}}_t[l], \tag{3}$$

with

$$\xi_{\mathrm{out},t}[l] = \langle \partial_1 f_{\mathrm{out}}[l](\tilde{\boldsymbol{z}}_t[l], \boldsymbol{y}[l]; v_t[l]) \rangle. \tag{4}$$

Here, the scalar function $f_{\mathrm{out}}[l](\cdot, \cdot; v_t[l]) : \mathbb{R}^2 \to \mathbb{R}$ is an outer denoiser. The notation ∂_1 denotes the partial derivative of $f_{\mathrm{out}}[l]$ with respect to the first variable. The parameter $\xi_{\mathrm{out},t}[l] \in \mathbb{R}$ has been designed so as to realize asymptotic Gaussianity of estimation errors before inner denoising. The outer module sends the messages $\boldsymbol{x}_t[l]$, $\hat{\boldsymbol{z}}_t[l]$, and $\xi_{\mathrm{out},t}[l]$ to the inner module.

The update rules in the outer module are similar to those in centralized GAMP [18]. Intuitively, the Onsager correction in (3) eliminates intractable memory terms in each iteration. Since clean messages after the Onsager correction are shared with adjacent nodes, consensus propagation does not affect the update rules in the outer module. The correctness of this intuition is proved via state evolution.

When the remainder of t divided by T is larger than or equal to the actual number of iterations $T[l]$, all messages are fixed to $\hat{\boldsymbol{z}}_t[l] = \hat{\boldsymbol{z}}_{iT+T[l]-1}[l]$, $\boldsymbol{x}_t[l] = \boldsymbol{x}_{iT+T[l]-1}[l]$, and $\xi_{\mathrm{out},t}[l] = \xi_{\mathrm{out},iT+T[l]-1}[l]$ for $i = \lfloor t/T \rfloor$. Thus, we define the set of iteration indices $\mathcal{T}_t[l]$ as $\mathcal{T}_0[l] = \emptyset$ and

$$\mathcal{T}_t[l] = \cup_{i=0}^{\lfloor t/T \rfloor}\{iT, \ldots, \min\{t, iT + T[l]\} - 1\} \tag{5}$$

for $t > 0$, by eliminating from $\{0, \ldots, t-1\}$ the indices for which the messages are fixed. For iteration t, the set $\mathcal{T}_t[l]$ contains all iterations in which the messages in node l are updated.

C. Consensus Propagation

The centralized GAMP [18] uses the messages $\tilde{\boldsymbol{x}}_t = \sum_{l\in\mathcal{L}} \boldsymbol{x}_t[l]$, $\eta_t = L$, and $\sigma_t^2 = L^{-1}\sum_{l\in\mathcal{L}} \xi_{\mathrm{out},t}^{-1}[l]$ in the inner module. However, computation of these messages requires a central node that receives the messages $\{\boldsymbol{x}_t[l] : l \in \mathcal{L}\}$ and $\{\xi_{\mathrm{out},t}[l] : l \in \mathcal{L}\}$ from all nodes. For a tree-structured network with no central nodes, consensus propagation [44] can be used to compute the messages $\tilde{\boldsymbol{x}}_t$, η_t, and σ_t^2 in a decentralized manner.

Consensus propagation with J inner iterations is performed in every T iterations. More precisely, node l shares messages with the adjacent nodes $\tilde{\mathcal{N}}[l]$ for consensus propagation if t is divisible by T.

Focus on inner iteration $j \in \{1, \ldots, J\}$ and suppose that node l has messages $\{\underline{\boldsymbol{x}}_{t,j-1}[l' \to l] \in \mathbb{R}^N : l' \in \mathcal{N}[l]\}$ and $\{\underline{\sigma}_{t,j-1}^2[l' \to l] \in \mathbb{R} : l' \in \mathcal{N}[l]\}$ sent from the adjacent nodes in the preceding inner iteration of consensus propagation and messages $\{\underline{\eta}_{t,j-1}[l' \to l] \in \mathbb{R} : l' \in \mathcal{N}[l]\}$ computed in node l, as well as the messages $\underline{\boldsymbol{x}}_t[l]$ and $\underline{\xi}_{\mathrm{out},t}[l]$ computed in the outer module. Node l computes the following messages $\underline{\boldsymbol{x}}_{t,j}[l \to l']$ and $\underline{\sigma}_{t,j}^2[l \to l']$, which are sent to node $l' \in \tilde{\mathcal{N}}[l]$, as well as $\underline{\eta}_{t,j}[l \to l']$.

$$\underline{\boldsymbol{x}}_{t,j}[l \to l'] = \underline{\boldsymbol{x}}_t[l] + \sum_{\tilde{l}' \in \mathcal{N}[l]\setminus\{l'\}} \underline{\boldsymbol{x}}_{t,j-1}[\tilde{l}' \to l], \tag{6}$$

$$\underline{\eta}_{t,j}[l \to l'] = 1 + \sum_{\tilde{l}' \in \mathcal{N}[l]\setminus\{l'\}} \underline{\eta}_{t,j-1}[\tilde{l}' \to l], \tag{7}$$

$$\underline{\sigma}_{t,j}^2[l \to l'] = \frac{1}{\underline{\xi}_{\mathrm{out},t}[l]} + \sum_{\tilde{l}' \in \mathcal{N}[l]\setminus\{l'\}} \underline{\sigma}_{t,j-1}^2[\tilde{l}' \to l], \tag{8}$$

with $\underline{\boldsymbol{x}}_{t,0}[l' \to l] = \underline{\boldsymbol{x}}_{t-T,J}[l' \to l]$, $\underline{\eta}_{t,0}[l' \to l] = \underline{\eta}_{t-T,J}[l' \to l]$, and $\underline{\sigma}_{t,0}^2[l' \to l] = \underline{\sigma}_{t-T,J}^2[l' \to l]$, which are the messages sent from the adjacent nodes in the preceding round of consensus propagation. As initial conditions, $\underline{\boldsymbol{x}}_{t-T,J}[l' \to l] = \boldsymbol{0}$, $\underline{\eta}_{t-T,J}[l' \to l] = 0$, and $\underline{\sigma}_{t-T,J}^2[l' \to l] = 0$ are used for all $t < T$.

After J inner iterations for consensus propagation, node l receives the messages $\{\underline{\boldsymbol{x}}_{t,J}[l' \to l] \in \mathbb{R}^N : l' \in \mathcal{N}[l]\}$ and $\{\underline{\sigma}_{t,J}^2[l' \to l] > 0 : l' \in \mathcal{N}[l]\}$ from the adjacent nodes. Then, the following messages $\overline{\underline{\boldsymbol{x}}_t[l]}$, $\overline{\underline{\eta}_t[l]}$, and $\overline{\underline{\sigma}_t^2[l]}$ are computed and sent to the inner module:

$$\overline{\underline{\boldsymbol{x}}_t[l]} = \sum_{l' \in \mathcal{N}[l]} \underline{\boldsymbol{x}}_{t,J}[l' \to l], \tag{9}$$

$$\overline{\underline{\eta}_t[l]} = \sum_{l' \in \mathcal{N}[l]} \underline{\eta}_{t,J}[l' \to l], \tag{10}$$

$$\overline{\underline{\sigma}_t^2[l]} = \sum_{l' \in \mathcal{N}[l]} \underline{\sigma}_{t,J}^2[l' \to l]. \tag{11}$$

The message $\overline{\underline{\boldsymbol{x}}_t[l]}$ is used in the inner module as an extrinsic estimate of $\boldsymbol{x}$ while $\overline{\underline{\sigma}_t^2[l]}$ is an estimator of the corresponding extrinsic MSE. After the aggregation of all messages, the messages $\overline{\underline{\boldsymbol{x}}_t[l]}$, $\overline{\underline{\eta}_t[l]}$, and $\overline{\underline{\sigma}_t^2[l]}$ converge to $\sum_{l'\neq l}\underline{\boldsymbol{x}}_t[l']$, $L-1$, and $\sum_{l'\neq l}\underline{\xi}_{\mathrm{out},t}^{-1}[l']$ for tree-structured networks, respectively.

D. Inner Module

In iteration t, suppose that the inner module has the messages $\boldsymbol{x}_t[l]$, $\hat{\boldsymbol{z}}_t[l]$, and $\xi_{\mathrm{out},t}[l]$ sent from the outer module and the messages $\overline{\underline{\boldsymbol{x}}_t[l]}$, $\overline{\underline{\eta}_t[l]}$, and $\overline{\underline{\sigma}_t^2[l]}$ computed in consensus propagation. The inner module first computes the following messages:

$$\tilde{\boldsymbol{x}}_t[l] = \boldsymbol{x}_t[l] + \overline{\underline{\boldsymbol{x}}_t[l]}, \tag{12}$$

$$\eta_t[l] = 1 + \overline{\underline{\eta}_t[l]}, \tag{13}$$

$$\sigma_t^2[l] = \frac{1}{L}\left(\frac{1}{\xi_{\mathrm{out},t}[l]} + \overline{\underline{\sigma}_t^2[l]}\right), \tag{14}$$

which should converge to $\tilde{\boldsymbol{x}}_t = \sum_{l'\in\mathcal{L}}\boldsymbol{x}_t[l']$, L, and $L^{-1}\sum_{l'\in\mathcal{L}}\xi_{\mathrm{out},t}^{-1}[l']$ for tree-structured networks after the aggregation of all messages, respectively, when t is divisible by T. Then, an estimator $\hat{\boldsymbol{x}}_{t+1}[l]\in\mathbb{R}^N$ of the signal vector $\boldsymbol{x}$ is computed as

$$\hat{\boldsymbol{x}}_{t+1}[l] = f_{\mathrm{in}}[l](\tilde{\boldsymbol{x}}_t[l];\eta_t[l],\sigma_t^2[l]), \tag{15}$$

where $f_{\mathrm{in}}[l](\cdot;\eta_t[l],\sigma_t^2[l]):\mathbb{R}\to\mathbb{R}$ denotes an inner denoiser. The parameter $\eta_t[l]$ denotes the number of messages aggregated in consensus propagation. The parameter $\sigma_t^2[l]$ corresponds to an estimator of the MSE for the message $\tilde{\boldsymbol{x}}_t[l]$. See Section IV for its precise meaning revealed via state evolution.

The estimator $\hat{\boldsymbol{x}}_{t+1}[l]$ depends on the node index l while the original signal vector $\boldsymbol{x}$ is independent of l. When there are no central nodes for aggregating $\{\tilde{\boldsymbol{x}}_t[l]: l\in\mathcal{L}\}$, the estimator $\hat{\boldsymbol{x}}_{t+1}[l]$ can be used when an estimator of $\boldsymbol{x}$ is needed in node l.

To refine the estimator of $\boldsymbol{x}$, the inner module computes the following messages:

$$\tilde{\boldsymbol{z}}_{t+1}[l] = \boldsymbol{A}[l]\hat{\boldsymbol{x}}_{t+1}[l] + \frac{N\xi_{\mathrm{in},t}[l]}{LM[l]\xi_{\mathrm{out},t}[l]}\hat{\boldsymbol{z}}_t[l], \tag{16}$$

$$v_{t+1}[l] = \frac{N}{M[l]}\frac{\sigma_t^2[l]\xi_{\mathrm{in},t}[l]}{\eta_t[l]}, \tag{17}$$

with

$$\xi_{\mathrm{in},t}[l] = \langle\partial_1 f_{\mathrm{in}}[l](\tilde{\boldsymbol{x}}_t[l];\eta_t[l],\sigma_t^2[l])\rangle. \tag{18}$$

The message $\tilde{\boldsymbol{z}}_t[l]$ is an estimator of $\boldsymbol{z}[l]$. The message $\xi_{\mathrm{in},t}[l]\in\mathbb{R}$ has been designed so as to realize asymptotic Gaussianity of estimation errors before outer denoising. The message $v_t[l]$ corresponds to an estimator of the MSE for $\tilde{\boldsymbol{z}}_t[l]$. See Section IV for its precise meaning revealed via state evolution.

For $t-\lfloor t/T\rfloor T\geq T[l]$, all messages are fixed to $\hat{\boldsymbol{x}}_{t+1}[l]=\hat{\boldsymbol{x}}_{iT+T[l]}[l]$, $\tilde{\boldsymbol{z}}_{t+1}[l]=\tilde{\boldsymbol{z}}_{iT+T[l]}[l]$, and $v_{t+1}[l]=v_{iT+T[l]}[l]$, as fixed in the outer module. The inner module feeds the messages $\hat{\boldsymbol{x}}_{t+1}[l]$, $\tilde{\boldsymbol{z}}_{t+1}[l]$, and $v_{t+1}[l]$ back to the outer module to refine the estimator of $\boldsymbol{x}$.

To understand D-GAMP, assume that the summation consensus $\overline{\underline{\boldsymbol{x}}_t[l]}=\sum_{l'\neq l}\underline{\boldsymbol{x}}_t[l']$, $\hat{\underline{\boldsymbol{x}}}_t[l']=\hat{\underline{\boldsymbol{x}}}_t[l]$, and $\underline{\xi}_{\mathrm{out},t}[l']=\underline{\xi}_{\mathrm{out},t}[l]$ have been achieved for all $l,l'\in\mathcal{L}$ when t is sufficiently large. Then, (12) reduces to $\tilde{\boldsymbol{x}}_t[l]=\boldsymbol{x}_t[l]+\sum_{l'\neq l}\underline{\boldsymbol{x}}_t[l']$. Using the definition (3) of $\boldsymbol{x}_t[l]$ yields

$$\tilde{\boldsymbol{x}}_{iT}[l] = \hat{\boldsymbol{x}}_{iT}[l] + \frac{1}{\xi_{\mathrm{out},iT}[l]}\sum_{l'\in\mathcal{L}}\boldsymbol{A}^{\mathrm{T}}[l']\hat{\boldsymbol{z}}_{iT}[l'] \tag{19}$$

for $t=iT\in\mathbb{N}$. The update rules (16) and (19) for $\tilde{\boldsymbol{z}}_t[l]$ and $\tilde{\boldsymbol{x}}_{iT}[l]$ are equivalent to those in centralized GAMP [18].

Conventional D-AMP [43] was designed under the implicit assumption of perfect consensus in each iteration. As a result, multiple inner iterations J for consensus propagation were considered to realize the summation consensus $\overline{\underline{\boldsymbol{x}}_t[l]}=\sum_{l'\in\mathcal{L}}\underline{\boldsymbol{x}}_t[l']$ for all $l\in\mathcal{L}$ approximately. However, such a protocol requires heavy communications between adjacent nodes.

D-GAMP with $T=1$ is equivalent to D-AMP [43] when the linear measurement model $g[l](z,w)=z+w$ is considered. Interestingly, state evolution in this paper reveals the correctness of the Onsager correction in D-GAMP. As a result, D-AMP [43] also has the correct Onsager correction while perfect consensus was implicitly assumed in designing D-AMP. Nonetheless, numerical simulations in [43] showed poor performance of D-AMP with a few inner iterations of consensus propagation. This conflict between theoretical and numerical results may be because small $M[l]=6$ was simulated in [43]. If much larger systems were simulated, good performance might be observed.

IV. Main Results

A. Definitions and Assumptions

The dynamics of D-GAMP is analyzed via state evolution in the large system limit for fixed L, where the dimensions $\{M[l]\}$ and N tend to infinity while the ratio $\delta[l]=M[l]/N$ is kept constant for all $l\in\mathcal{L}$. To present a rigorous result, we first define an empirical convergence in terms of separable, pseudo-Lipschitz, and proper functions [12], [55].

Definition 2 (Separability): A vector-valued function $\boldsymbol{f}=(f_1,\ldots,f_N)^{\mathrm{T}}$ with $f_n:\mathbb{R}^N\to\mathbb{R}$ is said to be separable if the nth function $f_n(\boldsymbol{x})$ depends only on the nth element of $\boldsymbol{x}=(x_1,\ldots,x_N)^{\mathrm{T}}\in\mathbb{R}^N$ for all n, i.e. $f_n(\boldsymbol{x})=f_n(x_n)$.

Definition 3 (Pseudo-Lipschitz): A function $f:\mathbb{R}^t\to\mathbb{R}$ is said to be pseudo-Lipschitz of order k if there is some Lipschitz constant $C>0$ such that the following inequality holds:

$$|f(\boldsymbol{x})-f(\boldsymbol{y})|\leq C\|\boldsymbol{x}-\boldsymbol{y}\|(1+|f(\boldsymbol{x})|^{k-1}+|f(\boldsymbol{y})|^{k-1}) \tag{20}$$

for all $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^t$.

By definition, any pseudo-Lipschitz function of order $k=1$ is Lipschitz-continuous. A separable vector-valued function $\boldsymbol{f}$ is said to be pseudo-Lipschitz of order k if all element functions of $\boldsymbol{f}$ are pseudo-Lipschitz of order k. In this paper, piecewise pseudo-Lipschitz functions are considered to include practical denoisers in the proposed framework of state evolution.

Definition 4 (Proper): A separable, pseudo-Lipschitz, and vector-valued function $\boldsymbol{f} = (f_1, \ldots, f_N)^{\mathrm{T}}$ is said to be proper if the Lipschitz constant $C_n > 0$ of the nth function $f_n : \mathbb{R} \to \mathbb{R}$ satisfies

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} C_n^k < \infty \quad \text{for all } k \in \mathbb{N}. \tag{21}$$

Definition 4 is used to analyze separable, pseudo-Lipschitz, and vector-valued functions $\boldsymbol{f}$ in the same manner as in the common function case [55]: $f_n = f$ for a pseudo-Lipschitz function f.

Definition 5: Random vectors $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t) \in \mathbb{R}^{N\times t}$ are said to converge jointly toward random variables $(V_1, \ldots, V_t)$ in the sense of kth-order pseudo-Lipschitz if the limit $\lim_{N\to\infty} N^{-1} \sum_{n=1}^{N} \mathbb{E}[f_n(V_1, \ldots, V_t)]$ exists and the following almost sure convergence holds:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \{f_n(v_{n,1}, \ldots, v_{n,t}) - \mathbb{E}[f_n(V_1, \ldots, V_t)]\} \overset{\text{a.s.}}{\to} 0 \tag{22}$$

for all separable and piecewise proper pseudo-Lipschitz functions $\boldsymbol{f} = [f_1, \ldots, f_N]^{\mathrm{T}}$ of order k. This convergence in the sense of kth-order pseudo-Lipschitz is called the $\mathrm{PL}(k)$ convergence and denoted by $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t) \overset{\mathrm{PL}(k)}{\to} (V_1, \ldots, V_t)$.

The goal of state evolution is to prove asymptotic Gaussianity for the messages $\tilde{\boldsymbol{z}}_t[l]$ and $\tilde{\boldsymbol{x}}_t[l]$ just before outer and inner denoising, respectively: The PL(2) convergence results $\tilde{\boldsymbol{z}}_t[l] \overset{\mathrm{PL}(2)}{\to} Z_t[l]$ and $\tilde{\boldsymbol{x}}_t[l] - L^{-1}\bar{\eta}_t[l]\boldsymbol{x} \overset{\mathrm{PL}(2)}{\to} H_t[l]$ hold in the large system limit for some Gaussian random variables $Z_t[l]$, $H_t[l]$, and deterministic variable $\bar{\eta}_t[l]$ defined shortly. We summarize assumptions to justify the PL(2) convergence.

Assumption 1: For some $\epsilon > 0$, the $\mathrm{PL}(2+\epsilon)$ convergence holds for the signal vector $\boldsymbol{x}$, i.e. $\boldsymbol{x} \overset{\mathrm{PL}(2+\epsilon)}{\to} X$ for some random variable X.

When $\boldsymbol{x}$ has i.i.d. elements with a bounded $(2+\epsilon)$th moment, the $\mathrm{PL}(2+\epsilon)$ convergence holds for X that follows the distribution for the elements of $\boldsymbol{x}$.

Assumption 2: For some $\epsilon > 0$, the $\mathrm{PL}(2+\epsilon)$ convergence holds for the noise vectors $\{\boldsymbol{w}[l] : l \in \mathcal{L}\}$ i.e. $\{\boldsymbol{w}[l]\} \overset{\mathrm{PL}(2+\epsilon)}{\to} \{W[l]\}$ for some random variables $\{W[l] : l \in \mathcal{L}\}$.

Independent Gaussian noise $\boldsymbol{w}[l] \sim \mathcal{N}(\boldsymbol{0}, \sigma^2[l]\boldsymbol{I}_{M[l]})$ with some variance $\sigma^2[l]$ satisfies Assumption 2 for independent Gaussian random variables $W[l] \sim \mathcal{N}(0, \sigma^2[l])$.

Assumption 3: The sensing matrices $\{\boldsymbol{A}[l] : l \in \mathcal{L}\}$ are independent. Each matrix $\boldsymbol{A}[l] \in \mathbb{R}^{M[l]\times N}$ has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$.

Assumption 3is an important assumption to analyze the dynamics of D-GAMP via state evolution. The independent assumption cannot be relaxed for AMP [14]. More precisely, the empirical eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}[l]\boldsymbol{A}[l]$ needs to converge in probability to that for zero-mean i.i.d. Gaussian sensing matrices.

Assumption 4: The composition $f_{\mathrm{out}}[l](\theta, g[l](z, w); v_t[l])$ of the measurement function $g[l]$ in (1) and outer denoiser $f_{\mathrm{out}}[l]$ in (2) is piecewise Lipschitz-continuous with respect to $(\theta, z, w) \in \mathbb{R}^3$. The inner denoiser $f_{\mathrm{in}}[l](u; \eta_t[l], \sigma_t^2[l])$ in (15) is piecewise Lipschitz-continuous with respect to $u \in \mathbb{R}$.

The everywhere Lipschitz-continuity was assumed in conventional state evolution analysis [12], [55]. Nonetheless, this paper postulates the piecewise Lipschitz-continuity to include practical outer denoisers in the proposed framework of state evolution. This slight generalization does not cause any gaps in state evolution analysis since any piecewise Lipschitz-continuous function has all properties required in state evolution, such as almost everywhere differentiability and the boundedness of derivatives—satisfied for all Lipschitz-continuous functions. Intuitively, there are no technically significant differences between the singularities at the origin of the two functions e.g. $f_1(x) = |x|$ and $f_2(x) = -x$ for all $x < 0$ and $f_2(x) = x + 1$ for all $x \geq 0$ unless $x = 0$ occurs with a finite probability.

Assumption 5: The graph $\mathfrak{G} = (\mathcal{L}, \mathcal{E})$ is a tree, i.e. an undirected and connected graph with no cycles.

Assumption 5 is used in justifying consensus propagation while it is not required in proving the asymptotic Gaussianity. Note that the incoming neighborhood $\mathcal{N}[l]$ is equal to the outgoing neighborhood $\tilde{\mathcal{N}}[l]$ under Assumption 5.

B. State Evolution

State evolution recursion for D-GAMP is given via four kinds of scalar zero-mean Gaussian random variables $\{Z[l]\}$, $\{Z_t[l]\}$, $\{H_t[l]\}$, and $\{\tilde{H}_t[l]\}$, associated with $\boldsymbol{z}[l]$, $\tilde{\boldsymbol{z}}_t[l]$, $\boldsymbol{x}_t[l]$, and $\tilde{\boldsymbol{x}}_t[l]$ in (1), (16), (3), and (12), respectively. The random variables $\{Z[l] : l \in \mathcal{L}\}$ are independent of $\{W[l]\}$ in Assumption 2 and independent zero-mean Gaussian random variables with variance

$$\mathbb{E}[Z^2[l]] = \frac{1}{L\delta[l]} \mathbb{E}[X^2], \tag{23}$$

with X defined in Assumption 1. To define statistical properties of the other random variables, we first define two variables $\bar{\xi}_{\mathrm{out},t}[l]$ and $\bar{\zeta}_t[l]$ in the outer module as

$$\bar{\xi}_{\mathrm{out},t}[l] = \mathbb{E}\left[\partial_1 f_{\mathrm{out}}[l](Z_t[l], Y[l]; \bar{v}_t[l])\right], \tag{24}$$

$$\bar{\zeta}_t[l] = -\mathbb{E}\left[\frac{\partial}{\partial z} f_{\mathrm{out}}[l](Z_t[l], g[l](z, W[l]); \bar{v}_t[l]) \Big|_{z=Z[l]} \right], \tag{25}$$

with $Y[l] = g[l](Z[l], W[l])$, in which $Z_t[l]$ and $\bar{v}_t[l]$ are defined shortly. The variable $\bar{\xi}_{\mathrm{out},t}[l]$ is the asymptotic alternative of $\xi_{\mathrm{out},t}[l]$ in (4) while $\bar{\zeta}_t[l]$ is used in the inner module.

We define random variables $\{H_t[l]\}$ and $\{\tilde{H}_t[l]\}$. The random variables $\{H_t[l]\}$ are independent of X in Assumption 1 and zero-mean Gaussian random variables with covariance

$$\mathbb{E}[H_\tau[l']H_t[l]] = \frac{\delta_{l,l'}}{L} \mathbb{E}\left[f_{\mathrm{out}}[l](Z_\tau[l], Y[l]; \bar{v}_\tau[l]) f_{\mathrm{out}}[l](Z_t[l], Y[l]; \bar{v}_t[l])\right] \tag{26}$$

for all $\tau \in \{0, \ldots, t\}$. Furthermore, $\tilde{H}_t[l]$ is defined recursively as follows:

$$\tilde{H}_t[l] = \frac{H_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]} + \sum_{l'\in\mathcal{N}[l]} \underline{H}_{t,J}[l' \to l], \tag{27}$$

with

$$\underline{H}_{t,j}[l \to l'] = \frac{\underline{H}_t[l]}{\underline{\bar{\xi}}_{\mathrm{out},t}[l]} + \sum_{\tilde{l}'\in\mathcal{N}[l]\setminus\{l'\}} \underline{H}_{t,j-1}[\tilde{l}' \to l] \tag{28}$$

for $j \in \{1, \ldots, J\}$. As an initial condition, $\underline{H}_{t,0}[l' \to l] = \underline{H}_{t-T,J}[l' \to l]$ is used, as well as $\underline{H}_{t-T,J}[l' \to l] = 0$ for all $t < T$. The random variables $\{\tilde{H}_t[l]\}$ are zero-mean Gaussian random variables since they are a linear combination of $\{H_t[l]\}$.

We next define four variables $\bar{\eta}_t[l]$, $\bar{\sigma}_t^2[l]$, $\bar{\xi}_{\mathrm{in},t}[l]$, and $\bar{v}_{t+1}[l]$ in the inner module. The variable $\bar{\eta}_t[l]$ corresponds to the effective amplitude of X, given by

$$\bar{\eta}_t[l] = \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]} + \sum_{l' \in \mathcal{N}[l]} \underline{\bar{\eta}}_{t,J}[l' \to l], \tag{29}$$

with

$$\underline{\bar{\eta}}_{t,j}[l \to l'] = \frac{\underline{\bar{\zeta}}_t[l]}{\underline{\bar{\xi}}_{\mathrm{out},t}[l]} + \sum_{\tilde{l}' \in \mathcal{N}[l] \setminus \{l'\}} \underline{\bar{\eta}}_{t,j-1}[\tilde{l}' \to l]. \tag{30}$$

As an initial condition, $\underline{\bar{\eta}}_{t,0}[l' \to l] = \underline{\bar{\eta}}_{t-T,J}[l' \to l]$ is used, as well as $\underline{\bar{\eta}}_{t-T,J}[l' \to l] = 0$ for all $t < T$. Similarly, $\bar{\sigma}_t^2[l]$ represents the asymptotic alternative of $\sigma_t^2[l]$ in (14),

$$\bar{\sigma}_t^2[l] = \frac{1}{L} \left(\frac{1}{\bar{\xi}_{\mathrm{out},t}[l]} + \sum_{l' \in \mathcal{N}[l]} \underline{\bar{\sigma}}_{t,J}^2[l' \to l] \right), \tag{31}$$

with

$$\underline{\bar{\sigma}}_{t,j}^2[l \to l'] = \frac{1}{\underline{\bar{\xi}}_{\mathrm{out},t}[l]} + \sum_{\tilde{l}' \in \mathcal{N}[l] \setminus \{l'\}} \underline{\bar{\sigma}}_{t,j-1}^2[\tilde{l}' \to l]. \tag{32}$$

As an initial condition, $\underline{\bar{\sigma}}_{t,0}^2[l' \to l] = \underline{\bar{\sigma}}_{t-T,J}^2[l' \to l]$ is used, as well as $\underline{\bar{\sigma}}_{t-T,J}^2[l' \to l] = 0$ for all $t < T$. The variable $\bar{\xi}_{\mathrm{in},t}[l]$ is the asymptotic alternative of $\xi_{\mathrm{in},t}[l]$ in (18), given by

$$\bar{\xi}_{\mathrm{in},t}[l] = \mathbb{E}\left[\partial_1 f_{\mathrm{in}}[l] \left(\frac{\bar{\eta}_t[l]}{L} X + \tilde{H}_t[l]; \bar{\eta}_t[l], \bar{\sigma}_t^2[l] \right) \right]. \tag{33}$$

The variable $\bar{v}_{t+1}[l]$ is the asymptotic alternative of $v_{t+1}[l]$ in (17), given by $\bar{v}_0[l] = (L\delta[l])^{-1}\mathbb{E}[X^2]$ and

$$\bar{v}_{t+1}[l] = \frac{\bar{\sigma}_t^2[l]\bar{\xi}_{\mathrm{in},t}[l]}{\delta[l]\eta_t[l]} \tag{34}$$

for $t \geq 0$, where $\eta_t[l]$ is defined in (13).

Finally, we define $Z_0[l] = 0$ and the random variables $\{Z_{t+1}[l]\}$, which are independent of $\{W[l]\}$ and correlated with $\{Z[l]\}$. More precisely, $\{Z_{t+1}[l]\}$ are zero-mean Gaussian random variables with covariance

$$\mathbb{E}[Z[l']Z_{t+1}[l]] = \frac{\delta_{l,l'}}{L\delta[l]} \mathbb{E}\left[X f_{\mathrm{in}}[l] \left(\frac{\bar{\eta}_t[l]}{L} X + \tilde{H}_t[l]; \eta_t[l], \bar{\sigma}_t^2[l] \right) \right], \tag{35}$$

$$\begin{aligned} &\mathbb{E}[Z_{\tau+1}[l']Z_{t+1}[l]] \\ &= \frac{\delta_{l,l'}}{L\delta[l]} \mathbb{E}\left[f_{\mathrm{in}}[l] \left(\frac{\bar{\eta}_\tau[l]}{L} X + \tilde{H}_\tau[l]; \eta_\tau[l], \bar{\sigma}_\tau^2[l] \right) \right. \\ &\quad \left. \cdot f_{\mathrm{in}}[l] \left(\frac{\bar{\eta}_t[l]}{L} X + \tilde{H}_t[l]; \eta_t[l], \bar{\sigma}_t^2[l] \right) \right] \end{aligned} \tag{36}$$

for all $\tau \in \{0, \ldots, t\}$.

The definitions (23)–(36) provide state evolution recursion for D-GAMP. The significance of these definitions is presented in the following theorem:

Theorem 1: Suppose that Assumptions 1, 2, 3, and 4 hold. Then, for all iterations $t = 0, 1, \ldots$ D-GAMP satisfies

$$\begin{aligned} &(\boldsymbol{z}[l], \{\tilde{\boldsymbol{z}}_t[l]\}_{l \in \mathcal{L}}, \{\boldsymbol{w}[l]\}_{l \in \mathcal{L}}) \\ &\overset{\mathrm{PL}(2)}{\to} (Z[l], \{Z_t[l]\}_{l \in \mathcal{L}}, \{W[l]\}_{l \in \mathcal{L}}), \end{aligned} \tag{37}$$

$$(\boldsymbol{x}, \{\tilde{\boldsymbol{x}}_t[l] - L^{-1}\bar{\eta}_t[l]\boldsymbol{x}\}_{l \in \mathcal{L}}) \overset{\mathrm{PL}(2)}{\to} (X, \{\tilde{H}_t[l]\}_{l \in \mathcal{L}}) \tag{38}$$

in the large system limit, where the zero-mean Gaussian random variables Z_t, $Z_t[l]$ and $\tilde{H}_t[l]$ are given via (23)–(36) to represent state evolution recursion.

Proof: See Appendix A. ■

Theorem 1 implies the asymptotic Gaussianity for the input messages to the outer and inner denoisers. In particular, the error covariance for D-GAMP converges almost surely to

$$\begin{aligned} &\frac{1}{N}(\hat{\boldsymbol{x}}_{\tau+1}[l] - \boldsymbol{x})^{\mathrm{T}}(\hat{\boldsymbol{x}}_{t+1}[l] - \boldsymbol{x}) \\ &\overset{\mathrm{a.s.}}{\to} \mathbb{E}\left[\left\{ f_{\mathrm{in}}[l] \left(\frac{\bar{\eta}_\tau[l]}{L} X + \tilde{H}_\tau[l]; \eta_\tau[l], \bar{\sigma}_\tau^2[l] \right) - X \right\} \right. \\ &\left. \cdot \left\{ f_{\mathrm{in}}[l] \left(\frac{\bar{\eta}_t[l]}{L} X + \tilde{H}_t[l]; \eta_t[l], \bar{\sigma}_t^2[l] \right) - X \right\} \right] \\ &\equiv \mathrm{cov}_{\tau+1,t+1}[l], \end{aligned} \tag{39}$$

$$\begin{aligned} &-\frac{1}{N}\boldsymbol{x}^{\mathrm{T}}(\hat{\boldsymbol{x}}_{t+1}[l] - \boldsymbol{x}) \\ &\overset{\mathrm{a.s.}}{\to} \mathbb{E}\left[X^2 - X f_{\mathrm{in}}[l] \left(\frac{\bar{\eta}_t[l]}{L} X + \tilde{H}_t[l]; \eta_t[l], \bar{\sigma}_t^2[l] \right) \right] \\ &\equiv \mathrm{cov}_{0,t+1}[l], \end{aligned} \tag{40}$$

$$\frac{1}{N}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \overset{\mathrm{a.s.}}{\to} \mathbb{E}[X^2] \equiv \mathrm{cov}_{0,0}[l] \tag{41}$$

in the large system limit.

It is possible to derive a closed form with respect to the covariance $\mathbb{E}[\tilde{H}_\tau[l]\tilde{H}_t[l]]$ when the network is a tree.

Theorem 2: Suppose that Assumption 5 holds and let $M_t[l] = f_{\mathrm{out}}[l](Z_t[l], Y[l]; \bar{v}_t[l])$. Then, the covariance $\bar{\Sigma}_{\tau,t}[l] = \mathbb{E}[\tilde{H}_\tau[l]\tilde{H}_t[l]]$ is given by

$$\bar{\Sigma}_{\tau,t}[l] = \frac{\mathbb{E}[M_\tau[l]M_t[l]]}{L\bar{\xi}_{\mathrm{out},\tau}[l]\bar{\xi}_{\mathrm{out},t}[l]} + \sum_{l' \in \mathcal{N}[l]} \underline{\bar{\Sigma}}_{\tau,t,J}[l' \to l] \tag{42}$$

for all $\tau \in \{0, \ldots, t\}$, with

$$\begin{aligned} \underline{\bar{\Sigma}}_{\tau,t,j}[l \to l'] &= \frac{\mathbb{E}[\underline{M}_\tau[l]\underline{M}_t[l]]}{L\underline{\bar{\xi}}_{\mathrm{out},\tau}[l]\underline{\bar{\xi}}_{\mathrm{out},t}[l]} \\ &\quad + \sum_{\tilde{l}' \in \mathcal{N}[l] \setminus \{l'\}} \underline{\bar{\Sigma}}_{\tau,t,j-1}[\tilde{l}' \to l]. \end{aligned} \tag{43}$$

As an initial condition, $\underline{\bar{\Sigma}}_{\tau,t,0}[l' \to l] = \underline{\bar{\Sigma}}_{\tau-T,t-T,J}[l' \to l]$ is used, as well as $\underline{\bar{\Sigma}}_{\tau-T,t-T,J}[l' \to l]] = 0$ for all $t < T$.

Proof: See Appendix B. ■

In evaluating the variance $\bar{\Sigma}_{t,t}[l]$ in (42) for $\tau = t$, the covariance $\mathbb{E}[Z_\tau[l]Z_t[l]]$ for $\tau \neq t$ is not needed. In other words, the variance variables $\mathbb{E}[Z^2[l]]$, $\mathbb{E}[Z[l]Z_t[l]]$, $\mathbb{E}[Z_t^2[l]]$, and $\bar{\Sigma}_{t,t}[l]$ satisfy closed-form state evolution recursion with respect to these variables. Nonetheless, we have evaluated the covariance between messages in all preceding iterations to follow the long-memory proof strategy [46].

We next investigate fixed points of the state evolution recursion for D-GAMP in tree-structured networks when homogeneous measurements are considered.

Theorem 3: Let $\delta[l]=\delta$ and $W[l]\sim W$ i.i.d. in Assumption 2 for all $l\in\mathcal{L}$, with some random variable W. Furthermore, consider $g[l]=g$, $f_{\mathrm{out}}[l]=f_{\mathrm{out}}$, and $f_{\mathrm{in}}[l]=f_{\mathrm{in}}$ with identical functions g, f_{out} and f_{in} for all nodes $l\in\mathcal{L}$. If Assumption 5 holds, then the state evolution recursion with respect to the variance variables for D-GAMP has the same fixed points as those for centralized GAMP [18].

Proof: See Appendix C. ∎

Theorem 3 implies that the consensus is achieved for tree-structured networks if D-GAMP converges. To realize the consensus for general ad hoc networks, one may replace consensus propagation with a distributed protocol in [32], as used in [33]–[37],

$$\tilde{\boldsymbol{x}}_t[l]=\boldsymbol{x}_t[l]+\gamma\sum_{l'\in\mathcal{N}[l]}(\boldsymbol{x}_t[l']-\boldsymbol{x}_t[l]) \tag{44}$$

for $T=1$, with $\gamma>0$. However, this naive protocol cannot achieve the same performance as that for centralized GAMP.

Theorem 4: Let $T=1$ and replace the update rule of $\tilde{\boldsymbol{x}}_t[l]$ in (12) with the distributed protocol (44). Suppose that Assumptions 1, 2, 3, and 4 hold. Then, the fixed points of state evolution recursion for D-GAMP with the distributed protocol (44) are different from those for centralized GAMP.

Proof: See Appendix D. ∎

The intuition of Theorem 4 is as follows: To achieve the performance of centralized GAMP, the effective signal-to-noise ratio (SNR) $L^{-2}\bar{\eta}_t^2[l]/\bar{\Sigma}_{t,t}[l]$ in the inner denoiser has to converge toward the same fixed point as that for centralized GAMP. This convergence is realizable for consensus propagation since both signal power $L^{-2}\bar{\eta}_t^2[l]$ and noise power $\bar{\Sigma}_{t,t}[l]$ are updated via consensus propagation, as shown in (29) and (42). However, the distributed protocol (44) results in different protocols for the signal and noise power. As a result, the consensus for the signal and noise power is not achievable simultaneously.

C. Bayes-Optimal Denoisers

We consider the Bayes-optimal denoisers in terms of MMSE. D-GAMP using the Bayes-optimal inner and outer denoisers—called Bayes-optimal D-GAMP—has three advantages: A first advantage is the optimality in terms of the MSE performance. A second advantage is that the state evolution recursion is simplified. This simplification is due to the fact that the update rules in D-GAMP are matched to the state evolution recursion. The last advantage is the convergence guarantee for the state evolution recursion. The convergence is systematically proved via the long-memory strategy [46].

We first focus on the Bayes-optimal inner denoiser. The inner denoiser is designed so as to minimize the MSE (39) with $\tau=t$. We know that the MSE is minimized if the inner denoiser is the posterior mean estimator of X given a scalar measurement $U_t[l]$,

$$U_t[l]=\frac{\bar{\eta}_t[l]}{L}X+\tilde{H}_t[l], \tag{45}$$

where $\tilde{H}_t[l]\sim\mathcal{N}(0,\bar{\Sigma}_{t,t}[l])$ is independent of X. Thus, the Bayes-optimal inner denoiser is defined as the posterior mean of X given $U_t[l]$,

$$f_{\mathrm{in}}[l](u;\bar{\eta}_t[l],\bar{\Sigma}_{t,t}[l])=\mathbb{E}[X|U_t[l]=u]. \tag{46}$$

We present an existing result [46, Lemma 2] for the Bayes-optimal inner denoiser (46), which is a key lemma to evaluate the covariance (36) in the long-memory proof strategy [46].

Lemma 1 ([46]): Consider the Bayes-optimal inner denoiser (46). For given $t\geq 0$, assume $\bar{\Sigma}_{\tau,t}[l]=\bar{\Sigma}_{t,t}[l]$ in (42) for all $\tau\in\{0,\ldots,t\}$ and $\bar{\Sigma}_{\tau',\tau'}[l]>\bar{\Sigma}_{\tau,\tau}[l]$ for all $\tau'<\tau\leq t$. If $\bar{\sigma}_t^2[l]$ in (31) is equal to $\bar{\Sigma}_{t,t}[l]$, then we have $\mathrm{cov}_{\tau+1,t+1}[l]=\mathrm{cov}_{t+1,t+1}[l]$ in (39) for all $\tau\in\{0,\ldots,t\}$.

The assumptions in Lemma 1 can be understood as follows: They imply the cascaded representation of $U_\tau[l]$ and $U_t[l]$:

$$U_\tau[l]=U_t[l]+\Delta\tilde{H}_{\tau,t}[l],\quad U_t[l]=\frac{\bar{\eta}[l]}{L}X+\tilde{H}_t[l], \tag{47}$$

where $\tilde{H}_t[l]\sim\mathcal{N}(0,\bar{\Sigma}_{t,t}[l])$ and $\Delta\tilde{H}_{\tau,t}[l]\sim\mathcal{N}(0,\bar{\Sigma}_{\tau,\tau}[l]-\bar{\Sigma}_{t,t}[l])$ are independent of all random variables. Since $U_\tau[l]$ is a noisy measurement of $U_t[l]$, the measurement $U_\tau[l]$ provides no additional information on X when $U_t[l]$ is observed. Thus, we have $\mathbb{E}[X|U_\tau[l],U_t[l]]=\mathbb{E}[X|U_t[l]]=f_{\mathrm{in}}[l](U_t[l];\bar{\eta}_t[l],\bar{\Sigma}_{t,t}[l])$, which is used to prove Lemma 1.

We next design the outer denoiser so as to maximize the signal-to-noise ratio (SNR) $L^{-2}\bar{\eta}_t^2[l]/\bar{\Sigma}_{t,t}[l]$ in the measurement model (45) for the inner denoiser. Using the definitions of $\bar{\eta}_t[l]$ and $\bar{\Sigma}_{t,t}[l]$ in (29) and (42), we find that the SNR $L^{-2}\bar{\eta}_t^2[l]/\bar{\Sigma}_{t,t}[l]$ after consensus propagation is maximized when the individual SNR $L\bar{\zeta}_t^2[l]/\mathbb{E}[M_t^2[l]]$ before consensus propagation is maximized for all $l\in\mathcal{L}$.

To solve this SNR maximization problem, we consider a scalar measurement model for the outer denoiser,

$$Y[l]=g[l](Z[l],W[l]), \tag{48}$$

$$Z_0[l]=0,\quad Z_\tau[l]=Z[l]+B_\tau[l],\quad Z_t[l]=Z[l]+B_t[l] \tag{49}$$

for $t\geq\tau>0$, where $B_\tau[l]$ and $B_t[l]$ are independent of $W[l]$ and zero-mean Gaussian random variables with covariance

$$\mathbb{E}[Z[l]B_t[l]]=\mathbb{E}[Z[l]Z_t[l]]-\mathbb{E}[Z^2[l]], \tag{50}$$

$$\mathbb{E}[B_\tau[l]B_t[l]]=\mathbb{E}[(Z_\tau[l]-Z[l])(Z_t[l]-Z[l])], \tag{51}$$

defined with (23), (35), and (36) for $t>0$. While the outer denoiser in iteration t is defined with only $Z_t[l]$, the two random variables $Z_\tau[l]$ and $Z_t[l]$ are considered to evaluate the covariance $\mathbb{E}[M_\tau[l]M_t[l]]$ in (42), which is needed in the long-memory proof strategy [46].

We evaluate the covariance in (50) and (51). Using the definitions (35), (36), (39), and (40) yields

$$-\mathbb{E}[Z[l]B_t[l]]=\frac{1}{L\delta[l]}\mathrm{cov}_{0,t}[l]\equiv\bar{v}_{0,t}[l], \tag{52}$$

$$\mathbb{E}[B_\tau[l]B_t[l]]=\frac{1}{L\delta[l]}\mathrm{cov}_{\tau,t}[l]\equiv\bar{v}_{\tau,t}[l], \tag{53}$$

with $\bar{v}_{0,0}[l] = (L\delta[l])^{-1}\mathrm{cov}_{0,0}[l]$. The following lemma presents the optimal solution to the SNR maximization problem:

Lemma 2: Let $\hat{Z}_t[l](\theta, y; \bar{v}_{t,t}[l])$ denote the posterior mean estimator of $Z[l]$ given $Z_t[l] = \theta$ and $Y[l] = y$,

$$\hat{Z}_t[l](\theta, y; \bar{v}_{t,t}[l]) = \frac{\int z P_{Y[l]|Z[l]}(y|z) e^{-\frac{(z-\theta)^2}{2\bar{v}_{t,t}[l]}} dz}{\int P_{Y[l]|Z[l]}(y|z) e^{-\frac{(z-\theta)^2}{2\bar{v}_{t,t}[l]}} dz}, \tag{54}$$

where $P_{Y[l]|Z[l]}(y|z)$ represents the conditional distribution[2] of $Y[l]$ given $Z[l]$, induced from the randomness of $W[l]$ through $Y[l] = g[l](Z[l], W[l])$. If $\bar{v}_{0,t}[l] = \bar{v}_{t,t}[l]$ holds, then for the individual SNR $\bar{\zeta}_{\mathrm{out},t}^2[l]/\mathbb{E}[M_t^2[l]]$ we have the following inequality:

$$\frac{\bar{\zeta}_{\mathrm{out},t}^2[l]}{\mathbb{E}[M_t^2[l]]} \leq \mathbb{E}\left[\left\{\frac{Z_t[l] - \hat{Z}_t[l](Z_t[l], Y[l]; \bar{v}_{t,t}[l])}{\bar{v}_{t,t}[l]}\right\}^2\right], \tag{55}$$

where the equality holds if and only if the outer denoiser is given by

$$f_{\mathrm{out}}[l](\theta, y; \bar{v}_{t,t}[l]) = C\left(\frac{\theta - \hat{Z}_t[l](\theta, y; \bar{v}_{t,t}[l])}{\bar{v}_{t,t}[l]}\right) \tag{56}$$

for any constant $C \in \mathbb{R}$.

Proof: See Appendix E-A. ■

Use of the Bayes-optimal inner denoiser (46) justifies the assumption $\bar{v}_{0,t}[l] = \bar{v}_{t,t}[l]$ for the Bayes-optimal inner denoiser (46), as shown shortly. This paper uses the Bayes-optimal outer denoiser (56) with $C = 1$ while [18] and [26] used $C = -1$ and $C = -\bar{v}_{t,t}^{1/2}[l]$, respectively. Of course, these choices of the arbitrary constant C do not provide any impacts on the performance of D-GAMP.

It is open whether any Bayes-optimal denoiser is piecewise Lipschitz-continuous. Thus, we postulate the following assumption instead of Assumption 4:

Assumption 6: The composition $f_{\mathrm{out}}[l](\theta, g[l](z, w); \bar{v}_{t,t}[l])$ of the measurement function $g[l]$ in (1) and Bayes-optimal outer denoiser $f_{\mathrm{out}}[l]$ in (56) is piecewise Lipschitz-continuous with respect to $(\theta, z, w) \in \mathbb{R}^3$. The Bayes-optimal inner denoiser $f_{\mathrm{in}}[l](u; \bar{\eta}_t[l], \bar{\Sigma}_{t,t}[l])$ in (46) is piecewise Lipschitz-continuous with respect to $u \in \mathbb{R}$.

The following lemma is a key result to evaluate the covariance $\bar{\Sigma}_{\tau,t}[l]$ given by (42) in the long-memory proof strategy [46].

Lemma 3: Suppose that Assumption 6 holds and consider the Bayes-optimal outer denoiser (56) with $C = 1$. For given $t \geq 0$, assume $\bar{v}_{\tau,t}[l] = \bar{v}_{t,t}[l]$ in (53) for all $\tau \in \{0, \ldots, t\}$ and $\bar{v}_{\tau',\tau'}[l] > \bar{v}_{\tau,\tau}[l]$ for all $\tau' < \tau \leq t$. If $\bar{v}_t[l]$ in (34) is equal to $\bar{v}_{t,t}[l]$, then we have

$$\mathbb{E}[M_\tau[l] M_t[l]] = \bar{\xi}_{\mathrm{out},\tau}[l] \tag{57}$$

for all $\tau \in \{0, \ldots, t\}$.

Proof: See Appendix E-B. ■

[2]The conditional probability density function should be used if $Y[l]$ is a continuous random variable.

The state evolution recursion for D-GAMP is simplified when the Bayes-optimal inner denoiser (46) and outer denoiser (56) are used. We first present the simplified state evolution recursion. For the outer module, we have

$$\mathbb{E}[M_t^2[l]] = \mathbb{E}[f_{\mathrm{out}}^2[l](Z_t[l], Y[l]; \bar{v}_{t,t}[l])], \tag{58}$$

$$\bar{\Sigma}_{\tau,t}[l] = \frac{1}{L\mathbb{E}[M_t^2[l]]} + \sum_{l' \in \mathcal{N}[l]} \underline{\bar{\Sigma}}_{t,t,J}[l' \to l] \tag{59}$$

for all $\tau \in \{0, \ldots, t\}$, with

$$\underline{\bar{\Sigma}}_{t,t,j}[l \to l'] = \frac{1}{L\mathbb{E}[\underline{M}_t^2[l]]} + \sum_{\tilde{l}' \in \mathcal{N}[l] \setminus \{l'\}} \underline{\bar{\Sigma}}_{t,t,j-1}[\tilde{l}' \to l]. \tag{60}$$

As initial conditions, $\bar{v}_{0,0}[l] = (L\delta[l])^{-1}\mathbb{E}[X^2]$ and $\underline{\bar{\Sigma}}_{t,t,0}[l' \to l] = \underline{\bar{\Sigma}}_{t-T,t-T,J}[l' \to l]$ are used, as well as $\underline{\bar{\Sigma}}_{t-T,t-T,J}[l' \to l] = 0$ for all $t < T$. Here, the expectation in (58) is over $Y[l] = g[l](Z[l], W[l])$ and $Z_t[l]$ defined in (23), (35), and (36).

For the inner module, we have

$$\mathrm{mse}_{t+1}[l] = \mathbb{E}\left[\left\{f_{\mathrm{in}}[l]\left(\frac{\eta_t[l]}{L}X + \tilde{H}_t[l]; \eta_t[l], \bar{\Sigma}_{t,t}[l]\right) - X\right\}^2\right], \tag{61}$$

$$\bar{v}_{\tau,t+1}[l] = \frac{1}{L\delta[l]}\mathrm{mse}_{t+1}[l] \tag{62}$$

for all $\tau \in \{0, \ldots, t+1\}$, with $\eta_t[l]$ given in (13), where $\tilde{H}_t[l] \sim \mathcal{N}(0, \bar{\Sigma}_{t,t}[l])$ is independent of X.

The following theorem shows that the update rules in Bayes-optimal D-GAMP are matched to the state evolution recursion. As a result, the state evolution recursion is simplified. Furthermore, the state evolution recursion for Bayes-optimal D-GAMP is guaranteed to converge toward a fixed point.

Theorem 5: Suppose that Assumptions 1, 2, 3, 5 and 6 hold and consider the Bayes-optimal inner denoiser (46) and outer denoiser (56) with $C = 1$. Then, we have the following results:

- Bayes-optimal D-GAMP is consistent: $\bar{v}_t[l] = \bar{v}_{t,t}[l]$, $\bar{\eta}_t[l] = \eta_t[l]$, and $\bar{\sigma}_t^2[l] = \bar{\Sigma}_{t,t}[l]$ hold for all t.
- The error covariance $N^{-1}(\hat{\boldsymbol{x}}_\tau[l] - \boldsymbol{x}[l])^{\mathrm{T}}(\hat{\boldsymbol{x}}_t[l] - \boldsymbol{x}[l])$ for Bayes-optimal D-GAMP converges almost surely to $\mathrm{mse}_t[l]$ in the large system limit for all $\tau \in \{0, \ldots, t\}$, in which $\mathrm{mse}_t[l]$ is given via the simplified state evolution recursion (58)–(62).
- The state evolution recursion (58)–(62) for Bayes-optimal D-GAMP converges to a fixed point as $t \to \infty$.

Proof: See Appendix F. ■

To the best of author's knowledge, Theorem 5 is the first result for the convergence guarantee of the state evolution recursion in general settings even for Bayes-optimal centralized GAMP [18]. Since we know the optimality of Bayes-optimal centralized GAMP [26], from Theorems 3 and 5 we can conclude that the state evolution recursion for Bayes-optimal D-GAMP converges to the Bayes-optimal fixed point for the homogeneous measurements in Theorem 3 when the fixed point is unique.

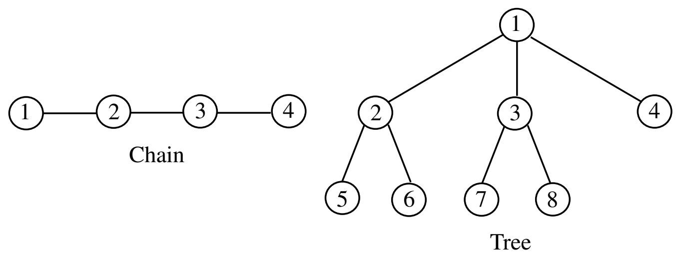

Fig. 1. Tree-structured networks with no central nodes.

V. Numerical Results

A. Numerical Conditions

In all numerical results, the i.i.d. Bernoulli-Gaussian signals with signal density $\rho \in (0,1]$ are considered: x_n is independently sampled from the Gaussian distribution $\mathcal{N}(0,\rho^{-1})$ with probability ρ. Otherwise, x_n is set to zero. This signal has the unit power $\mathbb{E}[x_n^2]=1$. The noise vector $\boldsymbol{w}[l] \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{M[l]})$ in the measurement model (1) has independent zero-mean Gaussian elements with variance $\sigma^2>0$. The sensing matrix $\boldsymbol{A}[l] \in \mathbb{R}^{M[l]\times N}$ in node l has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$. As examples of tree-structured networks, a one-dimensional chain and a tree with no central nodes in Fig. 1 are considered. These assumptions satisfy Assumptions 1, 2, 3, and 5.

This paper considers two measurement functions $g[l](z,w)$: One is the linear measurement $g[l](z,w)=z+w$ to test D-AMP. The other is clipping with threshold $A>0$,

$$g[l](z,w) = \begin{cases} A & \text{for } z+w>A \\ z+w & \text{for } |z+w|\leq A \\ -A & \text{for } z+w<-A, \end{cases} \tag{63}$$

which is used to evaluate D-GAMP.

Bayes-optimal D-GAMP is used. When the linear measurement $g[l](z,w)=z+w$ is considered, the posterior mean estimator (54) in the outer denoiser (56) reduces to

$$\hat{Z}_t[l](\theta,y;v_t[l]) = \theta + \frac{v_t[l]}{v_t[l]+\sigma^2}(y-\theta). \tag{64}$$

Thus, the Bayes-optimal outer denoiser (56) with $C=1$ is given by

$$f_{\mathrm{out}}[l](\theta,y;v_t[l]) = \frac{\theta-y}{v_t[l]+\sigma^2}. \tag{65}$$

In this case, D-GAMP is essentially[3] equivalent to D-AMP [43].

For the clipping case (63), we have the following Bayes-optimal outer denoiser (56) with $C=1$:

$$f_{\mathrm{out}}[l](\theta,y;v_t[l]) = \frac{\theta-y}{v_t[l]+\sigma^2} \quad \text{for } |y|<A, \tag{66}$$

$$f_{\mathrm{out}}[l](\theta,y;v_t[l]) = -\frac{p_{\mathrm{G}}(A-\theta, v_t[l]+\sigma^2)}{Q\left((A-\theta)(v_t[l]+\sigma^2)^{-1/2}\right)} \tag{67}$$

for $y>A$, and

$$f_{\mathrm{out}}[l](\theta,y;v_t[l]) = \frac{p_{\mathrm{G}}(A+\theta, v_t[l]+\sigma^2)}{Q\left((A+\theta)(v_t[l]+\sigma^2)^{-1/2}\right)} \tag{68}$$

for $y<-A$. In these expressions, $p_{\mathrm{G}}(\cdot;v)$ and $Q(x)$ denote the zero-mean Gaussian probability density function with variance v and complementary cumulative distribution function of the standard Gaussian distribution, respectively. It is an exercise to confirm the piecewise Lipschitz-continuity of the composition $f_{\mathrm{out}}[l](\theta, g[l](z,w); v_t[l])$ in Assumption 6, by using the well-known inequalities $x(1+x^2)^{-1}p_{\mathrm{G}}(x;1) < Q(x) < x^{-1}p_{\mathrm{G}}(x;1)$ for all $x>0$.

Damping [56]–[58] is a heuristic technique to improve the convergence property of message-passing algorithms for finite-sized systems. In this paper, damping was used just after inner denoising in each node: The update rules (15) and (17) were replaced with

$$\hat{\boldsymbol{x}}_{t+1}[l] = \chi f_{\mathrm{in}}[l](\tilde{\boldsymbol{x}}_t[l]; \eta_t[l], \sigma_t^2[l]) + (1-\chi)\hat{\boldsymbol{x}}_t[l], \tag{69}$$

$$v_{t+1}[l] = \chi \frac{N}{M[l]} \frac{\sigma_t^2[l]\xi_{\mathrm{in},t}[l]}{\eta_t[l]} + (1-\chi)v_t[l], \tag{70}$$

with damping factor $\chi \in (0,1]$. While it is possible to design t-dependent (or l-dependent) damping factors via deep learning [59], for simplicity, this paper considers the constant damping factor χ for all t and l, which was optimized via exhaustive search.

As a baseline, this paper considers centralized AMP [1] or GAMP [18] using $\sum_{l\in\mathcal{L}} M[l]$ measurements. The purpose of D-GAMP is to achieve the same MSE performance as the corresponding centralized GAMP. In all numerical results, 10^4 independent trials were simulated.

B. Chain Network

The one-dimensional chain network in Fig. 1 is considered. The linear measurement is first assumed to compare D-GAMP with conventional D-AMP [43]. D-GAMP with $T[l]=1$ is essentially equivalent to D-AMP [43] for the linear measurement.

Figure 2 shows numerical comparisons between D-GAMP and D-AMP [43] in terms of the total number of inner iterations for consensus propagation. As proved in Theorem 5, the state evolution recursion for D-GAMP converges to the fixed point of the state evolution recursion for the corresponding centralized AMP. Furthermore, D-GAMP for $T[l]=2$ and $J=1$ converges more quickly than D-AMP [43] with $J=1$ or $J=2$ while $J\geq 4$ was used in [43]. These observations imply that D-GAMP can reduce network traffic for consensus propagation compared to D-AMP [43].

We next consider the clipping case in Fig. 3. The basic observations are similar to those in Fig. 2. As a heterogeneous case, D-GAMP with $T[l]=2$ for odd l and $T[l]=1$ for even l is shown. This case corresponds to a heterogeneous situation in which the odd-numbered nodes have twice higher processing speed than the even-numbered nodes, so that they can repeat two GAMP iterations while the even-numbered nodes compute a single GAMP iteration. The performance of D-GAMP in the heterogeneous case is between that in the two homogeneous

[3] D-AMP [43] replaced the quantity $\xi_{\mathrm{out},t}^{-1}[l]$ given in (4) with the well-known estimator $M^{-1}[l]\|\tilde{\boldsymbol{z}}_t[l]-\boldsymbol{y}[l]\|^2$ in (16). However, numerical simulations showed that this estimator or its robust alternative had large errors for finite-sized systems with a non-negligible probability. Thus, this paper uses the original definition (4) in D-AMP.

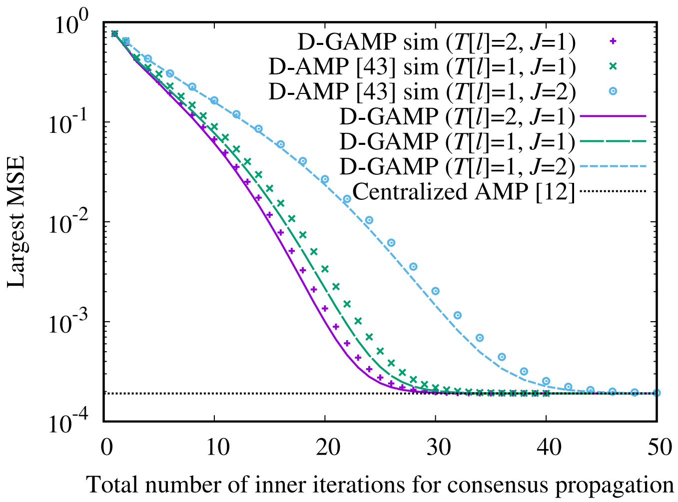

Fig. 2. Largest MSE versus the total number of inner iterations for consensus propagation in the linear measurements. One-dimensional chain network with $L = 4$ nodes, measurement dimension $M[l] = 480$, signal dimension $N = 6400$, signal density $\rho = 0.1$, SNR $1/\sigma^2 = 30$ dB, and damping factor $\chi = 1$. The solid curves show state evolution results while numerical simulations are plotted with markers.

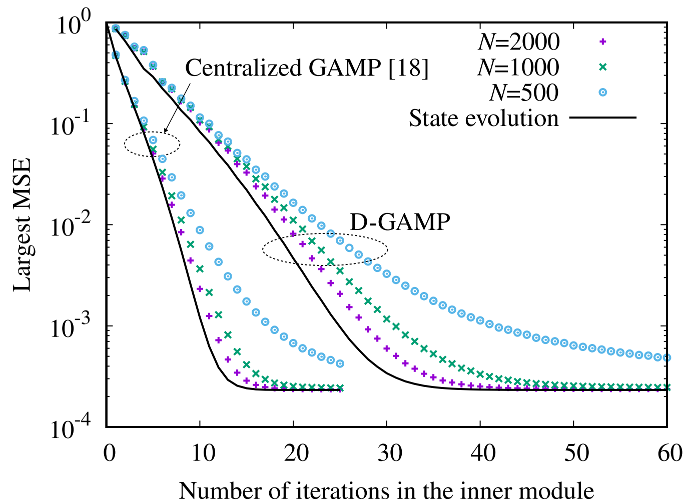

Fig. 4. Largest MSE versus the number of iterations in the inner module for the clipping case. Tree network with $L = 8$ nodes, compression ratio $M[l]/N = 0.05$, signal density $\rho = 0.1$, threshold $A = 2$, SNR $1/\sigma^2 = 30$ dB, $T[l] = 1$, and $J = 1$. For $N = 500, 1000, 2000$, D-GAMP used damping factors $\chi = 0.9, 1, 1$, respectively, while centralized GAMP used $\chi = 0.9, 0.95, 0.95$.

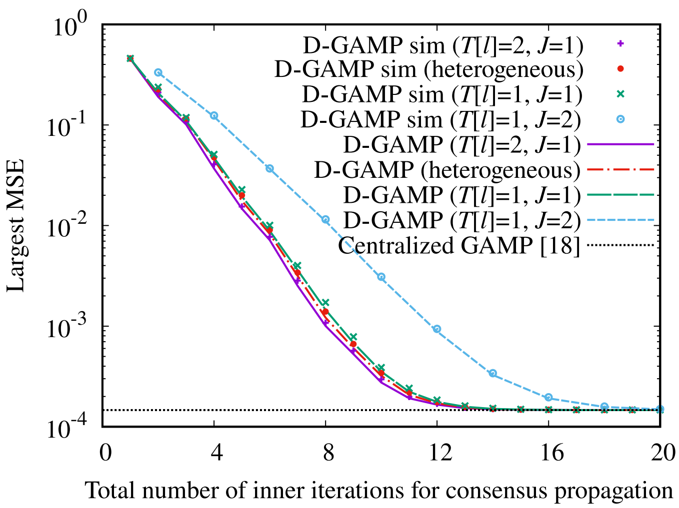

Fig. 3. Largest MSE versus the total number of inner iterations for consensus propagation in the clipping case. One-dimensional chain network with $L = 4$ nodes, measurement dimension $M[l] = 800$, signal dimension $N = 4000$, signal density $\rho = 0.1$, threshold $A = 2$, SNR $1/\sigma^2 = 30$ dB, and damping factor $\chi = 1$. The solid curves show state evolution results while numerical simulations are plotted with markers. As a heterogeneous case, $T[l] = 2$ for odd l, $T[l] = 1$ for even l, and $J = 1$ were considered.

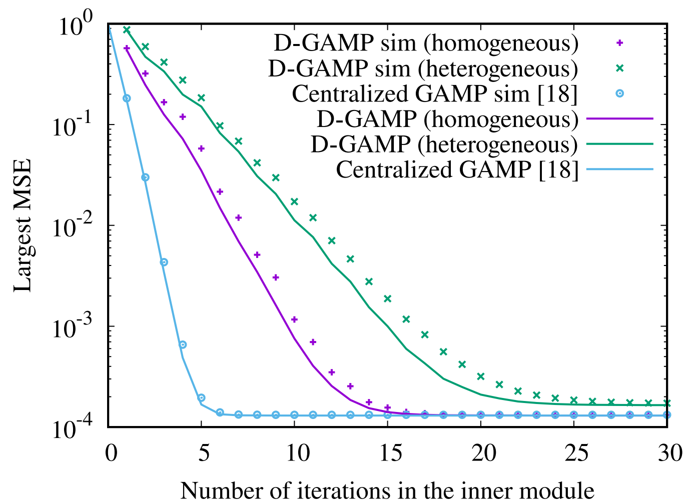

Fig. 5. Largest MSE versus the number of iterations in the inner module for the clipping case. Tree network with $L = 8$ nodes, signal dimension $N = 600$, signal density $\rho = 0.1$, threshold $A = 2$, SNR $1/\sigma^2 = 30$ dB, $T[l] = 1$, and $J = 1$. $M[l] = 90$ and $\chi = 0.95$ were considered in a homogeneous case while $M[l] = 150$ for odd l, $M[l] = 30$ for even l, and $\chi = 0.95$ were considered in a heterogeneous case. Centralized GAMP used damping factor $\chi = 1$.

cases for $J = 1$: For $T[l] = 1$ and $T[l] = 2$ the odd-numbered nodes wait for the completion of one and two GAMP iterations in the even-numbered nodes, respectively.

C. Tree Network

The tree network with $L = 8$ nodes in Fig. 1 is considered. D-GAMP with $T[l] = 1$ and $J = 1$ is compared to centralized GAMP [18] in terms of the number of iterations t. Figure 4 shows that D-GAMP converges to almost the same MSE performance as that of the corresponding centralized GAMP in the three cases $N = 500$, $N = 1000$, and $N = 2000$. As predicted from state evolution, D-GAMP needs more iterations than the corresponding centralized GAMP, because of iterations for consensus propagation. Interestingly, the optimized damping factors for D-GAMP are slightly larger than those for the corresponding centralized GAMP. This observation is because consensus propagation plays a role as kind of damping to slow down the convergence of GAMP. Thus, consensus propagation does not degrade the convergence property of GAMP for finite-sized systems.

We next consider heterogeneous measurements in Fig. 5. The odd-numbered nodes have $M[l] = 150$ measurements while the even-numbered nodes have $M[l] = 30$ measurements. D-GAMP for the heterogeneous case cannot approach the same MSE performance as that of the corresponding centralized GAMP while D-GAMP for the homogeneous case $M[l] = 90$ can achieve the same performance. This is because the Onsager-correction in (3) and (16) depends on the node index l. To achieve the performance of centralized GAMP, an additional protocol is needed to realize the convergence of the Onsager-correction in D-GAMP toward that in the corresponding centralized GAMP.

VI. Conclusions

This paper has proposed D-GAMP for signal reconstruction from distributed generalized linear measurements. D-GAMP is applicable to all tree-structured networks that do not necessarily have central nodes. State evolution has been used to analyze the asymptotic dynamics of D-GAMP for zero-mean i.i.d. Gaussian sensing matrices. The state evolution recursion for Bayes-optimal D-GAMP has been proved to converge toward the Bayes-optimal fixed point—achieved by the corresponding centralized GAMP—for homogeneous measurements with an identical dimension in all nodes when the fixed point is unique.

D-GAMP has two limitations: One limitation is that zero-mean i.i.d. sensing matrices are required. As long as GAMP is used, this assumption cannot be weaken. To solve this issue, GAMP needs to be replaced with another sophisticated message-passing algorithm.

The other limitation is in the assumption of tree-structured networks. To weaken this assumption, we need to replace consensus propagation with another sophisticated protocol for average consensus. As proved in Theorem 4, the conventional protocol (44) for average consensus cannot be used for this purpose.

Appendix A
Proof of Theorem 1

A. Preliminaries

We present a few technical lemmas required in proving Theorem 1 via state evolution. We first formulate a general error model that describes the dynamics of estimation errors for D-GAMP.

Definition 6 (General Error Model): For some $\bar{\eta}_t[l] \in \mathbb{R}$ and $C_{t,i}[l][l'] \in \mathbb{R}$, the general error model is defined with five random vectors $\{\boldsymbol{b}_t[l], \boldsymbol{m}_t[l], \boldsymbol{h}_t[l], \tilde{\boldsymbol{h}}_t[l], \boldsymbol{q}_{t+1}[l]\}$,

$$\boldsymbol{b}_t[l] = \frac{\xi_{\mathrm{in},t-1}[l]}{L\delta[l]}\left\{\frac{\boldsymbol{m}_{t-1}[l]}{\xi_{\mathrm{out},t-1}[l]} + \frac{C_{t,0}[l][l]-1}{\underline{\xi}_{\mathrm{out},t-1}[l]}\underline{\boldsymbol{m}}_{t-1}[l]\right\} + \boldsymbol{A}[l]\boldsymbol{q}_t[l] \tag{71}$$

$$\boldsymbol{m}_t[l] = f_{\mathrm{out}}[l]\left(\boldsymbol{b}_t[l] + \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]}\boldsymbol{z}[l], \boldsymbol{y}[l]; v_t[l]\right), \tag{72}$$

$$\boldsymbol{h}_t[l] = \frac{\xi_{\mathrm{out},t}[l]}{L}\boldsymbol{q}_t[l] - \boldsymbol{A}^{\mathrm{T}}[l]\boldsymbol{m}_t[l], \tag{73}$$

$$\tilde{\boldsymbol{h}}_t[l] = \frac{\boldsymbol{h}_t[l]}{\xi_{\mathrm{out},t}[l]} + \frac{C_{t,0}[l][l]-1}{\underline{\xi}_{\mathrm{out},t}[l]}\underline{\boldsymbol{h}}_t[l] + \sum_{i=0}^{\lfloor t/T\rfloor}\sum_{l'\neq l} C_{t,i}[l][l']\frac{\underline{\boldsymbol{h}}_{t-iT}[l']}{\underline{\xi}_{\mathrm{out},t-iT}[l']}, \tag{74}$$

$$\boldsymbol{q}_{t+1}[l] = f_{\mathrm{in}}[l]\left(\frac{\bar{\eta}_t[l]}{L}\boldsymbol{x} + \tilde{\boldsymbol{h}}_t[l]; \eta_t[l], \sigma_t^2[l]\right) - \frac{\bar{\zeta}_{t+1}[l]}{\bar{\xi}_{\mathrm{out},t+1}[l]}\boldsymbol{x}, \tag{75}$$

where $\xi_{\mathrm{out},t}[l]$, $\xi_{\mathrm{in},t}[l]$, $\bar{\xi}_{\mathrm{out},t}[l]$, and $\bar{\zeta}_t[l]$ are given in (4), (18), (24), and (25), respectively. As initial conditions, we use $\boldsymbol{q}_0[l] = -\bar{\xi}_{\mathrm{out},0}^{-1}[l]\bar{\zeta}_0[l]\boldsymbol{x}$, $\boldsymbol{m}_{-1}[l] = \underline{\boldsymbol{m}}_{-1}[l] = \boldsymbol{0}$. In particular, we have $\boldsymbol{b}_0[l] = -\bar{\xi}_{\mathrm{out},0}^{-1}[l]\bar{\zeta}_0[l]\boldsymbol{z}[l]$.

The vector $\tilde{\boldsymbol{h}}_t[l]$ represents the dynamics of protocols for average consensus. The other vectors describe the dynamics of GAMP iterations. By selecting $\{C_{t,i}[l][l']\}$ appropriately, we obtain the error model of D-GAMP under Assumption 5.

Lemma 4: Suppose that Assumption 5 holds, let $C_{t,0}[l][l] = 1$, and define

$$\boldsymbol{b}_t[l] = \tilde{\boldsymbol{z}}_t[l] - \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]}\boldsymbol{z}[l], \quad \boldsymbol{m}_t[l] = \hat{\boldsymbol{z}}_t[l],$$
$$\boldsymbol{h}_t[l] = \xi_{\mathrm{out},t}[l]\left(\boldsymbol{x}_t[l] - \frac{\bar{\zeta}_t[l]}{L\bar{\xi}_{\mathrm{out},t}[l]}\boldsymbol{x}\right),$$
$$\tilde{\boldsymbol{h}}_t[l] = \tilde{\boldsymbol{x}}_t[l] - \frac{\bar{\eta}_t[l]}{L}\boldsymbol{x}, \quad \boldsymbol{q}_t[l] = \hat{\boldsymbol{x}}_t[l] - \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]}\boldsymbol{x}, \tag{76}$$

with $\xi_{\mathrm{out},t}[l]$, $\bar{\xi}_{\mathrm{out},t}[l]$, $\bar{\zeta}_t[l]$, $\bar{\eta}_t[l]$, and $\underline{\bar{\eta}}_{t,j}[l\to l']$ given in (4), (24), (25), (29), and (30), respectively. Then, there are some $\{C_{t,i}[l][l']\}$ such that these vectors satisfy the dynamics (71)–(75) in the general error model.

Proof: The expression (71) with $C_{t,0}[l][l] = 1$ is obtained by using $\boldsymbol{z}[l] = \boldsymbol{A}[l]\boldsymbol{x}$ and the definition of $\tilde{\boldsymbol{z}}_t[l]$ in (16). The expression (72) follows from the definition of $\hat{\boldsymbol{z}}_t[l]$ in (2). Using the definition of $\boldsymbol{x}_t[l]$ in (3) yields (73). The expression (75) follows from the definition of $\hat{\boldsymbol{x}}_{t+1}[l]$ in (15).

We confirm (74). Let

$$\underline{\boldsymbol{h}}_{t,j}[l\to l'] = \underline{\boldsymbol{x}}_{t,j}[l\to l'] - \frac{\underline{\bar{\eta}}_{t,j}[l\to l']}{L}\boldsymbol{x}. \tag{77}$$

Using the definitions of $\overline{\underline{\boldsymbol{x}}_t[l]}$, $\tilde{\boldsymbol{x}}_t[l]$, and $\bar{\eta}_t[l]$ in (9), (12), and (29) yields

$$\tilde{\boldsymbol{h}}_t[l] = \frac{\boldsymbol{h}_t[l]}{\xi_{\mathrm{out},t}[l]} + \sum_{l'\in\mathcal{N}[l]}\underline{\boldsymbol{h}}_{t,J}[l'\to l]. \tag{78}$$

Furthermore, we use the definitions of $\underline{\boldsymbol{x}}_{t,j}[l\to l']$ and $\underline{\bar{\eta}}_{t,j}[l\to l']$ in (6) and (30) to obtain

$$\underline{\boldsymbol{h}}_{t,j}[l\to l'] = \frac{\underline{\boldsymbol{h}}_t[l]}{\underline{\xi}_{\mathrm{out},t}[l]} + \sum_{\tilde{l}'\in\mathcal{N}[l]\setminus\{l'\}}\underline{\boldsymbol{h}}_{t,j-1}[\tilde{l}'\to l], \tag{79}$$

with $\underline{\boldsymbol{h}}_{t,0}[l'\to l] = \underline{\boldsymbol{h}}_{t-T,J}[l'\to l]$. As an initial condition, $\underline{\boldsymbol{h}}_{t-T,J}[l'] = \boldsymbol{0}$ are used for all $t < T$. The tree assumption in Assumption 5 implies that $\underline{\boldsymbol{h}}_{t,J}[l'\to l]$ does not contain the messages $\underline{\boldsymbol{h}}_\tau[l]/\underline{\xi}_{\mathrm{out},\tau}[l]$ for any $\tau \leq t$ computed in node l. Comparing the expression of $\tilde{\boldsymbol{h}}_t[l]$ with (74), we have $C_{t,0}[l][l] = 1$ and find that there are some $\{C_{t,i}[l][l']\}$ such that $\tilde{\boldsymbol{h}}_t[l]$ reduce to (74). Thus, Lemma 4 holds. ■

The vector $\boldsymbol{b}_t[l]$ corresponds to the estimation errors of $\bar{\xi}_{\mathrm{out},t}^{-1}[l]\bar{\zeta}_t[l]\boldsymbol{z}[l]$ before outer denoising, while $\tilde{\boldsymbol{h}}_t[l]$ represents the estimation errors of $L^{-1}\bar{\eta}_t[l]\boldsymbol{x}$ before inner denoising.

In Bolthausen's conditioning technique [15], the dynamics of the vectors in the current iteration is evaluated via the conditional distribution of $\{\boldsymbol{A}[l]\}$ given those in all previous iterations, the signal vector $\boldsymbol{x}$, and the noise vector $\{\boldsymbol{w}[l]\}$. For notational convenience, we use the following notation in representing conditioned vectors:

Definition 7: For variables $\{a_\tau[l] \in \mathbb{R}\}$ associated with node l, the column vector $\boldsymbol{a}_t[l]$ contains the variables for the iterations $\mathcal{T}_t[l]$ in (5) where node l updates messages,

$$\boldsymbol{a}_t[l] = (\boldsymbol{a}_t^0[l], \ldots, \boldsymbol{a}_t^{\lfloor t/T \rfloor}[l])^{\mathrm{T}}, \tag{80}$$

with $\boldsymbol{a}_t^i[l] = (a_{iT}[l], \ldots, a_{\min\{t, iT+T[l]\}-1}[l])$. For column vectors $\{\boldsymbol{v}_\tau[l]\}$, the matrix $\boldsymbol{V}_t[l] = \{\boldsymbol{v}_\tau[l] : \tau \in \mathcal{T}_t[l]\}$ has $|\mathcal{T}_t[l]|$ column vectors aligned in the same manner.

We define the conditional distribution of $\{\boldsymbol{A}[l]\}$. Let $\boldsymbol{B}_t[l] \in \mathbb{R}^{M[l]\times|\mathcal{T}_t[l]|}$ denote the matrix defined from $\{\boldsymbol{b}_\tau[l] : \tau \in \mathcal{T}_t[l]\}$ in Definition 7. Similarly, we define $\boldsymbol{M}_t[l]$, $\boldsymbol{H}_t[l]$, $\tilde{\boldsymbol{H}}_t[l]$, and $\boldsymbol{Q}_t[l]$. Furthermore, let

$$\mathfrak{E}_{1,0} = \{\{\boldsymbol{b}_0[l] : l \in \mathcal{L}\}, \{\boldsymbol{m}_0[l] : l \in \mathcal{L}\}\}, \tag{81}$$

$$\begin{aligned}\mathfrak{E}_{t',t} = \Big\{&\{\boldsymbol{B}_{t'}[l] : l \in \mathcal{L}\}, \{\boldsymbol{M}_{t'}[l] : l \in \mathcal{L}\}, \{\boldsymbol{H}_t[l] : l \in \mathcal{L}\},\\ &\{\tilde{\boldsymbol{H}}_t[l] : l \in \mathcal{L}\}, \{\boldsymbol{Q}_{t+1}[l] : l \in \mathcal{L}\}\Big\}\end{aligned} \tag{82}$$

for $t' > 0$ and $t > 0$, where the columns in the five matrices satisfy (71)–(75) in the general error model. The set $\mathfrak{E}_{t,t}$ contains the messages that are computed just before updating $\boldsymbol{b}_t[l]$ in (71) while $\mathfrak{E}_{t+1,t}$ includes them just before updating $\boldsymbol{h}_t[l]$ in (73). The signal and noise vectors $\Theta = \{\boldsymbol{x}, \{\boldsymbol{w}[l] : l \in \mathcal{L}\}\}$ are always fixed. Thus, the conditional distribution of $\{\boldsymbol{A}[l]\}$ given $\mathfrak{E}_{t,t}$ and Θ is considered in evaluating the distribution of $\boldsymbol{b}_t[l]$ while the conditional distribution of $\{\boldsymbol{A}[l]\}$ given $\mathfrak{E}_{t+1,t}$ and Θ is considered in evaluating the distribution of $\boldsymbol{h}_t[l]$.

The conditional distributions of $\boldsymbol{A}[l]$ are evaluated via the following existing lemma:

Lemma 5 ([12]): Suppose that Assumption 3 holds. For some integers $t[l] \leq M[l]$ and $t'[l] \leq N$, let $\boldsymbol{X}[l] \in \mathbb{R}^{M[l]\times t[l]}$, $\boldsymbol{U}[l] \in \mathbb{R}^{N\times t[l]}$, $\boldsymbol{Y}[l] \in \mathbb{R}^{N\times t'[l]}$, and $\boldsymbol{V}[l] \in \mathbb{R}^{M[l]\times t'[l]}$ satisfy the following constraints:

$$\boldsymbol{X}[l] = \boldsymbol{A}[l]\boldsymbol{U}[l], \quad \boldsymbol{Y}[l] = \boldsymbol{A}[l]^{\mathrm{T}}\boldsymbol{V}[l]. \tag{83}$$

- If $\boldsymbol{U}[l]$ has full rank, then the conditional distribution of $\boldsymbol{A}[l]$ given $\boldsymbol{X}[l]$ and $\boldsymbol{U}[l]$ is represented as

$$\boldsymbol{A}[l] \sim \boldsymbol{X}[l]\boldsymbol{U}^\dagger[l] + \tilde{\boldsymbol{A}}[l]\boldsymbol{P}_{\boldsymbol{U}[l]}^\perp, \tag{84}$$

  where $\tilde{\boldsymbol{A}}[l]$ is independent of $\{\boldsymbol{X}[l], \boldsymbol{U}[l]\}$ and has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$.
- If Both $\boldsymbol{U}[l]$ and $\boldsymbol{V}[l]$ have full rank, then the conditional distribution of $\boldsymbol{A}[l]$ given $\mathfrak{E}[l] = \{\boldsymbol{X}[l], \boldsymbol{U}[l], \boldsymbol{Y}[l], \boldsymbol{V}[l]\}$ is represented as

$$\boldsymbol{A}[l] \sim \boldsymbol{A}_{\mathrm{bias}}[l] + \boldsymbol{P}_{\boldsymbol{V}[l]}^\perp\tilde{\boldsymbol{A}}[l]\boldsymbol{P}_{\boldsymbol{U}[l]}^\perp, \tag{85}$$

  with

$$\begin{aligned}\boldsymbol{A}_{\mathrm{bias}}[l] &= \boldsymbol{X}[l]\boldsymbol{U}^\dagger[l] + (\boldsymbol{V}^\dagger[l])^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}}[l]\boldsymbol{P}_{\boldsymbol{U}[l]}^\perp\\ &= (\boldsymbol{V}^\dagger[l])^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}}[l] + \boldsymbol{P}_{\boldsymbol{V}[l]}^\perp\boldsymbol{X}[l]\boldsymbol{U}^\dagger[l],\end{aligned} \tag{86}$$

  where $\tilde{\boldsymbol{A}}[l]$ is independent of $\mathfrak{E}[l]$ and has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$.

Lemma 5 and Assumption 3 imply that $\{\boldsymbol{A}[l]\}_{l\in\mathcal{L}}$ are conditionally independent given $\{\mathfrak{E}[l]\}_{l\in\mathcal{L}}$.

The following lemma is used to design the Onsager correction in D-GAMP.

Lemma 6 (Stein's Lemma [49], [60]): Suppose that $\{Z_\tau\}_{\tau=1}^t$ are zero-mean Gaussian random variables. Then, for any piecewise Lipschitz-continuous function $f : \mathbb{R}^t \to \mathbb{R}$ we have

$$\mathbb{E}[Z_{t'}f(Z_1, \ldots, Z_t)] = \sum_{\tau=1}^t \mathbb{E}[Z_{t'}Z_\tau]\mathbb{E}[\partial_\tau f(Z_1, \ldots, Z_t)], \tag{87}$$

where ∂_τ denotes the partial derivative with respect to the τth variable.

Proof: Confirm that it is possible to replace the Lipschitz-continuity in the proof of [49, Lemma 2] with the piecewise Lipschitz-continuity, without any changes in the proof. ■

Note that any piecewise Lipschitz-continuous function f is almost everywhere differentiable. Furthermore, the definition of the piecewise Lipschitz-continuity implies the bounds $|f(Z)| \leq A|Z| + B$ and $|f'(Z)| \leq C$ for some constants A, B, and C. Thus, both sides in (87) exist for any piecewise Lipschitz-continuous function f.

B. State Evolution

We define five kinds of random variables $\{B_t[l], M_t[l], H_t[l], \tilde{H}_t[l], Q_{t+1}[l]\}$ to represent the asymptotic dynamics of the general error model. Consider the initial condition $Q_0[l] = -\bar{\xi}_{\mathrm{out},0}^{-1}[l]\bar{\zeta}_0[l]X$, with X defined in Assumption 1. The random variables $\{B_t[l], M_t[l], H_t[l], \tilde{H}_t[l], Q_{t+1}[l]\}$ are defined recursively.

Let $\{B_t[l]\}$ denote the sequence of zero-mean Gaussian random variables with covariance

$$\mathbb{E}[B_{\tau'}[l']B_\tau[l]] = \frac{\delta_{l,l'}}{L\delta[l]}\mathbb{E}[Q_{\tau'}[l]Q_\tau[l]]. \tag{88}$$

The random variable $M_t[l]$ represents the asymptotic output of the outer module in iteration t, given by

$$M_t[l] = f_{\mathrm{out}}[l](Z_t[l], Y[l]; \bar{v}_t[l]), \tag{89}$$

with $Z[l] = -\bar{\xi}_{\mathrm{out},0}[l]\bar{\zeta}_0^{-1}[l]B_0[l]$ and

$$Z_t[l] = B_t[l] + \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]}Z[l], \tag{90}$$

where $\bar{\xi}_{\mathrm{out},t}[l]$, $\bar{\zeta}_t[l]$, and $\bar{v}_t[l]$ are given by (24), (25), and (34), respectively.

Similarly, let $\{H_t[l] \in \mathbb{R}\}$ denote zero-mean Gaussian random variables with covariance

$$\mathbb{E}[H_{\tau'}[l']H_\tau[l]] = \frac{\delta_{l,l'}}{L}\mathbb{E}[M_{\tau'}[l]M_\tau[l]]. \tag{91}$$

We define the random variable $\tilde{H}_t[l]$ recursively as

$$\tilde{H}_t[l] = \frac{H_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]} + \frac{C_{t,0}[l][l]-1}{\underline{\bar{\xi}}_{\mathrm{out},t}[l]}\underline{H}_t[l] + \sum_{i=0}^{\lfloor t/T \rfloor}\sum_{l'\neq l} C_{t,i}[l][l'] \frac{\underline{H}_{t-iT}[l']}{\underline{\bar{\xi}}_{\mathrm{out},t-iT}[l']}. \tag{92}$$

Then, the random variable $Q_{t+1}[l]$ describes the asymptotic output of the inner module, given by

$$Q_{t+1}[l] = f_{\mathrm{in}}[l]\left(\frac{\bar{\eta}_t[l]}{L}X + \tilde{H}_t[l]; \eta_t[l], \bar{\sigma}_t^2[l]\right) - \frac{\bar{\zeta}_{t+1}[l]}{\bar{\xi}_{\mathrm{out},t+1}[l]}X, \tag{93}$$

with $\bar{\xi}_{\mathrm{out},t}[l]$, $\bar{\zeta}_t[l]$, $\bar{\eta}_t[l]$ and $\bar{\sigma}_t^2[l]$ defined in (24), (25), (29) and (31). State evolution recursion is obtained via these random variables.

We are ready for presenting state evolution results for the general error model. Note that Assumption 5 is not required in state evolution analysis. Thus, the following theorem is applicable to general ad hoc networks.

Theorem 6: Postulate Assumptions 1, 2, 3, and 4. Then, the outer module satisfies the following properties for all $\tau \in \{0, 1, \ldots\}$ in the large system limit:

(Oa) Let $\boldsymbol{\beta}_\tau[l] = \boldsymbol{Q}_\tau^\dagger[l]\boldsymbol{q}_\tau[l]$ and $\boldsymbol{q}_\tau^\perp[l] = \boldsymbol{P}_{\boldsymbol{Q}_\tau[l]}^\perp \boldsymbol{q}_\tau[l]$. Then, for all $l \in \mathcal{L}$ and $\tau > 0$ we have

$$\boldsymbol{b}_\tau[l] \sim \boldsymbol{B}_\tau[l]\boldsymbol{\beta}_\tau[l] + \boldsymbol{M}_\tau[l]\boldsymbol{o}(1) + \tilde{\boldsymbol{A}}[l]\boldsymbol{q}_\tau^\perp[l] \tag{94}$$

conditioned on $\mathfrak{E}_{\tau,\tau}$ and Θ, where $\{\tilde{\boldsymbol{A}}[l]\}$ are independent random matrices and independent of $\{\mathfrak{E}_{\tau,\tau}, \Theta\}$. Each $\tilde{\boldsymbol{A}}[l]$ has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$.

(Ob) For all $l', l \in \mathcal{L}$, and $\tau' \in \mathcal{T}_{\tau+1}[l]$,

$$\frac{1}{M[l]}\boldsymbol{b}_{\tau'}^{\mathrm{T}}[l']\boldsymbol{b}_\tau[l] - \frac{\delta_{l,l'}}{NL\delta[l]}\boldsymbol{q}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{q}_\tau[l] \overset{\mathrm{a.s.}}{\to} 0. \tag{95}$$

(Oc) For $\mathcal{B}_{\tau+1}[l] = \{B_{\tau'}[l] \in \mathbb{R} : \tau' \in \mathcal{T}_{\tau+1}[l]\}$, suppose that $\{\mathcal{B}_{\tau+1}[l]\}_{l\in\mathcal{L}}$ are independent with respect to l and independent of $\{W[l]\}_{l\in\mathcal{L}}$. For each l, $\mathcal{B}_{\tau+1}[l]$ are zero-mean Gaussian random variables with covariance

$$\mathbb{E}[B_{\tau'}[l]B_\tau[l]] = \frac{1}{L\delta[l]}\mathbb{E}[Q_{\tau'}[l]Q_\tau[l]] \tag{96}$$

for all $\tau' \in \mathcal{T}_{\tau+1}[l]$. Then, we have

$$(\{\boldsymbol{b}_{\tau'}[l] : \tau' \in \mathcal{T}_{\tau+1}[l]\}_{l\in\mathcal{L}}, \{\boldsymbol{w}[l]\}_{l\in\mathcal{L}}) \overset{\mathrm{PL}(2)}{\to} (\{\mathcal{B}_{\tau+1}[l]\}_{l\in\mathcal{L}}, \{W[l]\}_{l\in\mathcal{L}}), \tag{97}$$

$$\xi_{\mathrm{out},\tau}[l] \overset{\mathrm{a.s.}}{\to} \bar{\xi}_{\mathrm{out},\tau}[l]. \tag{98}$$

(Od) For all $l \in \mathcal{L}$ and $\tau' \in \mathcal{T}_{\tau+1}[l]$, we have

$$\frac{1}{M[l]}\boldsymbol{b}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{m}_\tau[l] \overset{\mathrm{a.s.}}{\to} \frac{\bar{\xi}_{\mathrm{out},\tau}[l]}{L\delta[l]}\mathbb{E}[Q_{\tau'}[l]Q_\tau[l]]. \tag{99}$$

(Oe) For $\epsilon > 0$ used in Assumptions 1 and 2, the vector $\boldsymbol{m}_\tau[l]$ has bounded $(2+\epsilon)$th moments in the large system limit. Furthermore, the minimum eigenvalue of $M^{-1}[l]\boldsymbol{M}_{\tau+1}^{\mathrm{T}}[l]\boldsymbol{M}_{\tau+1}[l]$ is strictly positive in the large system limit.

On the other hand, the inner module satisfies the following properties for all $\tau \in \{0, 1, \ldots\}$ in the large system limit:

(Ia) Let $\boldsymbol{\alpha}_\tau[l] = \boldsymbol{M}_\tau^\dagger[l]\boldsymbol{m}_\tau[l]$ and $\boldsymbol{m}_\tau^\perp[l] = \boldsymbol{P}_{\boldsymbol{M}_\tau[l]}^\perp \boldsymbol{m}_\tau[l]$. Then, for all $l \in \mathcal{L}$ we have

$$\boldsymbol{h}_0[l] \sim \tilde{\boldsymbol{A}}^{\mathrm{T}}[l]\boldsymbol{m}_0[l] + o(1)\boldsymbol{q}_0[l] \tag{100}$$

conditioned on $\mathfrak{E}_{1,0}$ and Θ. For $\tau > 0$,

$$\boldsymbol{h}_\tau[l] \sim \boldsymbol{H}_\tau[l]\boldsymbol{\alpha}_\tau[l] + \boldsymbol{Q}_{\tau+1}[l]\boldsymbol{o}(1) + \tilde{\boldsymbol{A}}^{\mathrm{T}}[l]\boldsymbol{m}_\tau^\perp[l] \tag{101}$$

conditioned on $\mathfrak{E}_{\tau+1,\tau}$ and Θ. Here, $\{\tilde{\boldsymbol{A}}[l]\}$ are independent random matrices and independent of $\{\mathfrak{E}_{\tau+1,\tau}, \Theta\}$. Each $\tilde{\boldsymbol{A}}[l]$ has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$.

(Ib) For all $l', l \in \mathcal{L}$, and $\tau' \in \mathcal{T}_{\tau+1}[l]$,

$$\frac{1}{N}\boldsymbol{h}_{\tau'}^{\mathrm{T}}[l']\boldsymbol{h}_\tau[l] - \frac{\delta_{l,l'}}{LM[l]}\boldsymbol{m}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{m}_\tau[l] \overset{\mathrm{a.s.}}{\to} 0. \tag{102}$$

(Ic) For $\mathcal{H}_{\tau+1}[l] = \{H_{\tau'}[l] \in \mathbb{R} : \tau' \in \mathcal{T}_{\tau+1}[l]\}$, suppose that $\{\mathcal{H}_{\tau+1}[l]\}_{l\in\mathcal{L}}$ are independent with respect to l and independent of X. For each l, $\mathcal{H}_{\tau+1}[l]$ are zero-mean Gaussian random variables with covariance

$$\mathbb{E}[H_{\tau'}[l]H_\tau[l]] = \frac{1}{L}\mathbb{E}[M_{\tau'}[l]M_\tau[l]]. \tag{103}$$

Then, for all $l \in \mathcal{L}$ we have

$$(\{\boldsymbol{h}_{\tau'}[l] : \tau' \in \mathcal{T}_{\tau+1}[l]\}_{l\in\mathcal{L}}, \boldsymbol{x}) \overset{\mathrm{PL}(2)}{\to} (\{\mathcal{H}_{\tau+1}[l]\}_{l\in\mathcal{L}}, X), \tag{104}$$

$$\xi_{\mathrm{in},\tau}[l] \overset{\mathrm{a.s.}}{\to} \bar{\xi}_{\mathrm{in},\tau}[l]. \tag{105}$$

(Id) For all $l \in \mathcal{L}$ and $\tau' \in \mathcal{T}_{\tau+1}[l]$, we have

$$\frac{1}{N}\boldsymbol{h}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{q}_{\tau+1}[l] \overset{\mathrm{a.s.}}{\to} \frac{\bar{\xi}_{\mathrm{in},\tau}[l]}{L} \cdot \mathbb{E}\left[M_{\tau'}[l]\left(\frac{M_\tau[l]}{\bar{\xi}_{\mathrm{out},\tau}[l]} + \frac{C_{t,0}[l][l]-1}{\underline{\bar{\xi}}_{\mathrm{out},\tau}[l]}\underline{M}_\tau[l]\right)\right]. \tag{106}$$

(Ie) For $\epsilon > 0$ used in Assumption 1, the vector $\boldsymbol{q}_{\tau+1}[l]$ has bounded $(2+\epsilon)$th moments in the large system limit. Furthermore, the minimum eigenvalue of $N^{-1}\boldsymbol{Q}_{\tau+2}^{\mathrm{T}}[l]\boldsymbol{Q}_{\tau+2}[l]$ is strictly positive in the large system limit.

Proof: The proof by induction consists of four steps. A first step is the proof of Properties (Oa)–(Oe) for $\tau = 0$ presented in Appendix A-C, while a second step is the proof of Properties (Ia)–(Ie) for $\tau = 0$ presented in Appendix A-D.

For some $t \in \mathbb{N}$, suppose that Properties (Oa)–(Oe) and Properties (Ia)–(Ie) are correct for all $\tau < t$. We need to prove Properties (Oa)–(Oe) for $\tau = t$ as a third step. See Appendix A-E for the details.

The last step is the proof of Properties (Ia)–(Ie) for $\tau = t$ under the induction hypotheses (Oa)–(Oe) and (Ia)–(Ie) for all $\tau < t$, as well as Properties (Oa)–(Oe) for $\tau = t$ proved in the third step. See Appendix A-F for the details. From these four steps we arrive at Theorem 6. ■

Theorem 6 implies that the general error model (71)–(75) satisfies the asymptotic Gaussianity in each node, i.e. Properties (Oc) and (Ic). Since no additional assumptions on the network are postulated, Theorem 6 is available as a framework to design distributed protocols for average consensus in general networks, instead of consensus propagation for tree-structured networks.

We use Theorem 6 to prove Theorem 1.

Proof of Theorem 1: From Lemma 4 and Theorem 6, it is sufficient to prove that (88)–(93) to represent state evolution recursion for the general error model reduces to that in Theorem 1 for D-GAMP. The expression (26) follows from Property (Ic) in Theorem 6. Using (78) and (79) in Lemma 4, we find the equivalence between $\tilde{H}_t[l]$ in (27) and $\tilde{H}_t[l]$ in (92).

We derive the covariance for $\{Z[l]\}$ and $\{Z_t[l]\}$. Using the definition of $Z_t[l]$ in (90) and $Z[l] = -\bar{\xi}_{\mathrm{out},0}[l]\bar{\zeta}_0^{-1}[l]B_0[l]$ yields

$$\begin{aligned}
&\mathbb{E}[Z_\tau[l']Z_t[l]] = \delta_{l,l'}\mathbb{E}\left[\left(B_\tau[l] - \frac{\bar{\zeta}_\tau[l]\bar{\xi}_{\mathrm{out},0}[l]}{\bar{\xi}_{\mathrm{out},\tau}[l]\bar{\zeta}_0[l]}B_0[l]\right)\right. \\
&\left.\cdot\left(B_t[l] - \frac{\bar{\zeta}_t[l]\bar{\xi}_{\mathrm{out},0}[l]}{\bar{\xi}_{\mathrm{out},t}[l]\bar{\zeta}_0[l]}B_0[l]\right)\right] \\
&= \frac{\delta_{l,l'}}{L\delta[l]}\mathbb{E}\left[\left(Q_\tau[l] - \frac{\bar{\zeta}_\tau[l]}{\bar{\xi}_{\mathrm{out},\tau}[l]}X\right)\left(Q_t[l] - \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]}X\right)\right] \\
&= \frac{\delta_{l,l'}}{L\delta[l]}\mathbb{E}\left[f_{\mathrm{in},\tau-1}[l]f_{\mathrm{in},t-1}[l]\right], \qquad (107)
\end{aligned}$$

with $f_{\mathrm{in},\tau}[l] = f_{\mathrm{in}}[l](L^{-1}\bar{\eta}_\tau[l]X + \tilde{H}_\tau[l]; \eta_\tau[l], \bar{\sigma}_\tau^2[l])$. Here, the second equality follows from Property (Oc) in Theorem 6 and $Q_0[l] = -\bar{\xi}_{\mathrm{out},0}^{-1}[l]\bar{\zeta}_0[l]X$. The last equality is obtained from the definition of $Q_t[l]$ in (93).

Similarly, we obtain the other covariance,

$$\mathbb{E}[Z[l']Z[l]] = \frac{\delta_{l,l'}}{L\delta[l]}\mathbb{E}[X^2], \qquad (108)$$

$$\begin{aligned}
&\mathbb{E}[Z[l']Z_{t+1}[l]] \\
&= \frac{\delta_{l,l'}}{L\delta[l]}\mathbb{E}\left[Xf_{\mathrm{in}}[l]\left(\frac{\bar{\eta}_t[l]}{L}X + \tilde{H}_t[l]; \eta_t[l], \bar{\sigma}_t^2[l]\right)\right]. \qquad (109)
\end{aligned}$$

Thus, we arrive at Theorem 1. ∎

C. Outer Module for $\tau = 0$

Proof of Property (Ob): From Assumption 3 and the definition $\boldsymbol{b}_0[l] = \boldsymbol{A}[l]\boldsymbol{q}_0[l]$ in (71), the vectors $\{\boldsymbol{b}_0[l] : l \in \mathcal{L}\}$ conditioned on $\{\boldsymbol{q}_0[l] : l \in \mathcal{L}\}$ are independent. Furthermore, $\boldsymbol{b}_0[l]$ conditioned on $\{\boldsymbol{q}_0[l] : l \in \mathcal{L}\}$ has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}\|\boldsymbol{q}_0[l]\|^2$. Assumption 1 implies that the variance $(LM[l])^{-1}\|\boldsymbol{q}_0[l]\|^2 = (LM[l])^{-1}(\bar{\xi}_{\mathrm{out},0}^{-1}[l]\bar{\zeta}_0[l])^2\|\boldsymbol{x}\|^2$ converges almost surely to $(L\delta[l])^{-1}(\bar{\xi}_{\mathrm{out},0}^{-1}[l]\bar{\zeta}_0[l])^2\mathbb{E}[X^2]$ in the large system limit. Thus, the strong law of large numbers implies Property (Ob) for $\tau = 0$. ∎

Proof of Property (Oc): We first prove the former convergence in Property (Oc) with [55, Lemma 1]. In proving [55, Lemma 1], pseudo-Lipschitz functions were considered. We only present the main idea for generalizing them to piecewise pseudo-Lipschitz functions.

As an example, consider the expectation $\mathbb{E}[f(z)]$ of a piecewise pseudo-Lipschitz function $f(z)$ for an absolutely continuous random variable $z \in \mathbb{R}$. We separate the domain of f into two sets: the set of all discontinuous points $\mathcal{D}$ and the remainder $\mathbb{R}\backslash\mathcal{D}$. By definition, z is in $\mathbb{R}\backslash\mathcal{D}$ with probability 1. We evaluate the expectation as $\mathbb{E}[f(z)] = \mathbb{E}[f(z)|z \in \mathcal{D}]P(z \in \mathcal{D}) + \mathbb{E}[f(z)|z \notin \mathcal{D}]P(z \notin \mathcal{D})$. The former term does not contribute to the expectation, because of $P(z \in \mathcal{D}) = 0$. The latter term can be bounded in the same manner as in [55, Lemma 1] since conditioning $z \notin \mathcal{D}$ does not affect the distribution of absolutely continuous z. According to this argument, we can generalize [55, Lemma 1] to a lemma for piecewise pseudo-Lipschitz functions. Thus, we regard [55, Lemma 1] as a result for piecewise pseudo-Lipschitz functions.

From Assumption 3 and the definition $\boldsymbol{b}_0[l] = \boldsymbol{A}[l]\boldsymbol{q}_0[l]$ in (71), using [55, Lemma 1] and Property (Ob) for $\tau = 0$ yields $\{\boldsymbol{b}_0[l]\} \overset{\mathrm{PL}(2)}{\to} \{B_0[l]\}$. Since the noise vectors $\{\boldsymbol{w}[l]\}$ are independent of $\{\boldsymbol{b}_0[l]\}$, we use Assumption 2 to arrive at the former convergence in Property (Oc) for $\tau = 0$.

We next prove the latter convergence in Property (Oc). In proving [12, Lemma 5], two properties of Lipschitz-continuous functions were used: almost everywhere differentiability and the boundedness of derivatives, which are satisfied for any piecewise Lipschitz-continuous function. Thus, [12, Lemma 5] is available for all piecewise Lipschitz-continuous functions. Since $f_{\mathrm{out}}[l](0, g[l](z, w); v_0[l])$ is a piecewise Lipschitz-continuous function of (z, w) with $v_0[l] = \bar{v}_0[l]$, we use [12, Lemma 5] and the former convergence in Property (Oc) for $\tau = 0$ to find that $\xi_{\mathrm{out},0}[l]$ in (4) converges almost surely to $\bar{\xi}_{\mathrm{out},0}[l]$ in (24). ∎

Proof of Property (Od): Under Assumption 4, we can use Property (Oc) for $\boldsymbol{m}_0[l] = f_{\mathrm{out}}[l](\boldsymbol{0}, g[l](\boldsymbol{z}[l], \boldsymbol{w}[l]); v_0[l])$ with $v_0[l] = \bar{v}_0[l]$ and $\boldsymbol{z}[l] = -\bar{\xi}_{\mathrm{out},0}[l]\bar{\zeta}_0^{-1}[l]\boldsymbol{b}_0[l]$ to obtain

$$\begin{aligned}
&\frac{1}{M[l]}\boldsymbol{b}_0^{\mathrm{T}}[l]\boldsymbol{m}_0[l] \overset{\mathrm{a.s.}}{\to} \mathbb{E}\left[B_0[l]f_{\mathrm{out}}[l]\left(0, g[l]\left(Z[l], W[l]\right); \bar{v}_0[l]\right)\right] \\
&= -\frac{\bar{\xi}_{\mathrm{out},0}[l]}{\bar{\zeta}_0[l]}\mathbb{E}\left[B_0[l]B_0[l]\right](-\bar{\zeta}_0[l]) = \frac{\bar{\xi}_{\mathrm{out},0}[l]}{L\delta[l]}\mathbb{E}[(Q_0[l])^2] \qquad (110)
\end{aligned}$$

for $Z[l] = -\bar{\xi}_{\mathrm{out},0}[l]\bar{\zeta}_0^{-1}[l]B_0[l]$, where the first and second equalities in (110) follow from Lemma 6 the definition of $\bar{\zeta}_0[l]$ in (25) and from Property (Oc) for $\tau = 0$, respectively. Thus, Property (Od) holds for $\tau = 0$. ∎

Proof of Property (Oe): See [55, Proof of Property (A4) for $\tau = 0$ in Theorem 4]. According to the argument in the proof of the former property in Property (Oc) for $\tau = 0$, we can generalize pseudo-Lipschitz functions in [55] to piecewise pseudo-Lipschitz functions. ∎

D. Inner Module for $\tau = 0$

Proof of Property (Ia): We evaluate the distribution of $\boldsymbol{h}_0[l]$ in (73) conditioned on $\mathfrak{E}_{1,0}$ and Θ. We use Lemma 5 under the constraints $\boldsymbol{b}_0[l] = \boldsymbol{A}[l]\boldsymbol{q}_0[l]$ for all $l \in \mathcal{L}$ to obtain

$$\boldsymbol{A}[l] \sim \frac{\boldsymbol{b}_0[l]\boldsymbol{q}_0^{\mathrm{T}}[l]}{\|\boldsymbol{q}_0[l]\|^2} - \tilde{\boldsymbol{A}}[l]\boldsymbol{P}_{\boldsymbol{q}_0[l]}^{\perp} \qquad (111)$$

conditioned on $\mathfrak{E}_{1,0}$ and Θ, in which $\{\tilde{\boldsymbol{A}}[l]\}$ are independent matrices and independent of $\{\mathfrak{E}_{1,0}, \Theta\}$. Each $\tilde{\boldsymbol{A}}[l]$

has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$. Applying this expression to $\boldsymbol{h}_0[l]$ in (73), we have

$$\boldsymbol{h}_0[l] \sim \frac{\xi_{\text{out},0}[l]}{L}\boldsymbol{q}_0[l] - \frac{\boldsymbol{b}_0^{\text{T}}[l]\boldsymbol{m}_0[l]}{\|\boldsymbol{q}_0[l]\|^2}\boldsymbol{q}_0[l] + \boldsymbol{P}_{\boldsymbol{q}_0[l]}^{\perp}\tilde{\boldsymbol{A}}^{\text{T}}[l]\boldsymbol{m}_0[l] \tag{112}$$

conditioned on $\mathfrak{E}_{1,0}$ and Θ. From the definition $\boldsymbol{P}_{\boldsymbol{q}_0[l]}^{\perp} = \boldsymbol{I} - \|\boldsymbol{q}_0[l]\|^{-2}\boldsymbol{q}_0[l]\boldsymbol{q}_0^{\text{T}}[l]$, we use [12, Lemma 3(c)] to have

$$\boldsymbol{P}_{\boldsymbol{q}_0[l]}^{\perp}\tilde{\boldsymbol{A}}^{\text{T}}[l]\boldsymbol{m}_0[l] \overset{\text{a.s.}}{=} \tilde{\boldsymbol{A}}^{\text{T}}[l]\boldsymbol{m}_0[l] + o(1)\boldsymbol{q}_0[l]. \tag{113}$$

To prove Property (Ia) for $\tau = 0$, it is sufficient to prove that $\boldsymbol{b}_0^{\text{T}}[l]\boldsymbol{m}_0[l]/\|\boldsymbol{q}_0[l]\|^2$ converges almost surely to $L^{-1}\xi_{\text{out},0}[l]$ in the large system limit. Using Property (Od) for $\tau = 0$ yields $N^{-1}\boldsymbol{b}_0^{\text{T}}[l]\boldsymbol{m}_0[l] \overset{\text{a.s.}}{\to} L^{-1}\bar{\xi}_{\text{out},0}[l]\mathbb{E}[Q_0^2[l]]$. Furthermore, Assumption 1 implies $N^{-1}\|\boldsymbol{q}_0[l]\|^2 \overset{\text{a.s.}}{\to} \mathbb{E}[(Q_0[l])^2]$. From these observations we find that $\boldsymbol{b}_0^{\text{T}}[l]\boldsymbol{m}_0[l]/\|\boldsymbol{q}_0[l]\|^2$ converges almost surely to $L^{-1}\bar{\xi}_{\text{out},0}[l]$ in the large system limit, which is almost surely equal to $L^{-1}\xi_{\text{out},0}[l]$, because of Property (Oc) for $\tau = 0$. Thus, Property (Ia) holds for $\tau = 0$. ■

Proof of Property (Ib): Property (Ia) for $\tau = 0$ implies the conditional independence of $\{\boldsymbol{h}_0[l] : l \in \mathcal{L}\}$ given $\mathfrak{E}_{1,0}$ and Θ in the large system limit. Thus, we use Property (Ia) for $\tau = 0$ and [49, Lemma 3] to obtain

$$\frac{1}{N}\boldsymbol{h}_0^{\text{T}}[l']\boldsymbol{h}_0[l] \overset{\text{a.s.}}{=} \frac{\delta_{l,l'}}{LM[l]}\|\boldsymbol{m}_0[l]\|^2 + o(1). \tag{114}$$

Thus, Property (Ib) holds for $\tau = 0$. ■

Proof of Property (Ic): We find the almost sure convergence $\sigma_0^2[l] \overset{\text{a.s.}}{\to} \bar{\sigma}_0^2[l]$ for (14) and (31), because of Property (Oc) for $\tau = 0$. We only prove the former convergence because the latter convergence $\xi_{\text{in},0}[l] \overset{\text{a.s.}}{\to} \bar{\xi}_{\text{in},0}[l]$ follows from the former convergence in Property (Ic) for $\tau = 0$, $\sigma_0^2[l] \overset{\text{a.s.}}{\to} \bar{\sigma}_0^2[l]$, and [12, Lemma 5]. Using Properties (Ia) and (Ib) for $\tau = 0$, Assumption 1, and [55, Lemma 1], we obtain the convergence $(\{\boldsymbol{h}_0[l]\}_{l\in\mathcal{L}}, \boldsymbol{x}) \overset{\text{PL}(2)}{\to} (\{H_0[l]\}_{l\in\mathcal{L}}, X)$, in which $\{H_0[l]\}$ are independent of X and zero-mean Gaussian random variables with covariance $\mathbb{E}[H_0[l]H_0[l']] = L^{-1}\delta_{l,l'}\mathbb{E}[M_0^2[l]]$. Thus, Property (Ic) holds for $\tau = 0$. ■

Proof of Property (Id): Using the definition of $\boldsymbol{q}_1[l]$ in (75) yields

$$\begin{aligned}
&\frac{1}{N}\boldsymbol{h}_0^{\text{T}}[l]\boldsymbol{q}_1[l] = \frac{1}{N}\boldsymbol{h}_0^{\text{T}}[l]f_{\text{in}}[l]\left(\frac{\bar{\eta}_0[l]}{L}\boldsymbol{x} + \tilde{\boldsymbol{h}}_0[l]; \eta_0[l], \sigma_0^2[l]\right) \\
&- \frac{\bar{\zeta}_1[l]}{\bar{\xi}_{\text{out},1}[l]}\frac{1}{N}\boldsymbol{h}_0^{\text{T}}[l]\boldsymbol{x} \\
&\overset{\text{a.s.}}{\to} \mathbb{E}\left[H_0[l]f_{\text{in}}[l]\left(\frac{\bar{\eta}_0[l]}{L}X + \tilde{H}_0[l]; \eta_t[l], \bar{\sigma}_0^2[l]\right)\right] \\
&- \frac{\bar{\zeta}_1[l]}{\bar{\xi}_{\text{out},1}[l]}\mathbb{E}[H_0[l]X],
\end{aligned} \tag{115}$$

where the last follows from the definition of $\tilde{\boldsymbol{h}}_0[l]$ in (74) and Property (Ic) for $\tau = 0$. Since we have $\mathbb{E}[H_0[l]X] = \mathbb{E}[H_0[l]]\mathbb{E}[X] = 0$ in the second term, we use the definition of $\tilde{H}_0[l]$ in (92) and Lemma 6 to evaluate the first term as

$$\begin{aligned}
&\mathbb{E}\left[H_0[l]f_{\text{in}}[l](L^{-1}\bar{\eta}_0[l]X + \tilde{H}_0[l]; \eta_0[l], \bar{\sigma}_0^2[l])\right] \\
&= \frac{\bar{\xi}_{\text{in},0}[l]}{\bar{\xi}_{\text{out},0}[l]}C_{0,0}[l][l]\mathbb{E}[H_0^2[l]] \\
&+ \bar{\xi}_{\text{in},0}[l]\sum_{l'\neq l}\frac{C_{0,0}[l][l']}{\bar{\xi}_{\text{out},0}[l']}\mathbb{E}[H_0[l]H_0[l']] \\
&= C_{0,0}[l][l]\frac{\bar{\xi}_{\text{in},0}[l]}{L\bar{\xi}_{\text{out},0}[l]}\mathbb{E}[M_0^2[l]],
\end{aligned} \tag{116}$$

where the last equality follows from Property (Ib) for $\tau = 0$. Thus, we arrive at Property (Id) for $\tau = 0$. ■

Proof of Property (Ie): See [55, Proof of Property (B4) for $\tau = 0$ in Theorem 4]. ■

E. Outer Module for $\tau > 0$

For some $t \in \mathbb{N}$, suppose that Properties (Oa)–(Oe) and Properties (Ia)–(Ie) are correct for all $\tau < t$. We prove Properties (Oa)–(Oe) for $\tau = t$ under these induction hypotheses.

Proof of Property (Oa): Let

$$\boldsymbol{\Lambda}_{\text{in},t}[l] = \text{diag}\left\{\frac{\xi_{\text{in},\tau}[l]}{L\delta[l]\xi_{\text{out},\tau}[l]} : \tau \in \mathcal{T}_t[l]\right\}, \tag{117}$$

$$\underline{\boldsymbol{\Lambda}}_{\text{in},t}[l] = (C_{t,0}[l][l] - 1)\text{diag}\left\{\frac{\xi_{\text{in},\tau}[l]}{L\delta[l]\underline{\xi}_{\text{out},\tau}[l]} : \tau \in \mathcal{T}_t[l]\right\}, \tag{118}$$

$$\boldsymbol{\Lambda}_{\text{out},t}[l] = \text{diag}\left\{\frac{\xi_{\text{out},\tau}[l]}{L} : \tau \in \mathcal{T}_t[l]\right\}. \tag{119}$$

From the definitions of $\boldsymbol{b}_t[l]$ and $\boldsymbol{h}_t[l]$ in (71) and (73), $\boldsymbol{A}[l]$ conditioned on $\mathfrak{E}_{t,t}$ and Θ satisfies the following constraints just before updating $\boldsymbol{b}_t[l]$ for all $l \in \mathcal{L}$:

$$\begin{aligned}
&\boldsymbol{B}_t[l] - \left[\boldsymbol{0}, \boldsymbol{M}_{t-1}[l]\boldsymbol{\Lambda}_{\text{in},t-1}[l] + \underline{\boldsymbol{M}}_{t-1}[l]\underline{\boldsymbol{\Lambda}}_{\text{in},t-1}[l]\right] \\
&= \boldsymbol{A}[l]\boldsymbol{Q}_t[l],
\end{aligned} \tag{120}$$

$$\boldsymbol{Q}_t[l]\boldsymbol{\Lambda}_{\text{out},t}[l] - \boldsymbol{H}_t[l] = \boldsymbol{A}^{\text{T}}[l]\boldsymbol{M}_t[l] \tag{121}$$

for all $l \in \mathcal{L}$.

We let $\boldsymbol{U}[l] = \boldsymbol{Q}_t[l]$ and $\boldsymbol{V}[l] = \boldsymbol{M}_t[l]$ in Lemma 5. Since the induction hypotheses (Oe) and (Ie) for $\tau = t - 1$ imply that $\boldsymbol{M}_t[l]$ and $\boldsymbol{Q}_t[l]$ have full rank, we can use Lemma 5 to obtain

$$\begin{aligned}
\boldsymbol{A}[l] \sim &- \left[\boldsymbol{0}, \boldsymbol{M}_{t-1}[l]\boldsymbol{\Lambda}_{\text{in},t-1}[l] - \underline{\boldsymbol{M}}_{t-1}[l]\underline{\boldsymbol{\Lambda}}_{\text{in},t-1}[l]\right]\boldsymbol{Q}_t^{\dagger}[l] \\
&+ \boldsymbol{B}_t[l]\boldsymbol{Q}_t^{\dagger}[l] - (\boldsymbol{M}_t^{\dagger}[l])^{\text{T}}\boldsymbol{H}_t^{\text{T}}[l]\boldsymbol{P}_{\boldsymbol{Q}_t[l]}^{\perp} \\
&+ \boldsymbol{P}_{\boldsymbol{M}_t[l]}^{\perp}\tilde{\boldsymbol{A}}[l]\boldsymbol{P}_{\boldsymbol{Q}_t[l]}^{\perp}
\end{aligned} \tag{122}$$

conditioned on $\mathfrak{E}_{t,t}$ and Θ, where we have used $\boldsymbol{Q}_t^{\text{T}}[l]\boldsymbol{P}_{\boldsymbol{Q}_t[l]}^{\perp} = \boldsymbol{O}$. Here, $\{\tilde{\boldsymbol{A}}[l]\}$ are independent matrices and independent of $\{\mathfrak{E}_{t,t}, \Theta\}$. Each $\tilde{\boldsymbol{A}}[l]$ has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$.

Let $\boldsymbol{\beta}_t[l] = \boldsymbol{Q}_t^{\dagger}[l]\boldsymbol{q}_t[l]$ and $\boldsymbol{q}_t^{\perp}[l] = \boldsymbol{P}_{\boldsymbol{Q}_t[l]}^{\perp}\boldsymbol{q}_t[l]$. Furthermore, we write the τth element of $\boldsymbol{\beta}_t[l]$ as $\beta_{t,\tau}[l]$. Substituting the expression of $\boldsymbol{A}[l]$ into the definition of $\boldsymbol{b}_t[l]$ in (71) yields

$$\begin{aligned}\boldsymbol{b}_t[l] &\sim \boldsymbol{B}_t[l]\boldsymbol{\beta}_t[l] - \sum_{\tau\in\mathcal{T}_t[l]\setminus\{0\}} \frac{\beta_{t,\tau}[l]\xi_{\mathrm{in},\tau-1}[l]}{L\delta[l]\xi_{\mathrm{out},\tau-1}[l]}\boldsymbol{m}_{\tau-1}[l] \\ &- (C_{t,0}[l][l]-1)\sum_{\tau\in\mathcal{T}_t[l]\setminus\{0\}} \beta_{t,\tau}[l]\frac{\xi_{\mathrm{in},\tau-1}[l]}{L\delta[l]\underline{\xi}_{\mathrm{out},\tau-1}[l]}\underline{\boldsymbol{m}}_{\tau-1}[l] \\ &- (\boldsymbol{M}_t^{\dagger}[l])^{\mathrm{T}}\boldsymbol{H}_t^{\mathrm{T}}[l]\boldsymbol{q}_t^{\perp}[l] + \boldsymbol{P}_{\boldsymbol{M}_t[l]}^{\perp}\tilde{\boldsymbol{A}}[l]\boldsymbol{q}_t^{\perp}[l] \\ &+ \frac{\xi_{\mathrm{in},t-1}[l]}{L\delta[l]}\left\{\frac{\boldsymbol{m}_{t-1}[l]}{\xi_{\mathrm{out},t-1}[l]} + \frac{C_{t,0}[l][l]-1}{\underline{\xi}_{\mathrm{out},t-1}[l]}\underline{\boldsymbol{m}}_{t-1}[l]\right\} \end{aligned} \tag{123}$$

conditioned on $\mathfrak{E}_{t,t}$ and Θ.

We simplify the expression of $\boldsymbol{b}_t[l]$. Using $\boldsymbol{q}_t^{\perp}[l] = \boldsymbol{q}_t[l] - \boldsymbol{Q}_t[l]\boldsymbol{\beta}_t[l]$ and the induction hypothesis (Id) for $\tau = t-1$ yields

$$\begin{aligned}&(\boldsymbol{M}_t^{\dagger}[l])^{\mathrm{T}}\boldsymbol{H}_t^{\mathrm{T}}[l]\boldsymbol{q}_t^{\perp}[l] \\ &\overset{\mathrm{a.s.}}{=} \frac{\bar{\xi}_{\mathrm{in},t-1}[l]}{L\delta[l]}\left\{\frac{\boldsymbol{m}_{t-1}[l]}{\bar{\xi}_{\mathrm{out},t-1}[l]} + \frac{C_{t,0}[l][l]-1}{\bar{\underline{\xi}}_{\mathrm{out},t-1}[l]}\underline{\boldsymbol{m}}_{t-1}[l]\right\} \\ &+ \boldsymbol{M}_t[l]\boldsymbol{o}(1) - \sum_{\tau\in\mathcal{T}_t[l]\setminus\{0\}} \frac{\beta_{t,\tau}[l]\bar{\xi}_{\mathrm{in},\tau-1}[l]}{L\delta[l]\bar{\xi}_{\mathrm{out},\tau-1}[l]}\boldsymbol{m}_{\tau-1}[l] \\ &- (C_{t,0}[l][l]-1)\sum_{\tau\in\mathcal{T}_t[l]\setminus\{0\}} \frac{\beta_{t,\tau}[l]\bar{\xi}_{\mathrm{in},\tau-1}[l]}{L\delta[l]\bar{\underline{\xi}}_{\tau-1}[l]}\underline{\boldsymbol{m}}_{\tau-1}[l]. \end{aligned} \tag{124}$$

Substituting this expression into the representation of $\boldsymbol{b}_t[l]$ in (123) and using the induction hypotheses (Oc) and (Ic) for all $\tau < t$, we arrive at

$$\boldsymbol{b}_t[l] \sim \boldsymbol{B}_t[l]\boldsymbol{\beta}_t[l] + \boldsymbol{M}_t[l]\boldsymbol{o}(1) + \boldsymbol{P}_{\boldsymbol{M}_t[l]}^{\perp}\tilde{\boldsymbol{A}}[l]\boldsymbol{q}_t^{\perp}[l] \tag{125}$$

conditioned on $\mathfrak{E}_{t,t}$ and Θ, which is equivalent to Property (Oa) for $\tau = t$, because $\boldsymbol{P}_{\boldsymbol{M}_t[l]}^{\perp}\tilde{\boldsymbol{A}}[l]\boldsymbol{q}_t^{\perp}[l] = \tilde{\boldsymbol{A}}[l]\boldsymbol{q}_t^{\perp}[l] + \boldsymbol{M}_t[l]\boldsymbol{o}(1)$ holds due to [12, Lemma 3(c)]. ■

Proof of Property (Ob): For all $\tau' < t$, using Property (Oa) for $\tau = t$ yields

$$\begin{aligned}\frac{1}{M[l]}\boldsymbol{b}_{\tau'}^{\mathrm{T}}[l']\boldsymbol{b}_t[l] &\overset{\mathrm{a.s.}}{=} \frac{1}{M[l]}\boldsymbol{b}_{\tau'}^{\mathrm{T}}[l']\boldsymbol{B}_t[l]\boldsymbol{\beta}_t[l] + o(1) \\ &\overset{\mathrm{a.s.}}{=} \frac{\delta_{l,l'}}{NL\delta[l]}\boldsymbol{q}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{Q}_t[l]\boldsymbol{\beta}_t[l] + o(1) \\ &= \frac{\delta_{l,l'}}{NL\delta[l]}\boldsymbol{q}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{q}_t[l] + o(1), \end{aligned} \tag{126}$$

where the second and last equalities follow from the induction hypothesis (Ob) for $\tau < t$ and $\boldsymbol{\beta}_t[l] = \boldsymbol{Q}_t^{\dagger}[l]\boldsymbol{q}_t[l]$, respectively.

For $\tau' = t$ we use Property (Oa) for $\tau = t$ and [49, Lemma 3] to obtain

$$\begin{aligned}\frac{\boldsymbol{b}_t^{\mathrm{T}}[l']\boldsymbol{b}_t[l]}{M[l]} &\overset{\mathrm{a.s.}}{=} \frac{1}{M[l]}\boldsymbol{\beta}_t^{\mathrm{T}}[l']\boldsymbol{B}_t^{\mathrm{T}}[l']\boldsymbol{B}_t[l]\boldsymbol{\beta}_t[l] \\ &+ \frac{\delta_{l,l'}}{NL\delta[l]}\|\boldsymbol{q}_t^{\perp}[l]\|^2 + o(1) \\ &\overset{\mathrm{a.s.}}{=} \frac{\delta_{l,l'}}{NL\delta[l]}(\|\boldsymbol{q}_t^{\parallel}[l]\|^2 + \|\boldsymbol{q}_t^{\perp}[l]\|^2) + o(1) \\ &\overset{\mathrm{a.s.}}{=} \frac{\delta_{l,l'}}{NL\delta[l]}\|\boldsymbol{q}_t[l]\|^2 + o(1), \end{aligned} \tag{127}$$

with $\boldsymbol{q}_t^{\parallel}[l] = \boldsymbol{P}_{\boldsymbol{Q}_t[l]}^{\parallel}\boldsymbol{q}_t[l]$, where the second equality follows from the induction hypothesis (Ob) for $\tau < t$ and the definition $\boldsymbol{\beta}_t[l] = \boldsymbol{Q}_t^{\dagger}[l]\boldsymbol{q}_t[l]$. Thus, Property (Ob) holds for $\tau = t$. ■

Proof of Property (Oc): The former convergence in Property (Oc) for $\tau = t$ follows from Property (Ob) for $\tau = t$, [55, Lemma 1], and Assumption 2. We find the convergence $v_t[l] \overset{\mathrm{a.s.}}{\to} \bar{v}_t[l]$ for (17) and (34) from the induction hypotheses (Oc) and (Ic) for all $\tau < t$. The latter convergence $\xi_{\mathrm{out},t}[l] \overset{\mathrm{a.s.}}{\to} \bar{\xi}_{\mathrm{out},t}[l]$ follows from the former convergence in Property (Oc) for $\tau = t$, $v_t[l] \overset{\mathrm{a.s.}}{\to} \bar{v}_t[l]$, and [12, Lemma 5]. ■

Proof of Property (Od): Using the definition of $\boldsymbol{m}_t[l]$ in (72) and Property (Oc) for $\tau = t$ yields

$$\begin{aligned}&\frac{1}{M[l]}\boldsymbol{b}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{m}_t[l] \\ &\overset{\mathrm{a.s.}}{\to} \mathbb{E}\left[B_{\tau'}[l]f_{\mathrm{out}}[l]\left(B_t[l] + \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]}Z[l], Y[l]; \bar{v}_t[l]\right)\right] \\ &= \bar{\xi}_{\mathrm{out},t}[l]\mathbb{E}[B_{\tau'}[l]B_t[l]] \\ &+ \left(\frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]}\bar{\xi}_{\mathrm{out},t}[l] - \bar{\zeta}_t[l]\right)\mathbb{E}[B_{\tau'}[l]Z[l]] \\ &= \frac{\bar{\xi}_{\mathrm{out},t}[l]}{L\delta[l]}\mathbb{E}[Q_{\tau'}[l]Q_t[l]], \end{aligned} \tag{128}$$

with $\bar{\zeta}_t[l]$ given in (25), where the first and second equalities follow from Lemma 6 and $Y[l] = g[l](Z[l], W[l])$ and from Property (Oc) for $\tau = t$, respectively. Thus, Property (Od) holds for $\tau = t$. ■

Proof of Property (Oe): See [55, Proof of Property (A4) for $\tau = t$ in Theorem 4]. ■

F. Inner Module for $\tau > 0$

Suppose that Properties (Oa)–(Oe) and (Ia)–(Ie) for all $\tau < t$ are correct. To complete the proof of Theorem 6 by induction, we prove Properties (Ia)–(Ie) for $\tau = t$ under these induction hypotheses, as well as Properties (Oa)–(Oe) for $\tau = t$.

Proof of Property (Ia): From the definitions of $\boldsymbol{b}_t[l]$ and $\boldsymbol{h}_t[l]$ in (71) and (73), $\boldsymbol{A}[l]$ conditioned on $\mathfrak{E}_{t+1,t}$ and Θ satisfies the following constraints just before updating $\boldsymbol{h}_t[l]$:

$$\boldsymbol{B}_{t+1}[l] - \left[\boldsymbol{0}, \boldsymbol{M}_t[l]\boldsymbol{\Lambda}_{\mathrm{in},t}[l] + \underline{\boldsymbol{M}}_t[l]\underline{\boldsymbol{\Lambda}}_{\mathrm{in},t}[l]\right] = \boldsymbol{A}[l]\boldsymbol{Q}_{t+1}[l], \tag{129}$$

$$\boldsymbol{Q}_t[l]\boldsymbol{\Lambda}_{\mathrm{out},t}[l] - \boldsymbol{H}_t[l] = \boldsymbol{A}^{\mathrm{T}}[l]\boldsymbol{M}_t[l] \tag{130}$$

for all $l \in \mathcal{L}$. Since the induction hypotheses (Oe) and (Ie) for $\tau = t-1$ imply that $\boldsymbol{M}_t[l]$ and $\boldsymbol{Q}_{t+1}[l]$ have full rank, we use Lemma 5 to obtain

$$\begin{aligned}\boldsymbol{A}[l] &\sim (\boldsymbol{M}_t^{\dagger}[l])^{\mathrm{T}}(\boldsymbol{Q}_t[l]\boldsymbol{\Lambda}_{\mathrm{out},t}[l] - \boldsymbol{H}_t[l])^{\mathrm{T}} \\ &+ \boldsymbol{P}_{\boldsymbol{M}_t[l]}^{\perp}\boldsymbol{B}_{t+1}[l]\boldsymbol{Q}_{t+1}^{\dagger}[l] - \boldsymbol{P}_{\boldsymbol{M}_t[l]}^{\perp}\tilde{\boldsymbol{A}}[l]\boldsymbol{P}_{\boldsymbol{Q}_{t+1}[l]}^{\perp} \end{aligned} \tag{131}$$

conditioned on $\mathfrak{E}_{t+1,t}$ and Θ. Here, $\{\tilde{\boldsymbol{A}}[l]\}$ are independent matrices and independent of $\{\mathfrak{E}_{t+1,t}, \Theta\}$. Each $\tilde{\boldsymbol{A}}[l]$ has independent zero-mean Gaussian elements with variance $(LM[l])^{-1}$.

Let $\boldsymbol{\alpha}_t[l] = \boldsymbol{M}_t^{\dagger}[l]\boldsymbol{m}_t[l]$ and $\boldsymbol{m}_t^{\perp}[l] = \boldsymbol{P}_{\boldsymbol{M}_t[l]}^{\perp}\boldsymbol{m}_t[l]$. Substituting the expression of $\boldsymbol{A}[l]$ into the definition of $\boldsymbol{h}_t[l]$ in (73), we have

$$\boldsymbol{h}_t[l] \sim \boldsymbol{H}_t[l]\boldsymbol{\alpha}_t[l] + \boldsymbol{h}_t^{\mathrm{bias}}[l] + \boldsymbol{P}_{\boldsymbol{Q}_{t+1}[l]}^{\perp}\tilde{\boldsymbol{A}}^{\mathrm{T}}[l]\boldsymbol{m}_t^{\perp}[l] \tag{132}$$

conditioned on $\mathfrak{E}_{t+1,t}$ and Θ, with

$$\boldsymbol{h}_t^{\mathrm{bias}}[l] = \frac{\xi_{\mathrm{out},t}[l]}{L}\boldsymbol{q}_t[l] - \boldsymbol{Q}_t[l]\boldsymbol{\Lambda}_{\mathrm{out},t}[l]\boldsymbol{\alpha}_t[l] - (\boldsymbol{Q}_{t+1}^{\dagger}[l])^{\mathrm{T}}\boldsymbol{B}_{t+1}^{\mathrm{T}}[l]\boldsymbol{m}_t[l] + (\boldsymbol{Q}_{t+1}^{\dagger}[l])^{\mathrm{T}}\boldsymbol{B}_{t+1}^{\mathrm{T}}[l]\boldsymbol{M}_t[l]\boldsymbol{\alpha}_t[l], \tag{133}$$

where we have used $\boldsymbol{m}_t^{\perp}[l] = \boldsymbol{m}_t[l] - \boldsymbol{M}_t[l]\boldsymbol{\alpha}_t[l]$. In particular, we use [12, Lemma 3(c)] to obtain

$$\boldsymbol{P}_{\boldsymbol{Q}_{t+1}[l]}^{\perp}\tilde{\boldsymbol{A}}^{\mathrm{T}}[l]\boldsymbol{m}_t^{\perp}[l] \overset{\mathrm{a.s.}}{=} \tilde{\boldsymbol{A}}^{\mathrm{T}}[l]\boldsymbol{m}_t^{\perp}[l] + \boldsymbol{Q}_{t+1}[l]\boldsymbol{o}(1). \tag{134}$$

Thus, it is sufficient to prove $\boldsymbol{h}_t^{\mathrm{bias}}[l] \overset{\mathrm{a.s.}}{=} \boldsymbol{Q}_{t+1}[l]\boldsymbol{o}(1)$.

We evaluate $\boldsymbol{h}_t^{\mathrm{bias}}[l]$. Using Property (Od) for $\tau = t$ yields

$$\begin{aligned}&(\boldsymbol{Q}_{t+1}^{\dagger}[l])^{\mathrm{T}}\boldsymbol{B}_{t+1}^{\mathrm{T}}[l]\boldsymbol{m}_t[l] \\ &\overset{\mathrm{a.s.}}{=} \frac{\bar{\xi}_{\mathrm{out},t}[l]}{L}(\boldsymbol{Q}_{t+1}^{\dagger}[l])^{\mathrm{T}}\boldsymbol{Q}_{t+1}^{\mathrm{T}}[l]\boldsymbol{q}_t[l] + \boldsymbol{Q}_{t+1}[l]\boldsymbol{o}(1) \\ &= \frac{\bar{\xi}_{\mathrm{out},t}[l]}{L}\boldsymbol{q}_t[l] + \boldsymbol{Q}_{t+1}[l]\boldsymbol{o}(1).\end{aligned} \tag{135}$$

Similarly, we obtain

$$\begin{aligned}&(\boldsymbol{Q}_{t+1}^{\dagger}[l])^{\mathrm{T}}\boldsymbol{B}_{t+1}^{\mathrm{T}}[l]\boldsymbol{M}_t[l]\boldsymbol{\alpha}_t[l] \\ &= \sum_{\tau\in\mathcal{T}_t[l]}\alpha_{t,\tau}[l](\boldsymbol{Q}_{t+1}^{\dagger}[l])^{\mathrm{T}}\boldsymbol{B}_{t+1}^{\mathrm{T}}[l]\boldsymbol{m}_{\tau}[l] \\ &\overset{\mathrm{a.s.}}{=} \sum_{\tau\in\mathcal{T}_t[l]}\alpha_{t,\tau}[l]\frac{\bar{\xi}_{\mathrm{out},\tau}[l]}{L}\boldsymbol{q}_{\tau}[l] + \boldsymbol{Q}_{t+1}[l]\boldsymbol{o}(1),\end{aligned} \tag{136}$$

with $\alpha_{t,\tau}[l]$ denoting the τth element of $\boldsymbol{\alpha}_t[l]$. Substituting these expressions into the definition of $\boldsymbol{h}_t^{\mathrm{bias}}[l]$ in (133) and using Property (Oc) for all $\tau \leq t$, we arrive at $\boldsymbol{h}_t^{\mathrm{bias}}[l] \overset{\mathrm{a.s.}}{=} \boldsymbol{Q}_{t+1}[l]\boldsymbol{o}(1)$. Thus, Property (Ia) holds for $\tau = t$. ∎

Proof of Property (Ib): Property (Ia) for $\tau = t$ implies the conditional independence of $\{\boldsymbol{h}_t[l] : l \in \mathcal{L}\}$ given $\mathfrak{E}_{t+1,t}$ and Θ. We first consider the case of $\tau' < t$. Using Property (Ia) for $\tau = t$ and the induction hypothesis (Ib) for $\tau < t$ yields

$$\begin{aligned}\frac{1}{N}\boldsymbol{h}_{\tau'}^{\mathrm{T}}[l']\boldsymbol{h}_t[l] &\overset{\mathrm{a.s.}}{=} \frac{\delta_{l,l'}}{LM[l]}\boldsymbol{m}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{M}_t[l]\boldsymbol{\alpha}_t[l] + o(1) \\ &= \frac{\delta_{l,l'}}{LM[l]}\boldsymbol{m}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{m}_t[l] + o(1)\end{aligned} \tag{137}$$

for all $\tau' < t$, where the last equality follows from the definition $\boldsymbol{\alpha}_t[l] = \boldsymbol{M}_t^{\dagger}[l]\boldsymbol{m}_t[l]$.

We next consider the case of $\tau' = t$. Using Property (Ia) for $\tau = t$ and the induction hypothesis (Ib) for $\tau < t$ yields

$$\begin{aligned}\frac{\boldsymbol{h}_t^{\mathrm{T}}[l']\boldsymbol{h}_t[l]}{N} &\overset{\mathrm{a.s.}}{=} \frac{\delta_{l,l'}}{LM[l]}\boldsymbol{\alpha}_t^{\mathrm{T}}[l]\boldsymbol{M}_t^{\mathrm{T}}[l]\boldsymbol{M}_t[l]\boldsymbol{\alpha}_t[l] \\ &\quad + \frac{\delta_{l,l'}}{LM[l]}\|\boldsymbol{m}_t^{\perp}[l]\|^2 + o(1) \\ &= \frac{\delta_{l,l'}}{LM[l]}\left(\|\boldsymbol{P}_{\boldsymbol{M}_t[l]}^{\parallel}\boldsymbol{m}_t[l]\|^2 + \|\boldsymbol{m}_t^{\perp}[l]\|^2\right) + o(1) \\ &= \frac{\delta_{l,l'}}{LM[l]}\|\boldsymbol{m}_t[l]\|^2 + o(1).\end{aligned} \tag{138}$$

Thus, Property (Ib) holds for $\tau = t$. ∎

Proof of Property (Ic): The former convergence in Property (Ic) for $\tau = t$ follows from Assumption 1, Properties (Ia) and (Ib) for $\tau = t$, and [55, Lemma 1]. We find the convergence $\sigma_t^2[l] \overset{\mathrm{a.s.}}{\to} \bar{\sigma}_t^2[l]$ for (14) and (31) from Property (Oc) for $\tau = t$. The latter convergence $\xi_{\mathrm{in},t}[l] \overset{\mathrm{a.s.}}{\to} \bar{\xi}_{\mathrm{in},t}[l]$ is obtained from the former convergence in Property (Ic) for $\tau = t$, $\sigma_t^2[l] \overset{\mathrm{a.s.}}{\to} \bar{\sigma}_t^2[l]$, and [12, Lemma 5]. ∎

Proof of Property (Id): Using the definition of $\boldsymbol{q}_{t+1}[l]$ in (75) and Property (Ic) for $\tau = t$ yields

$$\begin{aligned}\frac{\boldsymbol{h}_{\tau'}^{\mathrm{T}}[l]\boldsymbol{q}_{t+1}[l]}{N} &\overset{\mathrm{a.s.}}{\to} \mathbb{E}\left[H_{\tau'}[l]f_{\mathrm{in}}[l]\left(\frac{\bar{\eta}_t}{L}X + \tilde{H}_t[l]; \eta_t[l], \bar{\sigma}_t^2[l]\right)\right] \\ &= \bar{\xi}_{\mathrm{in},t}[l]\mathbb{E}[H_{\tau'}[l]\tilde{H}_t[l]] + o(1) \\ &= \bar{\xi}_{\mathrm{in},t}[l]\mathbb{E}\left[H_{\tau'}[l]\left\{\frac{H_t[l]}{\bar{\xi}_{\mathrm{out},t}[l]} + \frac{C_{t,0}[l][l]-1}{\underline{\bar{\xi}}_{\mathrm{out},t}[l]}\underline{H}_t[l]\right\}\right] \\ &\quad + \bar{\xi}_{\mathrm{in},t}[l]\sum_{i=0}^{\lfloor t/T\rfloor}\sum_{l'\neq l}\frac{C_{t,i}[l][l']}{\underline{\bar{\xi}}_{\mathrm{out},t-iT}[l']}\mathbb{E}[H_{\tau'}[l]\underline{H}_{t-iT}[l']] + o(1),\end{aligned} \tag{139}$$

where the first and second equalities follow from Lemma 6 and the definition of $\tilde{H}_t[l]$ in (92), respectively. We use Property (Ib) for $\tau \leq t$ to arrive at Property (Id) for $\tau = t$. ∎

Proof of Property (Ie): See [55, Proof of Property (B4) for $\tau = t$ in Theorem 4]. ∎

Appendix B
Proof of Theorem 2

The tree assumption in Assumption 5 is used in the proof of Theorem 2. Focus a root node $l_0 \in \mathcal{L}$ and consider message-passing that aggregate messages from leaf nodes toward the root node l_0. We analyze properties of message-passing with respect to a path that passes through the root node l_0.

Let $(l, l', \tilde{l}', \tilde{l})$ denote four different nodes l, l', $\tilde{l}'$, and $\tilde{l}$ that are located in this order on a path. We first prove the following lemma for the path length longer than or equal to 3:

Lemma 7: For four different nodes $(l, l', \tilde{l}', \tilde{l})$ on a path,

$$\mathbb{E}\left[\underline{H}_{\tau,j}[l \to l']\underline{H}_{\tau',j}[\tilde{l} \to \tilde{l}']\right] = 0 \tag{140}$$

with all $\tau' \in \{0, \ldots, \tau\}$.

Proof: The proof is by induction with respect to the total number of inner iterations for consensus propagation. For the first inner iteration $\tau = 0$ and $j = 1$, the definition of $\underline{H}_{0,1}[l \to l']$ in (28) implies $H_{0,1}[l \to l'] = \underline{\bar{\xi}}_{\mathrm{out},0}^{-1}[l]H_0[l]$. Thus, we use Property (Ic) in Theorem 6 to obtain (140) for $\tau = 0$ and $j = 1$ due to $l \neq \tilde{l}$. For some integers $t \geq T$ and $i \leq J + 1$, suppose that (140) is correct for $\tau = t$ and $j = i - 1$. If $i \leq J$ holds, we need to prove (140) for $\tau = t$ and $j = i$. Otherwise, we need to prove (140) for $\tau = t + 1$ and $j = 1$.

We only consider the case of $\tau = t$ and $j = i$ since (140) for $\tau = t + 1$ and $j = 1$ can be proved in the same manner. Since we have a path that connect the nodes l, l', $\tilde{l}'$, and $\tilde{l}$, the tree assumption in Assumption 5 implies that $\underline{H}_{\tau,i-1}[l'' \to l]$ does not contain the message $\underline{H}_{\tau'}[\tilde{l}]/\underline{\bar{\xi}}_{\mathrm{out},\tau'}[\tilde{l}]$ computed in node $\tilde{l}$ for any $l'' \in \mathcal{N}[l]\backslash\{l'\}$ and that $\underline{H}_{\tau',i-1}[\tilde{l}'' \to \tilde{l}]$ does not include the message $\underline{H}_{\tau}[l]/\underline{\bar{\xi}}_{\mathrm{out},\tau}[l]$ computed in node l

for any $\tilde{l}'' \in \mathcal{N}[\tilde{l}]\backslash\{l'\}$. Using the definition of $\underline{H}_{t,i}[l \to l']$ in (28) and Property (Ic) in Theorem 6 yields

$$\begin{aligned}&\mathbb{E}\left[\underline{H}_{t,i}[l \to l']\underline{H}_{\tau',i}[\tilde{l} \to \tilde{l}']\right] \\ &= \sum_{l''\in\mathcal{N}[l]\backslash\{l'\}} \sum_{\tilde{l}''\in\mathcal{N}[\tilde{l}]\backslash\{\tilde{l}'\}} \mathbb{E}\left[\underline{H}_{t,i-1}[l'' \to l]\underline{H}_{\tau',i-1}[\tilde{l}'' \to \tilde{l}]\right].\end{aligned} \tag{141}$$

Since Assumption 5 implies $\mathcal{N}[l] \cap \mathcal{N}[\tilde{l}] = \emptyset$ for the nodes $(l, l', \tilde{l}', \tilde{l})$, we have $l'' \neq \tilde{l}''$ for the indices of the summation. For four nodes $(l, l', \tilde{l}', \tilde{l}) = (l'', l, \tilde{l}, \tilde{l}'')$ on a path, we use the induction hypothesis (140) for $\tau = t$ and $j = i - 1$ to obtain (140) for $\tau = t$ and $j = i$. Thus, (140) is correct for all τ and j. ∎

We use Lemma 7 to prove the following lemma for a path of length 2.

Lemma 8: For three different nodes l, l', and $\tilde{l}$ that are located in this order on a path of length 2, we have

$$\mathbb{E}\left[\underline{H}_{t,j}[l \to l']\underline{H}_{\tau,j}[\tilde{l} \to l']\right] = 0 \tag{142}$$

for all $\tau \in \{0, \dots, t\}$.

Proof: Since Assumption 5 implies that $\underline{H}_{t,j-1}[\tilde{l}' \to l]$ does not contain $\underline{H}_t[\tilde{l}]/\bar{\underline{\xi}}_{\mathrm{out},t}[\tilde{l}]$ for any $\tilde{l}' \in \mathcal{N}[l]\backslash\{l'\}$ and that $\underline{H}_{t,j-1}[\tilde{l}' \to \tilde{l}]$ does not include $\underline{H}_t[l]/\bar{\underline{\xi}}_{\mathrm{out},t}[l]$ for $\tilde{l}' \in \mathcal{N}[\tilde{l}]\backslash\{l'\}$, we use the definition of $\underline{H}_{t,j}[l \to l']$ in (28) and Property (Ib) in Theorem 6 to obtain

$$\begin{aligned}&\mathbb{E}\left[\underline{H}_{t,j}[l \to l']\underline{H}_{\tau,j}[\tilde{l} \to l']\right] \\ &= \sum_{l''\in\mathcal{N}[l]\backslash\{l'\}} \sum_{\tilde{l}''\in\mathcal{N}[\tilde{l}]\backslash\{l'\}} \mathbb{E}\left[\underline{H}_{t,j-1}[l'' \to l]\underline{H}_{\tau,j-1}[\tilde{l}'' \to \tilde{l}]\right].\end{aligned} \tag{143}$$

Assumption 5 implies $\mathcal{N}[l] \cap \mathcal{N}[\tilde{l}] = \{l'\}$, so that we have $l'' \neq \tilde{l}''$ for the indices of the summation. For four nodes $(l, l', \tilde{l}', \tilde{l}) = (l'', l, \tilde{l}, \tilde{l}'')$ on a path, we use Lemma 7 to arrive at (142). ∎

We prove Theorem 2. Since Assumption 5 implies that $\underline{H}_{t,J}[l' \to l]$ does not contain the message $H_t[l]/\bar{\xi}_{\mathrm{out},t}[l]$ computed in node l, we use the definition of $\tilde{H}_t[l]$ in (27) and Lemma 8 yields

$$\begin{aligned}\mathbb{E}[\tilde{H}_\tau[l]\tilde{H}_t[l]] &= \frac{\mathbb{E}[M_\tau[l]M_t[l]]}{L\bar{\xi}_{\mathrm{out},\tau}[l]\bar{\xi}_{\mathrm{out},t}[l]} \\ &+ \sum_{l'\in\mathcal{N}[l]} \mathbb{E}\left[\underline{H}_{\tau,J}[l' \to l]\underline{H}_{t,J}[l' \to l]\right].\end{aligned} \tag{144}$$

Similarly, for $\underline{H}_{t,j}[l \to l']$ in (28) we obtain

$$\begin{aligned}\mathbb{E}[\underline{H}_{\tau,j}[l \to l']\underline{H}_{t,j}[l \to l']] &= \frac{\mathbb{E}[\underline{M}_\tau[l]\underline{M}_t[l]]}{L\bar{\underline{\xi}}_{\mathrm{out},\tau}[l]\bar{\underline{\xi}}_{\mathrm{out},t}[l]} \\ &+ \sum_{\tilde{l}'\in\mathcal{N}[l]\backslash\{l'\}} \mathbb{E}[\underline{H}_{\tau,j-1}[\tilde{l}' \to l]\underline{H}_{t,j-1}[\tilde{l}' \to l]].\end{aligned} \tag{145}$$

Thus, Theorem 2 holds.

Appendix C
Proof of Theorem 3

To represent a fixed point of the state evolution recursion with respect to the variance variables, we replace the iteration index t with an asterisk for all variables. We first focus on the inner module. Since a unique fixed point of (unnormalized) consensus propagation [44] is summation consensus under Assumption 5, from (29), (31), and Theorem 2 we have

$$\frac{1}{L}\bar{\eta}_*[l] = \frac{1}{L}\sum_{l'\in\mathcal{L}} \frac{\bar{\zeta}_*[l']}{\bar{\xi}_{\mathrm{out},*}[l']} \equiv \bar{\eta}_*, \tag{146}$$

$$\bar{\sigma}_*^2[l] = \frac{1}{L}\sum_{l'\in\mathcal{L}} \frac{1}{\bar{\xi}_{\mathrm{out},*}[l']} \equiv \bar{\sigma}_*^2, \tag{147}$$

$$\bar{\Sigma}_*[l] = \sum_{l'\in\mathcal{L}} \frac{\mathbb{E}[M_*^2[l']]}{L\bar{\xi}_{\mathrm{out},*}^2[l']} \equiv \bar{\Sigma}_* \tag{148}$$

for all $l \in \mathcal{L}$. Thus, all variables fed back from the inner module are identical for all l, i.e. for (33), (34), (35), and (36)

$$\bar{\xi}_{\mathrm{in},*}[l] = \mathbb{E}[\partial_1 f_{\mathrm{in},*}(\bar{\eta}_* X + \tilde{H}_*; L, \bar{\sigma}_*^2)] \equiv \bar{\xi}_{\mathrm{in},*}, \tag{149}$$

$$\bar{v}_*[l] = \frac{\bar{\sigma}_*^2\bar{\xi}_{\mathrm{in},*}}{L\delta} \equiv \bar{v}_*, \tag{150}$$

$$\mathbb{E}[Z[l]Z_*[l]] = \frac{1}{L\delta}\mathbb{E}\left[Xf_{\mathrm{in}}\left(\bar{\eta}_* X + \tilde{H}_*; L, \bar{\sigma}_*^2\right)\right] \equiv \frac{\mathbb{E}[ZZ_*]}{L\delta}, \tag{151}$$

$$\mathbb{E}[Z_*^2[l]] = \frac{1}{L\delta}\mathbb{E}\left[f_{\mathrm{in}}^2\left(\bar{\eta}_* X + \tilde{H}_*; L, \bar{\sigma}_*^2\right)\right] \equiv \frac{\mathbb{E}[Z_*^2]}{L\delta}, \tag{152}$$

with $Z \sim \mathcal{N}(0, (L\delta)^{-1}\mathbb{E}[X^2])$ independent of W and $\tilde{H}_* \sim \mathcal{N}(0, \bar{\Sigma}_*)$ independent of X. In the derivation of these expressions, we have used $\eta_*[l] = L$ obtained from the definition of $\eta_t[l]$ in (13).

We next focus on the outer module. Let Z and Z_* denote zero-mean Gaussian random variable with covariance $\mathbb{E}[Z^2]$, $\mathbb{E}[ZZ_*]$, and $\mathbb{E}[Z_*^2]$, independent of W. Using (24) and (25) yields

$$\bar{\xi}_{\mathrm{out},*}[l] = \mathbb{E}\left[\partial_1 f_{\mathrm{out}}(Z_*, Y; \bar{v}_*)\right] \equiv \bar{\xi}_{\mathrm{out},*}, \tag{153}$$

$$\bar{\zeta}_*[l] = -\mathbb{E}\left[\frac{\partial}{\partial z} f_{\mathrm{out}}(Z_*, g(z, W); \bar{v}_*)\bigg|_{z=Z}\right] \equiv \bar{\zeta}_*, \tag{154}$$

with $Y = g(Z, W)$. Substituting these expressions into (146), (147), and (148), we have $\bar{\eta}_* = \bar{\zeta}_*/\bar{\xi}_{\mathrm{out},*}$, $\bar{\sigma}_*^2 = \bar{\xi}_{\mathrm{out},*}^{-1}$, and

$$\bar{\Sigma}_* = \frac{1}{\bar{\xi}_{\mathrm{out},*}^2}\mathbb{E}\left[f_{\mathrm{out}}^2(Z_*, Y; \bar{v}_*)\right]. \tag{155}$$

From these observations, we find that the fixed point of the state evolution recursion for D-GAMP is equivalent to that for centralized GAMP [18].

APPENDIX D
PROOF OF THEOREM 4

We first derive the error model of D-GAMP with the distributed protocol (44). Replace $\bar{\eta}_t[l]$ in Lemma 4 with

$$\bar{\eta}_t[l] = \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\text{out},t}[l]} + \gamma \sum_{l' \in \mathcal{N}[l]} \left(\frac{\bar{\zeta}_t[l']}{\bar{\xi}_{\text{out},t}[l']} - \frac{\bar{\zeta}_t[l]}{\bar{\xi}_{\text{out},t}[l]} \right). \quad (156)$$

Applying $\tilde{\boldsymbol{x}}_t[l]$ in (44) to the definition of $\tilde{\boldsymbol{h}}_t[l]$ in Lemma 4 for $T = 1$ yields

$$\tilde{\boldsymbol{h}}_t[l] = \frac{\boldsymbol{h}_t[l]}{\xi_{\text{out},t}[l]} + \gamma \sum_{l' \in \mathcal{N}[l]} \left(\frac{\boldsymbol{h}_t[l']}{\xi_{\text{out},t}[l']} - \frac{\boldsymbol{h}_t[l]}{\xi_{\text{out},t}[l]} \right), \quad (157)$$

with $\boldsymbol{h}_t[l]$ defined in Lemma 4. Thus, the error model of D-GAMP with the distributed protocol (44) is equal to the error model (71)–(75) with $C_{t,0}[l][l] = 1$, in which $\tilde{\boldsymbol{h}}_t[l]$ in (74) is replaced with (157).

Define $\Sigma_{t,t}[l] = N^{-1}\|\tilde{\boldsymbol{h}}_t[l]\|^2$. It is sufficient to confirm that the effective SNR $L^{-2}\bar{\eta}_t^2[l]/\Sigma_{t,t}[l]$ for the inner denoiser is different from that of centralized GAMP. Suppose that $\bar{\xi}_{\text{out},t}[l]$ and $\bar{\zeta}_t[l]$ converge to $\bar{\xi}_{\text{out},*}[l]$ and $\bar{\zeta}_*[l]$ as $t \to \infty$, respectively. When $\gamma > 0$ is set appropriately [32], the definition of $\bar{\eta}_t[l]$ in (156) implies that the following consensus is achieved:

$$\lim_{t\to\infty} \frac{\bar{\eta}_t[l]}{L} = \frac{1}{L} \sum_{l' \in \mathcal{L}} \frac{\bar{\zeta}_*[l']}{\bar{\xi}_{\text{out},*}[l']} = \bar{\eta}_* \quad (158)$$

for all $l \in \mathcal{L}$, with $\bar{\eta}_*$ given in (146).

We next evaluate $\Sigma_{t,t}[l] = N^{-1}\|\tilde{\boldsymbol{h}}_t[l]\|^2$. Applying Properties (Oc) and (Ib) in Theorem 6 to $\tilde{\boldsymbol{h}}_t[l]$ in (157) yields

$$\Sigma_{t,t}[l] \overset{\text{a.s.}}{\to} (1 - \gamma|\mathcal{N}[l]|)^2 \frac{\mathbb{E}[M_t^2[l]]}{L\bar{\xi}_{\text{out},t}^2[l]} + \gamma^2 \sum_{l' \in \mathcal{N}[l]} \frac{\mathbb{E}[M_t^2[l']]}{L\bar{\xi}_{\text{out},t}^2[l']}, \quad (159)$$

which cannot converge to $\bar{\Sigma}_*[l]$ in (148), because (159) is different from the distributed protocol for average consensus [32]. These observations imply that the fixed point of the effective SNR $L^{-2}\bar{\eta}_t^2[l]/\Sigma_{t,t}[l]$ is different from that for centralized GAMP. Thus, Theorem 4 holds.

APPENDIX E
PROPERTIES FOR THE BAYES-OPTIMAL DENOISERS

A. Proof of Lemma 2

We first confirm the following proposition:

Proposition 1:

- $\bar{v}_{0,t}[l] = \bar{v}_{t,t}[l]$ implies $\mathbb{E}[Z[l]Z_t[l]] = \mathbb{E}[Z_t^2[l]]$.
- $\bar{v}_{0,\tau}[l] = \bar{v}_{\tau,\tau}[l]$, $\bar{v}_{0,t}[l] = \bar{v}_{t,t}[l]$, and $\bar{v}_{\tau,t}[l] = \bar{v}_{t,t}[l]$ imply $\mathbb{E}[Z_\tau[l]Z_t[l]] = \mathbb{E}[Z_\tau^2[l]]$.

Proof: We first prove the former property. Applying the assumption $\bar{v}_{0,t}[l] = \bar{v}_{t,t}[l]$ to the definitions of (52) and (53) yields $\mathbb{E}[(Z[l] + B_t[l])B_t[l]] = 0$, which implies $\mathbb{E}[Z_t[l]B_t[l]] = 0$. Thus, we have $\mathbb{E}[Z_t^2[l]] = \mathbb{E}[Z_t[l]Z[l]]$.

We next prove the latter property. Using the definitions of (52) and (53) yields

$$\begin{aligned}&\mathbb{E}[Z_\tau^2[l]] - \mathbb{E}[Z_\tau[l]Z_t[l]] = \mathbb{E}[(Z[l] + B_\tau[l])(B_\tau[l] - B_t[l])] \\ &= -\bar{v}_{0,\tau}[l] + \bar{v}_{0,t}[l] + \bar{v}_{\tau,\tau}[l] - \bar{v}_{\tau,t}[l] = 0, \end{aligned} \quad (160)$$

where the last equality follows from the assumptions. ■

We prove Lemma 2. From Proposition 1 under the assumption $\bar{v}_{0,t}[l] = \bar{v}_{t,t}[l]$, it is straightforward to find the representation $Z[l] \sim Z_t[l] + N_t[l]$ for all $t \geq 0$, with $N_t[l] \sim \mathcal{N}(0, \bar{v}_{t,t}[l])$ independent of $Z_t[l]$. This representation justifies the expression of the posterior mean estimator $\hat{Z}_t[l]$ in (54).

Let $f_{\text{out},t}[l] = f_{\text{out}}[l](Z_t[l], g[l](Z[l], W[l]); \bar{v}_{t,t}[l])$ with $\bar{v}_t[l] = \bar{v}_{t,t}[l]$. We use the representation $Z[l] \sim Z_t[l] + N_t[l]$, Lemma 6, and the definition of $\bar{\zeta}_t[l]$ in (25) to obtain

$$\mathbb{E}[N_t[l]f_{\text{out},t}[l]] = -\bar{v}_{t,t}[l]\bar{\zeta}_t[l]. \quad (161)$$

Thus, we have the following identity:

$$\begin{aligned}\bar{\zeta}_t[l] &= -\mathbb{E}\left[\frac{Z[l] - Z_t[l]}{\bar{v}_{t,t}[l]} f_{\text{out},t}[l]\right] \\ &= \mathbb{E}\left[\frac{Z_t[l] - \hat{Z}_t[l]}{\bar{v}_{t,t}[l]} f_{\text{out},t}[l]\right], \end{aligned} \quad (162)$$

with the posterior mean estimator $\hat{Z}_t[l] = \mathbb{E}[Z[l]|Z_t[l], Y[l]]$. Using the Cauchy–Schwarz inequality for $\zeta_t^2[l]$, we arrive at

$$\frac{\bar{\zeta}_t^2[l]}{\mathbb{E}[M_t^2[l]]} \leq \mathbb{E}\left[\left(\frac{Z_t[l] - \hat{Z}_t[l]}{\bar{v}_{t,t}[l]}\right)^2\right], \quad (163)$$

where the equality holds if and only if there is some constant $C \in \mathbb{R}$ such that $f_{\text{out},t}[l] = C(Z_t[l] - \hat{Z}_t[l])/\bar{v}_{t,t}[l]$ is satisfied. Thus, Lemma 2 holds.

B. Proof of Lemma 3

We first prove basic properties of the Bayes-optimal denoisers.

Lemma 9 ([18]): Consider estimation of X based on the Gaussian measurement $Y = aX + Z$ with $Z \sim \mathcal{N}(0, \sigma^2)$ independent of X. Let the posterior mean estimator $f(y) = \mathbb{E}[X|Y = y]$. Then, we have

$$\mathbb{E}[Xf(Y)] = \mathbb{E}[f^2(Y)], \quad (164)$$

$$f'(y) = \frac{a\mathbb{E}[\{X - f(Y)\}^2|Y = y]}{\sigma^2}. \quad (165)$$

Proof: The former identity is trivial: $\mathbb{E}[Xf(Y)] = \mathbb{E}[\mathbb{E}[Xf(Y)|Y]] = \mathbb{E}[f^2(Y)]$. The latter identity is obtained from direct computation,

$$\begin{aligned}f'(y) &= \frac{d}{dy} \frac{\int x e^{-\frac{(y-ax)^2}{2\sigma^2}} dP(x)}{\int e^{-\frac{(y-ax)^2}{2\sigma^2}} dP(x)} \\ &= -\frac{\langle x \rangle y - a\langle x^2 \rangle}{\sigma^2} + \langle x \rangle \frac{y - a\langle x \rangle}{\sigma^2} = a\frac{\langle x^2 \rangle - \langle x \rangle^2}{\sigma^2}, \end{aligned} \quad (166)$$

with

$$\langle x^i \rangle = \frac{\int x^i e^{-\frac{(y-ax)^2}{2\sigma^2}} dP(x)}{\int e^{-\frac{(y-ax)^2}{2\sigma^2}} dP(x)}. \quad (167)$$

Thus, the latter identity holds. ■

We follow [18] to represent the Bayes-optimal outer denoiser (56) with $C = 1$ as the (negative) score function

$$f_{\mathrm{out}}[l](\theta, y; \bar{v}_{t,t}[l]) = -\frac{\partial}{\partial\theta}\log P[l](y|\theta; \bar{v}_{t,t}[l]), \tag{168}$$

with

$$P[l](y|\theta; v) = \int P_{Y[l]|Z[l]}(y|z)\frac{1}{\sqrt{2\pi v}}e^{-\frac{(z-\theta)^2}{2v}}dz, \tag{169}$$

where $P_{Y[l]|Z[l]}(y|z)$ denotes the conditional distribution of $Y[l]$ given $Z[l]$. Note that the distribution $P[l](y|\theta; v)$ is normalized with respect to y. It is straightforward to confirm the equivalence between the two representations.

We reproduce an existing lemma [18].

Lemma 10 ([18]): Suppose that Assumption 6 holds and consider the Bayes-optimal outer denoiser (168) with $C = 1$. Furthermore, suppose that $\bar{v}_t[l]$ in (34) is equal to $\bar{v}_{t,t}[l]$ in (53). If $\bar{v}_{0,t}[l]$ is equal to $\bar{v}_{t,t}[l]$, then we have

$$\bar{\zeta}_t[l] = \bar{\xi}_{\mathrm{out},t}[l]. \tag{170}$$

Proof: We first consider the case of $t > 0$. Utilizing the well-known identity for the score function

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\log P[l](Y[l]|\theta; \bar{v}_{t,t}[l])\middle| Z_t[l] = \theta\right] = 0 \tag{171}$$

and the representation of the Bayes-optimal outer denoiser in (168) with $\bar{v}_t[l] = \bar{v}_{t,t}[l]$ yields

$$\mathbb{E}[Z_t[l]f_{\mathrm{out}}[l](Z_t[l], Y[l]; \bar{v}_{t,t}[l])] = 0. \tag{172}$$

We use Lemma 6 and the definitions of $\bar{\xi}_{\mathrm{out},t}[l]$ and $\bar{\zeta}_t[l]$ in (24) and (25) to obtain

$$\begin{aligned} 0 &= \mathbb{E}[Z_t[l]f_{\mathrm{out}}[l](Z_t[l], g[l](Z[l], W[l]); \bar{v}_{t,t}[l])] \\ &= \mathbb{E}[Z_t^2[l]]\bar{\xi}_{\mathrm{out},t}[l] - \mathbb{E}[Z_t[l]Z[l]]\bar{\zeta}_t[l] \\ &= \mathbb{E}[Z_t^2[l]](\bar{\xi}_{\mathrm{out},t}[l] - \bar{\zeta}_t[l]), \end{aligned} \tag{173}$$

where the last equality follows from Proposition 1 under the assumption $\bar{v}_{0,t}[l] = \bar{v}_{t,t}[l]$. Since $\mathbb{E}[Z_t^2] > 0$ holds for $t > 0$, we arrive at Lemma 10 for $t > 0$.

We next consider the case of $t = 0$, in which we have $Z_0[l] = 0$. To extract information about the partial derivative of $f_{\mathrm{out}}[l](\theta, Y[l]; \bar{v}_{0,0}[l])$, we inject an independent weak Gaussian noise $Z_0^{(n)}[l] \sim \mathcal{N}(0, n^{-1})$ to θ for sufficiently large $n \in \mathbb{N}$,

$$\bar{\xi}_{\mathrm{out},0}^{(n)}[l] = \mathbb{E}\left[\partial_1 f_{\mathrm{out}}[l](Z_0^{(n)}[l], Y[l]; \bar{v}_0[l])\right], \tag{174}$$

$$\bar{\zeta}_0^{(n)}[l] = -\mathbb{E}\left[\frac{\partial}{\partial z}f_{\mathrm{out}}[l](Z_0^{(n)}[l], g[l](z, W[l]); \bar{v}_0[l])\bigg|_{z=Z[l]}\right], \tag{175}$$

associated with $\bar{\xi}_{\mathrm{out},0}[l]$ and $\bar{\zeta}_0[l]$ in (24) and (25), respectively. Repeating the proof of Lemma 10 for $t > 0$, we obtain $\bar{\xi}_{\mathrm{out},0}^{(n)}[l] = \bar{\zeta}_0^{(n)}[l]$ for any $n \in \mathbb{N}$.

To prove $\bar{\xi}_{\mathrm{out},0}[l] = \bar{\zeta}_0[l]$, we show $\lim_{n\to\infty}\bar{\xi}_{\mathrm{out},0}^{(n)}[l] = \bar{\xi}_{\mathrm{out},0}[l]$ and $\lim_{n\to\infty}\bar{\zeta}_0^{(n)}[l] = \bar{\zeta}_0[l]$. Assumption 6 implies that $f_{\mathrm{out},0}[l]$ is differentiable almost everywhere and has bounded partial derivatives. Thus, we use the dominated convergence theorem to arrive at these limits. ■

We prove Lemma 3. From the assumption $\bar{v}_{\tau',\tau'}[l] > \bar{v}_{\tau,\tau}[l]$ and Proposition 1 under the assumption $\bar{v}_{\tau,t}[l] = \bar{v}_{t,t}[l]$ for all $\tau \in \{0, \ldots, t\}$, we have the following cascaded representation of $Z_\tau[l]$ and $Z_t[l]$:

$$Z[l] = Z_t[l] + N_t[l], \quad Z_t[l] = Z_\tau[l] + \Delta N_{\tau,t}[l] \tag{176}$$

for $\tau \geq 0$, where $N_t[l] \sim \mathcal{N}(0, \bar{v}_{t,t}[l])$ and $\Delta N_{\tau,t}[l] \sim \mathcal{N}(0, \bar{v}_{\tau,\tau}[l] - \bar{v}_{t,t}[l])$ are independent of all random variables. It is straightforward to confirm $\mathbb{E}[Z_\tau[l]Z_t[l]] = \mathbb{E}[Z_\tau^2[l]]$ and $\mathbb{E}[(Z_\tau[l] - Z[l])^2] = \bar{v}_{\tau,\tau}[l]$. Since $Z_\tau[l]$ depends on $Z[l]$ only through $Z_t[l]$ for all $\tau > 0$, as well as $Z_0[l] = 0$, we have

$$\mathbb{E}\left[Z|Z_t[l], Z_\tau[l], Y[l]\right] = \mathbb{E}\left[Z|Z_t[l], Y[l]\right]. \tag{177}$$

For $M_t[l] = f_{\mathrm{out}}[l](Z_t[l], Y[l]; \bar{v}_{t,t}[l])$, we use the definition of the Bayes-optimal outer denoiser in (56) with $C = 1$ to evaluate $\mathbb{E}[M_\tau[l]M_t[l]]$ as

$$\begin{aligned} &\bar{v}_{t,t}[l]\mathbb{E}[M_\tau[l]M_t[l]] = \mathbb{E}\left[M_\tau[l](Z_t[l] - \hat{Z}_t[l])\right] \\ &= \mathbb{E}\left[M_\tau[l]\left(Z_t[l] - \mathbb{E}[Z[l]|Z_\tau[l], Z_t[l], Y[l]]\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[M_\tau[l](Z_t[l] - Z[l])\,|Z_\tau[l],\ Z_t[l], Y[l]\right]\right] \\ &= \mathbb{E}\left[M_\tau[l](Z_t[l] - Z[l])\right], \end{aligned} \tag{178}$$

where the second and third equalities follow from the identity (177) and from the fact that $M_\tau[l]$ is a deterministic function of $Z_\tau[l]$ and $Y[l]$. Using Lemma 6 yields

$$\begin{aligned} &\bar{v}_{t,t}[l]\mathbb{E}[M_\tau[l]M_t[l]] \\ &= \mathbb{E}[(Z_t[l] - Z[l])Z_\tau[l]]\bar{\xi}_{\mathrm{out},\tau}[l] - \mathbb{E}[(Z_t[l] - Z[l])Z[l]]\bar{\zeta}_\tau[l] \\ &= \bar{\xi}_{\mathrm{out},\tau}[l]\mathbb{E}[(Z_t[l] - Z[l])(Z_\tau[l] - Z[l])] \\ &= \bar{\xi}_{\mathrm{out},\tau}[l]\bar{v}_{\tau,t}[l] \end{aligned} \tag{179}$$

with $\bar{\xi}_{\mathrm{out},\tau}[l]$ and $\bar{\zeta}_\tau[l]$ given in (24) and (25), where the second and last equalities follow from Lemma 10 and the definition of $\bar{v}_{\tau,t}[l]$ in (53), respectively. Using the assumption $\bar{v}_{\tau,t}[l] = \bar{v}_{t,t}[l]$ for all $\tau \in \{0, \ldots, t\}$, we arrive at Lemma 3.

Appendix F
Proof of Theorem 5

A. Long-Memory Proof Strategy

In the long-memory proof strategy [46], the covariance matrix of estimation errors for D-GAMP is utilized to prove the convergence of the state evolution recursion with respect to the variance of the estimation errors. When the covariance matrix has a special structure, the positive definiteness of the covariance matrix implies the convergence of its diagonal elements from basic properties in linear algebra. Furthermore, the Bayes-optimality of the denoisers produces the special structure in the covariance matrix naturally. As a result, the convergence of the state evolution recursion can be proved, without utilizing concrete properties in the measurement model.

As shown in Theorem 6, rigorous state evolution with respect to the MSE has already required evaluation of the error covariance matrix and a guarantee for its positive definiteness. To prove the convergence in the long-memory strategy, thus, the only additional tasks are evaluation of the error covariance matrices for the Bayes-optimal denoisers, as presented in

Lemmas 1 and 3. In this sense, we are ready for proving Theorem 5.

The following lemma is a technical result to guarantee the monotonicity for the diagonal elements of a covariance matrix:

Lemma 11 ([46]): Suppose that a symmetric matrix $\boldsymbol{M}_t \in \mathbb{R}^{(t+1)\times(t+1)}$ is strictly positive definite. Let $m_{\tau,t}$ denote the (τ, t) element of $\boldsymbol{M}_t$. Then,

- $m_{\tau',\tau} = m_{\tau,\tau}$ for all $\tau \in \{0,\ldots,t\}$ and $\tau' \in \{0,\ldots,\tau\}$ implies $m_{\tau',\tau'} > m_{\tau,\tau}$ for all $\tau' < \tau \leq t$.
- $m_{\tau',\tau} = m_{\tau',\tau'}$ for all $\tau \in \{0,\ldots,t\}$ and $\tau' \in \{0,\ldots,\tau\}$ implies $m_{\tau',\tau'} < m_{\tau,\tau}$ for all $\tau' < \tau \leq t$.

Proof: We only prove the latter property since the former property was proved in [46, Lemma 3]. The proof is by induction. For $t = 1$, we use $\det \boldsymbol{M}_1 = m_{0,0}m_{1,1} - m_{0,0}^2$ and the positive definiteness of $\boldsymbol{M}_1$ to obtain $m_{1,1} > m_{0,0} > 0$. Suppose that the latter property is correct for some t. We need to prove the latter property for $\boldsymbol{M}_{t+1}$.

The positive definiteness of $\boldsymbol{M}_{t+1}$ implies that of $\boldsymbol{M}_1$. Thus, we have $0 < m_{0,0} < m_{1,1}$. Subtracting the first row in $\boldsymbol{M}_{t+1}$ from the other rows, we find $\det \boldsymbol{M}_{t+1} = m_{0,0} \det \tilde{\boldsymbol{M}}_t$, with $\tilde{m}_{\tau',\tau} \equiv [\tilde{\boldsymbol{M}}_t]_{\tau',\tau} = m_{\tau'+1,\tau+1} - m_{0,0}$ for $\tau', \tau \in \{0,\ldots,t\}$. The positive definiteness of $\boldsymbol{M}_{t+1}$ implies the positive definiteness of $\tilde{\boldsymbol{M}}_t$. Since $\tilde{m}_{\tau',\tau} = \tilde{m}_{\tau',\tau'}$ holds for $\tau' \leq \tau$, we use the induction hypothesis for $\tilde{\boldsymbol{M}}_t$ to obtain $m_{\tau'+1,\tau'+1} < m_{\tau+1,\tau+1}$ for all $0 < \tau' < \tau < t$. Combining these results, we arrive at the latter property for $\boldsymbol{M}_{t+1}$. ■

As presented in Lemmas 1 and 3, the structure $m_{\tau',\tau} = m_{\tau,\tau}$ or $m_{\tau',\tau} = m_{\tau',\tau'}$ in Lemma 11 appears when the Bayes-optimal denoisers are used. To prove Theorem 5, we need the following lemma:

Lemma 12: Suppose that Assumption 6 holds and consider the Bayes-optimal inner denoiser (46) and outer denoiser (56) with $C = 1$. Then, for all $t = 0, 1, \ldots$ we have the following properties for the outer module:

- $\bar{\Sigma}_{\tau',t}[l]$ satisfies (59). In particular, $\bar{\Sigma}_{\tau',t}[l] = \bar{\Sigma}_{t,t}[l]$ holds for all $\tau' \in \{0,\ldots,t\}$.
- $\bar{\eta}_t[l] = \eta_t[l]$ and $\bar{\sigma}_t^2[l] = \bar{\Sigma}_{t,t}[l]$ hold.
- $\bar{\Sigma}_{\tau',\tau'} > \bar{\Sigma}_{\tau,\tau}[l]$ holds for all $\tau' < \tau \leq t$

On the other hand, for the inner module we have

- $\mathrm{cov}_{\tau',t+1}[l] = \mathrm{cov}_{t+1,t+1}[l]$ holds for all $\tau' \in \{0,\ldots,t+1\}$.
- $\bar{v}_{t+1}[l] = \bar{v}_{t+1,t+1}[l]$ and $\bar{\zeta}_{t+1}[l] = \bar{\xi}_{\mathrm{out},t+1}[l]$ hold.
- $\mathrm{cov}_{\tau',\tau'}[l] > \mathrm{cov}_{\tau,\tau}[l]$ holds for all $\tau' < \tau \leq t+1$.

Proof: The proof of Lemma 12 is by induction with respect to t. The proof for $t = 0$ is omitted because it is the same as for the general case. For some τ, suppose that Lemma 12 is correct for all $t < \tau$. We need to prove Lemma 12 for $t = \tau$. See Appendices F-B and F-C for the proofs of the properties in the outer and inner modules, respectively. ■

We prove Theorem 5. The first properties in Theorem 5 are part of Lemma 12. The second property follows from Theorems 1, 2, and Lemma 12. The last property is obtained from the monotonicity in Lemma 12: $\{\mathrm{cov}_{t,t}[l]\}$ are a monotonically decreasing sequence with respect to t. Since the MSE $\mathrm{cov}_{t,t}[l]$ is non-negative, we conclude that $\mathrm{cov}_{t,t}[l]$ converges to a non-negative constant as $t \to \infty$.

B. Proof for the Outer Module

For some τ, suppose that Lemma 12 is correct for all $t < \tau$. We confirm the conditions in Lemma 3 for all $t \leq \tau$. Applying the induction hypotheses $\mathrm{cov}_{\tau',t}[l] = \mathrm{cov}_{t,t}[l]$ for all $t \leq \tau$ and $\tau' \in \{0,\ldots,t\}$ and $\mathrm{cov}_{\tau',\tau'}[l] > \mathrm{cov}_{t,t}[l]$ for all $\tau' < t \leq \tau$ to the relationship between $\bar{v}_{\tau',t}[l]$ and $\mathrm{cov}_{\tau',t}[l]$ in (53), respectively, we obtain $\bar{v}_{\tau',t}[l] = \bar{v}_{t,t}[l]$ for all $t \leq \tau$ and $\tau' \in \{0,\ldots,t\}$ and $\bar{v}_{\tau',\tau'}[l] > \bar{v}_{t,t}[l]$ for all $\tau' < t \leq \tau$. From these properties and the induction hypothesis $\bar{v}_t[l] = \bar{v}_{t,t}[l]$ for all $t \leq \tau$, as well as Assumption 6, we can use Lemma 3 for all $t \leq \tau$.

We first derive the representation of $\bar{\Sigma}_{\tau',\tau}[l]$ in (59) for $\tau' \in \{0,\ldots,\tau\}$. Lemma 3 for $t = \tau$ implies that (42) and (43) reduce to

$$\bar{\Sigma}_{\tau',\tau}[l] = \frac{1}{L\bar{\xi}_{\mathrm{out},\tau}[l]} + \sum_{l' \in \mathcal{N}[l]} \underline{\bar{\Sigma}}_{\tau',\tau,J}[l' \to l], \tag{180}$$

with

$$\underline{\bar{\Sigma}}_{\tau',\tau,j}[l \to l'] = \frac{1}{L\underline{\bar{\xi}}_{\mathrm{out},\tau}[l]} + \sum_{\tilde{l}' \in \mathcal{N}[l]\setminus\{l'\}} \underline{\bar{\Sigma}}_{\tau',\tau,j-1}[\tilde{l}' \to l], \tag{181}$$

which are equivalent to (59) and (60), because of the identity $\bar{\xi}_{\mathrm{out},\tau}[l] = \mathbb{E}[M_\tau^2[l]]$ obtained from Lemma 3. The identity $\bar{\Sigma}_{\tau',\tau}[l] = \bar{\Sigma}_{\tau,\tau}[l]$ is trivial for $\tau' \in \{0,\ldots,\tau\}$ from the definitions of $\bar{\Sigma}_{\tau',\tau}[l]$ and $\underline{\bar{\Sigma}}_{\tau',\tau}[l \to l']$.

We next prove $\bar{\eta}_\tau[l] = \eta_\tau[l]$ and $\bar{\sigma}_\tau^2[l] = \bar{\Sigma}_{\tau,\tau}[l]$. We use the induction hypothesis $\bar{\zeta}_\tau[l] = \bar{\xi}_{\mathrm{out},\tau}[l]$ for the definitions of $\bar{\eta}_\tau[l]$ and $\underline{\bar{\eta}}_{\tau,j}[l \to l']$ in (29) and (30) to obtain $\bar{\eta}_\tau[l] = \eta_\tau[l]$ given in (13). The identity $\bar{\sigma}_\tau^2[l] = \bar{\Sigma}_{\tau,\tau}[l]$ follows from the definitions of $\bar{\sigma}_\tau^2[l]$, $\underline{\bar{\sigma}}_{\tau,j}^2[l \to l']$, $\bar{\Sigma}_{\tau,\tau}[l]$, and $\underline{\bar{\Sigma}}_{\tau,\tau,j}[l \to l']$ in (31), (32), (59), and (60), as well as Lemma 3.

Finally, we prove $\bar{\Sigma}_{\tau',\tau'} > \bar{\Sigma}_{t,t}[l]$ for all $\tau' < t \leq \tau$. From the definition of $\bar{\Sigma}_{\tau',\tau}[l]$ in (59), it is sufficient to prove $\mathbb{E}[M_{\tau'}^2[l]] < \mathbb{E}[M_t^2[l]]$ for all $\tau' < t \leq \tau$. From Properties (Oc) and (Oe) in Theorem 6, we find the positive definiteness of the covariance matrix that has $\mathbb{E}[M_{\tau'}[l]M_t[l]]$ as the (τ', t) element for all $\tau', t \in \{0,\ldots,\tau\}$. Using Lemma 11 for this covariance matrix with $\mathbb{E}[M_{\tau'}[l]M_t[l]] = \mathbb{E}[M_{\tau'}^2[l]]$ for all $\tau' \in \{0,\ldots,t\}$, obtained from Lemma 3 for all $t \leq \tau$, we arrive at $\mathbb{E}[M_{\tau'}^2[l]] < \mathbb{E}[M_t^2[l]]$ for all $\tau' < t \leq \tau$, which implies $\bar{\Sigma}_{\tau',\tau'} > \bar{\Sigma}_{t,t}[l]$ for all $\tau' < t \leq \tau$.

C. Proof for the Inner Module

We prove the identity $\mathrm{cov}_{\tau',\tau+1}[l] = \mathrm{cov}_{\tau+1,\tau+1}[l]$ for all $\tau' \in \{0,\ldots,\tau+1\}$. We first consider the case of $\tau' = 0$. Let $f_{\mathrm{in},\tau}[l] = f_{\mathrm{in}}[l](L^{-1}\bar{\eta}_\tau[l]X + \tilde{H}_\tau[l]; \eta_\tau[l], \bar{\Sigma}_{\tau,\tau}[l])$. Using the definition of $\mathrm{cov}_{0,\tau+1}[l]$ in (40) and $\bar{\sigma}_\tau^2[l] = \bar{\Sigma}_{\tau,\tau}[l]$ yields

$$\begin{aligned}\mathrm{cov}_{0,\tau+1}[l] &= \mathbb{E}\left[(X - f_{\mathrm{in},\tau}[l] + f_{\mathrm{in},\tau}[l])(X - f_{\mathrm{in},\tau}[l])\right] \\ &= \mathbb{E}[(X - f_{\mathrm{in},\tau}[l])^2],\end{aligned} \tag{182}$$

where the last equality follows from the well-known property $\mathbb{E}[f_{\mathrm{in},\tau}[l](X - f_{\mathrm{in},\tau}[l])] = 0$ for the Bayes-optimal inner denoiser (46). Thus, we use the definition of $\mathrm{cov}_{\tau+1,\tau+1}[l]$ in (39) to have $\mathrm{cov}_{0,\tau+1}[l] = \mathrm{cov}_{\tau+1,\tau+1}[l]$.

We next consider the case of $\tau' > 0$. From $\bar{\Sigma}_{\tau',\tau}[l] = \bar{\Sigma}_{\tau,\tau}[l]$ for all $\tau' \in \{0,\ldots,\tau\}$, $\bar{\Sigma}_{\tau',\tau'} > \bar{\Sigma}_{t,t}$ for all $\tau' <$

$t \leq \tau$, and $\bar{\sigma}_{\tau}^{2}[l] = \bar{\Sigma}_{\tau,\tau}[l]$, we use Lemma 1 to obtain $\mathrm{cov}_{\tau'+1,\tau+1}[l] = \mathrm{cov}_{\tau+1,\tau+1}[l]$ for all $\tau' \in \{0,\ldots,\tau\}$. Combining these results, we arrive at $\mathrm{cov}_{\tau',\tau+1}[l] = \mathrm{cov}_{\tau+1,\tau+1}[l]$ for all $\tau' \in \{0,\ldots,\tau+1\}$.

Let us prove $\bar{v}_{\tau+1}[l] = \bar{v}_{\tau+1,\tau+1}[l]$. From $\bar{\eta}_{\tau}[l] = \eta_{\tau}[l]$ and $\bar{\sigma}_{\tau}^{2}[l] = \bar{\Sigma}_{\tau,\tau}[l]$, we use Lemma 9 for the Bayes-optimal inner denoiser (46) and the definition of $\bar{\xi}_{\mathrm{in},\tau}[l]$ in (33) to obtain

$$\bar{\xi}_{\mathrm{in},\tau}[l] = \frac{\eta_{\tau}[l]\mathrm{cov}_{\tau+1,\tau+1}[l]}{L\bar{\Sigma}_{\tau,\tau}[l]}, \tag{183}$$

with $\mathrm{cov}_{\tau+1,\tau+1}[l]$ given in (39). Applying this identity and $\bar{\sigma}_{\tau}^{2}[l] = \bar{\Sigma}_{\tau,\tau}[l]$ to $\bar{v}_{\tau+1}[l]$ in (34), we have

$$\bar{v}_{\tau+1}[l] = \frac{1}{L\delta[l]}\mathrm{cov}_{\tau+1,\tau+1}[l], \tag{184}$$

which is equal to $\bar{v}_{\tau+1,\tau+1}[l]$ in (53).

We prove the identity $\bar{\zeta}_{\tau+1}[l] = \bar{\xi}_{\mathrm{out},\tau+1}[l]$. Using $\mathrm{cov}_{0,\tau+1}[l] = \mathrm{cov}_{\tau+1,\tau+1}[l]$ and the definition (52) yields $\bar{v}_{0,\tau+1}[l] = \bar{v}_{\tau+1,\tau+1}[l]$. From this identity, $\bar{v}_{\tau+1}[l] = \bar{v}_{\tau+1,\tau+1}[l]$, and Assumption 6, we use Lemma 10 to obtain the identity $\bar{\zeta}_{\tau+1}[l] = \bar{\xi}_{\mathrm{out},\tau+1}[l]$.

Finally, we prove $\mathrm{cov}_{\tau',\tau'}[l] > \mathrm{cov}_{t,t}[l]$ for all $\tau' < t \leq \tau+1$. Using $\bar{\sigma}_{\tau}[l] = \bar{\Sigma}_{\tau,\tau}[l]$ and $\bar{\zeta}_{\tau+1}[l] = \bar{\xi}_{\mathrm{out},\tau+1}[l]$, we represent the random variable $Q_{\tau'}$ in (93) as $Q_{0}[l] = -X$ and

$$Q_{\tau'+1}[l] = f_{\mathrm{in}}[l]\left(\frac{\bar{\eta}_{\tau'}[l]}{L}X + \tilde{H}_{\tau'}[l]; \eta_{\tau'}[l], \bar{\Sigma}_{\tau',\tau'}[l]\right) - X \tag{185}$$

for $\tau' \geq 0$, which imply $\mathbb{E}[Q_{\tau'}[l]Q_{t}[l]] = \mathrm{cov}_{\tau',t}[l]$ from the definitions of $\mathrm{cov}_{\tau',t}[l]$ in (39), (40), and (41). Properties (Ic) and (Ie) in Theorem 6 imply the positive definiteness of the covariance matrix that has $\mathrm{cov}_{\tau',t}[l]$ as the (τ',t) element for all $\tau', t \in \{0,\ldots,\tau+1\}$. From this positive definiteness and $\mathrm{cov}_{\tau',\tau+1}[l] = \mathrm{cov}_{\tau+1,\tau+1}[l]$ for all $\tau' \in \{0,\ldots,\tau+1\}$, as well as the induction hypothesis $\mathrm{cov}_{\tau',t}[l] = \mathrm{cov}_{t,t}[l]$ for all $t \in \{0,\ldots,\tau\}$ and $\tau' \in \{0,\ldots,t\}$, we use Lemma 11 to obtain $\mathrm{cov}_{\tau',\tau'}[l] > \mathrm{cov}_{t,t}[l]$ for all $\tau' < t \leq \tau+1$. Thus, Lemma 12 holds for $t = \tau$.

REFERENCES

[1] 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.

[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. 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.

[5] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a Markov-tree prior," *IEEE J. Sel. Topics Signal Process.*, vol. 60, no. 7, pp. 3439–3448, Jul. 2012.

[6] J. Tan, Y. Ma, and D. Baron, "Compressive imaging via approximate message passing with image denoising," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 2085–2092, Apr. 2015.

[7] L. Anitori, A. Maleki, M. Otten, R. G. Baraniuk, and P. Hoogeboom, "Design and analysis of compressed sensing radar detectors," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 813–827, Feb. 2013.

[8] C. Rush, A. Greig, and R. Venkataramanan, "Capacity-achieving sparse superposition codes via approximate message passing decoding," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1476–1500, Mar. 2017.

[9] J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity achieving sparse superposition codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 4894–4927, Aug. 2017.

[10] T. Lesieur, F. Krzakala, and L. Zdeborová, "Constrained low-rank matrix estimation: phase transitions, approximate message passing and applications," *J. Stat. Mech.: Theory Exp.*, vol. 2017, no. 7, p. 073403, Jul. 2017.

[11] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," *Ann. Stat.*, vol. 49, no. 1, pp. 321–345, Feb. 2021.

[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] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and message passing algorithms," *Ann. Appl. Probab.*, vol. 25, no. 2, pp. 753–822, Apr. 2015.

[14] K. Takeuchi, "A unified framework of state evolution for message-passing algorithms," in *Proc. 2019 IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019, pp. 151–155.

[15] 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.

[16] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for random linear estimation with Gaussian matrices is exact," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2252–2283, Apr. 2019.

[17] J. Barbier, N. Macris, M. Dia, and F. Krzakala, "Mutual information and optimality of approximate message-passing in random linear estimation," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4270–4303, Jul. 2020.

[18] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Aug. 2011, pp. 2168–2172.

[19] U. S. Kamilov, A. Bourquard, A. Amini, and M. Unser, "One-bit measurements with adaptive thresholds," *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 607–610, Oct. 2012.

[20] U. S. Kamilov, V. K. Goyal, and S. Rangan, "Message-passing dequantization with applications to compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6270–6281, Dec. 2012.

[21] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1043–1055, Feb. 2015.

[22] J. Ma, J. Xu, and A. Maleki, "Optimization-based AMP for phase retrieval: The impact of initialization and ℓ_2 regularization," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3600–3629, Jun. 2019.

[23] H. Bao, J. Fang, Z. Chen, H. Li, and S. Li, "An efficient Bayesian PAPR reduction method for OFDM-based massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4183–4195, Jun. 2016.

[24] J.-C. Chen, C.-J. Wang, K.-K. Wong, and C.-K. Wen, "Low-complexity precoding design for massive multiuser MIMO systems using approximate message passing," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5707–5714, Jul. 2016.

[25] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *Inf. Inference: A Journal of the IMA*, vol. 2, no. 2, pp. 115–144, Dec. 2013.

[26] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," *Proc. Nat. Acad. Sci.*, vol. 116, no. 12, pp. 5451–5460, Mar. 2019.

[27] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.

[28] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.

[29] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[30] S. Patterson, Y. C. Eldar, and I. Keidar, "Distributed sparse signal recovery for sensor networks," in *Proc. 2013 IEEE Int. Conf. Acoust. Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 4494–4498.

[31] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.

[32] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Contr. Lett.*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[33] G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed sparse linear regression," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5262–5276, Oct. 2010.
[34] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.
[35] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Püschel, "Distributed basis pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1942–1956, Apr. 2012.
[36] ——, "D-ADMM: A communication-efficient distributed algorithm for separable optimization," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2718–2723, May 2013.
[37] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the ADMM in decentralized consensus optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1750–1761, Apr. 2014.
[38] P. Han, R. Niu, M. Ren, and Y. C. Eldar, "Distributed approximate message passing for sparse signal recovery," in *Proc. 2014 IEEE Global Conf. Signal Inf. Process.*, Atlanta, GA, USA, Dec. 2014, pp. 497–501.
[39] P. Han, J. Zhu, R. Niu, and D. Baron, "Multi-processor approximate message passing using lossy compression," in *Proc. 2016 IEEE Int. Conf. Acoust. Speech Signal Process.*, Shanghai, China, Mar. 2016, pp. 6240–6244.
[40] Y. Ma, Y. M. Lu, and D. Baron, "Multiprocessor approximate message passing with column-wise partitioning," in *Proc. 2017 IEEE Int. Conf. Acoust. Speech Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 4765–4769.
[41] M. Guo and M. C. Gursoy, "Joint activity detection and channel estimation in cell-free massive MIMO networks with massive connectivity," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 317–331, Jan. 2022.
[42] J. Bai and E. G. Larsson, "Activity detection in distributed MIMO: Distributed AMP via likelihood ratio fusion," *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2200–2204, Oct. 2022.
[43] R. Hayakawa, A. Nakai, and K. Hayashi, "Distributed approximate message passing with summation propagation," in *Proc. 2018 IEEE Int. Conf. Acoust. Speech Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 4104–4108.
[44] C. C. Moallemi and B. V. Roy, "Consensus propagation," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 4753–4766, Nov. 2006.
[45] K. Takeuchi, "On the convergence of orthogonal/vector AMP: Long-memory message-passing strategy," in *Proc. 2022 IEEE Int. Symp. Inf. Theory*, Espoo, Finland, Jun.–Jul. 2022, pp. 1366–1371.
[46] ——, "On the convergence of orthogonal/vector AMP: Long-memory message-passing strategy," *IEEE Trans. Inf. Theory*, vol. 68, no. 12, pp. 8121–8138, Dec. 2022.
[47] L. Liu, S. Huang, and B. M. Kurkoski, "Sufficient statistic memory approximate message passing," in *Proc. 2022 IEEE Int. Symp. Inf. Theory*, Espoo, Finland, Jun.–Jul. 2022, pp. 1378–1383.
[48] K. Takeuchi, "Convolutional approximate message-passing," *IEEE Signal Process. Lett.*, vol. 27, pp. 416–420, 2020.
[49] ——, "Bayes-optimal convolutional AMP," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4405–4428, Jul. 2021.
[50] Z. Fan, "Approximate message passing algorithms for rotationally invariant matrices," *Ann. Statist.*, vol. 50, no. 1, pp. 197–224, Feb. 2022.
[51] R. Venkataramanan, K. Kögler, and M. Mondelli, "Estimation in rotationally invariant generalized linear models via approximate message passing," in *Proc. 39th Int. Conf. Mach. Learn.*, Baltimore, MD, USA, Jul. 2022.
[52] N. Skuratovs and M. E. Davies, "Compressed sensing with upscaled vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 68, no. 7, pp. 4818–4836, Jul. 2022.
[53] L. Liu, S. Huang, and B. M. Kurkoski, "Memory AMP," *IEEE Trans. Inf. Theory*, vol. 68, no. 12, pp. 8015–8039, Dec. 2022.
[54] K. Takeuchi, "Decentralized generalized approximate message-passing for tree-structured networks," submitted to *2024 IEEE Int. Conf. Acoust. Speech Signal Process.*
[55] ——, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 368–386, Jan. 2020.
[56] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proc. 15th Conf. Uncertain. Artif. Intell.*, Stockholm, Sweden, Jul.–Aug. 1999, pp. 467–475.
[57] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *Proc. 2015 IEEE Int. Conf. Acoust. Speech Signal Process.*, South Brisbane, Australia, Apr. 2015, pp. 2021–2025.
[58] S. Rangan, P. Schniter, A. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5339–5351, Sep. 2019.
[59] T. Yoshida and K. Takeuchi, "Deep learning of damped AMP decoding networks for sparse superposition codes via annealing," *IEICE Trans. Fundamentals.*, vol. E106-A, no. 3, pp. 414–421, Mar. 2023.
[60] 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.