On the Convergence of Orthogonal/Vector AMP: Long-Memory Message-Passing Strategy

Keigo Takeuchi[iD], *Member, IEEE*

Abstract— Orthogonal/vector approximate message-passing (AMP) is a powerful message-passing (MP) algorithm for signal reconstruction in compressed sensing. This paper proves the convergence of Bayes-optimal orthogonal/vector AMP in the large system limit. The proof strategy is based on a novel long-memory (LM) MP approach: A first step is a construction of LM-MP that is guaranteed to converge systematically. A second step is a large-system analysis of LM-MP via an existing framework of state evolution. A third step is to prove the convergence of state evolution recursions for Bayes-optimal LM-MP via a new statistical interpretation of existing LM damping. The last is an exact reduction of the state evolution recursions for Bayes-optimal LM-MP to those for Bayes-optimal orthogonal/vector AMP. The convergence of the state evolution recursions for Bayes-optimal LM-MP implies that for Bayes-optimal orthogonal/vector AMP. Numerical simulations are presented to show the verification of state evolution results for damped orthogonal/vector AMP and a negative aspect of LM-MP in finite-sized systems.

Index Terms— Compressed sensing, orthogonal/vector approximate message-passing (AMP), long memory, damping, state evolution, convergence analysis.

I. INTRODUCTION

A. Motivation

CONSIDER a reconstruction of an N-dimensional sparse signal vector $\boldsymbol{x} \in \mathbb{R}^N$ with $N^{-1}\mathbb{E}[\|\boldsymbol{x}\|^2] = 1$ from compressed, noisy, and linear measurements $\boldsymbol{y} \in \mathbb{R}^M$ [1], [2] with $M \leq N$, given by

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}. \tag{1}$$

In (1), $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is a known sensing matrix. The vector $\boldsymbol{w} \in \mathbb{R}^M$ denotes an additive noise vector with $\sigma^2 = M^{-1}\mathbb{E}[\|\boldsymbol{w}\|^2]$. A goal of compressed sensing is to reconstruct the signal vector $\boldsymbol{x}$ from the knowledge on the sensing matrix $\boldsymbol{A}$ and the measurement vector $\boldsymbol{y}$.

Compressed sensing has been applied to several practical issues, such as magnetic resonance imaging [3], radar [4], image super-resolution [5], channel estimation [6], spectrum sensing [7], error-correcting codes [8], and multiuser

detection [9]. The prior distribution of the sparse signal vector is known in specific instances, such as error-correcting codes or multiuser detection, while it is unknown in many signal processing instances. This paper is more relevant to such specific instances—the prior distribution is known—since the knowledge on the prior distribution is utilized in the theoretical analysis of iterative reconstruction algorithms.

Approximate message-passing (AMP) [10], [11] is a powerful message-passing (MP) algorithm for signal reconstruction. AMP with Bayes-optimal denoiser—called Bayes-optimal AMP—can be regarded as an exact approximation of loopy belief propagation in the large system limit [12], where both M and N tend to infinity while the compression ratio $\delta = M/N \in (0, 1]$ is kept constant. Furthermore, Bayes-optimal AMP was proved to achieve the Bayes-optimal performance [13]–[16] for zero-mean, independent and identically distributed (i.i.d.), and sub-Gaussian sensing matrices [17], [18] if the fixed-point (FP) equations describing the Bayes-optimal performance have a unique solution. Spatial coupling [19] is required when the FP equations have multiple solutions. See [20]–[23] for the details.

The main weakness of AMP is that AMP does not converge for sensing matrices beyond zero-mean i.i.d. matrices, such as non-zero mean matrices [24] or ill-conditioned matrices [25]. To tackle this convergence issue, several modified MP algorithms [26]–[31] were proposed. In particular, orthogonal AMP (OAMP) [30] or equivalently vector AMP (VAMP) [31] is a promising algorithm to solve the weakness of AMP. Since they are equivalent to each other in the linear measurement system (1), the two MP algorithms are referred to as OAMP in this paper.

The main feature of OAMP is the use of *extrinsic* messages to eliminate intractable dependencies between the current message and all preceding messages. The extrinsic messages were originally used in a single loop algorithm to solve a FP of the expectation consistency (EC) free energy [32], inspired by expectation propagation [33]. In fact, Bayes-optimal OAMP can be regarded as an exact approximation of EP in the large system limit [34], [35].

The asymptotic performance of OAMP can be characterized via rigorous state evolution [31], [35]. Bayes-optimal OAMP was proved to achieve the Bayes-optimal performance [36]–[38] for all orthogonally invariant sensing matrices if the FP equations describing the Bayes-optimal performance have a unique solution. The class of orthogonally invariant matrices is a general class beyond zero-mean i.i.d. sensing matrices.

Strictly speaking, state evolution [17], [18], [31], [35] does not guarantee the convergence[1] of MP algorithms for AMP or OAMP while it describes the asymptotic dynamics of these MP algorithms rigorously. As a result, we need to investigate the convergence of state evolution recursions to a FP for individual problems, e.g. [39]–[41].

For finite-sized systems, damping [25] is used to improve the convergence property of MP algorithms. Conventional state evolution cannot describe the dynamics of damped AMP or OAMP anymore. To the best of author's knowledge, theoretical analysis for damped MP algorithms are very limited [25], [42].

This paper addresses these two issues by proposing a novel damped MP algorithm that realizes the convergence *in principle* and the exact description for its dynamics. The damping issue has been already solved in [43] via the framework of long-memory (LM) MP [44]. LM-MP utilizes all preceding messages in computing the current message while conventional MP uses messages only in the latest iteration to update the current message. LM-MP was originally used to solve the Thouless-Anderson-Palmer (TAP) equation for Ising models with orthogonally invariant spin interaction via non-rigorous dynamical functional theory [45]. In this direction, Fan [46] proposed an LM-MP algorithm and analyzed the dynamics of the LM-MP algorithm rigorously. See [47] for a generalization of [46].

The state evolution framework for LM-MP in [44] is fully general, so that rigorous state evolution results for LM-MP can be obtained just by confirming whether the model of estimation errors for the LM-MP is included into a general error model proposed in [44]. In this direction, convolutional AMP (CAMP) [44], [48] was proposed as an LM generalization of AMP. The framework was used to construct VAMP with warm-started conjugate gradient (WS-CG) in [49]. To solve an convergence issue in CAMP for sensing matrices with high condition numbers [50], memory AMP (MAMP) [43] was proposed. State evolution results in [43] can describe the exact dynamics of MAMP even if damping is used. Since the main purpose of these LM-MP algorithms is to reduce the complexity of OAMP, such LM-MP algorithms are out of the scope of this paper.

The "convergence in principle" is the main idea to prove the convergence of OAMP. It means that the convergence of an iterative algorithm is intuitively trivial while the convergence of conventional MP algorithms is non-trivial and therefore needs to be proved for individual problems. The current message in LM-MP can be regarded as an additional measurement for the signal vector that is correlated with all preceding messages. When all measurements are utilized in a *Bayes-optimal* manner to estimate the signal vector, using the additional measurements never degrades the estimation performance. As a result, the performance of the obtained Bayes-optimal LM-MP is monotonically convergent in principle as the iteration proceeds.

To design the Bayes-optimal estimation of the signal vector in each iteration, we start with an LM generalization of OAMP—called LM-OAMP. The estimation errors in LM-OAMP are proved to be jointly Gaussian-distributed for all iterations in the large system limit via the unified framework of state evolution [44]. Thus, the design issue of Bayes-optimal LM-OAMP reduces to a solvable issue, i.e. the Bayes-optimal estimation of the signal vector from correlated additive white Gaussian noise (AWGN) measurements. The convergence of Bayes-optimal OAMP is guaranteed by proving the equivalence between Bayes-optimal LM-OAMP and OAMP.

B. Contributions

The main contributions of this paper are threefold. In particular, the first two contributions were presented in [51].

A first contribution is a statistical interpretation for the optimization of damping in [43]. This paper derives the optimized damping in terms of a sufficient statistic, without using state evolution recursions, while it was obtained via an optimization problem for state evolution recursions [43]. This statistical interpretation is a key technical tool to prove the convergence of Bayes-optimal LM-OAMP rigorously.

A second contribution is a derivation of state evolution recursions for LM-OAMP (Theorems 1 and 2) and a rigorous convergence analysis of state evolution recursions for Bayes-optimal LM-OAMP (Theorem 3), as well as an exact reduction of Bayes-optimal LM-OAMP to conventional Bayes-optimal OAMP [30]. These results provide a rigorous proof for the convergence of conventional Bayes-optimal OAMP in the large system limit.

The last contribution is a negative aspect of LM-OAMP. Numerical simulations show that LM-OAMP requires larger systems than conventional OAMP for state evolution to provide a good approximation for finite-sized systems. As a result, damped LM-OAMP is slightly inferior to OAMP with heuristic damping for small-to-moderate system sizes while it is consistent with state evolution for large systems. This observation explains why MAMP with the optimized LM damping [43] needs large systems to realize good convergence properties.

C. Organization

After summarizing the notation used in this paper, Section II presents the statistical interpretation and technical results in the first contribution as preliminaries to propose LM-OAMP in Section III. Section IV presents state evolution results for LM-OAMP. The convergence of state evolution recursions for Bayes-optimal LM-OAMP is analyzed in Section V. Numerical simulations for damped LM-OAMP are presented in Section VI. Section VII concludes this paper. Several theorems and lemmas are proved in the appendices.

D. Notation

Throughout this paper, M^{T}, $\mathrm{Tr}(M)$, and $\det M$ denote the transpose, trace, and determinant of a matrix M, respectively. The notation $[M]_{t',t}$ represents the (t', t) element of M.

---

[1] Throughout this paper, we only consider the convergence of state evolution recursions for MP algorithms in the large system limit, which may be phrased as the convergence of MP algorithms in the large system limit or more simply as the convergence of MP algorithms. Thus, convergence for finite-sized systems is out of the scope of this paper.

For a vector $\boldsymbol{v}_t$ with a subscript t, the nth element of $\boldsymbol{v}_t$ is written as $v_{n,t}$. The notation $\|\cdots\|$ denotes the Euclidean norm. The vector $\boldsymbol{1}$ represents a vector whose elements are all one while $\boldsymbol{e}_n$ is the nth column of the identity matrix $\boldsymbol{I}$. The notations $\otimes$ and $\delta_{i,j}$ denote the Kronecker product and delta, respectively.

The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is written as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The notations $\overset{\text{a.s.}}{\to}$ and $\overset{\text{a.s.}}{=}$ represent the almost sure convergence and equivalence, respectively.

For a function $f : \mathbb{R}^t \to \mathbb{R}$ of t variables, the notation $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ means the element-wise application of f, i.e. $[f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)]_n = f([\boldsymbol{x}_1]_n, \ldots, [\boldsymbol{x}_t]_n)$. For a vector $\boldsymbol{v} \in \mathbb{R}^N$, the arithmetic mean of $\boldsymbol{v}$ is written as $\langle \boldsymbol{v} \rangle = N^{-1} \sum_{n=1}^{N} v_n$. Combining these notations, we find that $\langle f(\boldsymbol{x}, \boldsymbol{y}) \rangle$ means $N^{-1} \sum_{n=1}^{N} f(x_n, y_n)$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$.

For any finite set $\mathcal{T}$ of non-negative integers, we define the one-to-one mapping $i = I_{\mathcal{T}}(t) \in \{1, \ldots, |\mathcal{T}|\}$ to represent that $t \in \mathcal{T}$ is the ith minimum element in $\mathcal{T}$, e.g. for $\mathcal{T} = \{2, 3, 4\}$ we have $I_{\mathcal{T}}(2) = 1$, $I_{\mathcal{T}}(3) = 2$, and $I_{\mathcal{T}}(4) = 3$.

II. PRELIMINARIES

A. Measurement Model

The purpose of Section II is to present the technical background of LM-OAMP proposed in Section III. This section is separated from the other sections in terms of the notation.

Consider correlated $(t + 1)$ AWGN measurements $\{Y_\tau \in \mathbb{R}\}_{\tau=0}^{t}$ for a signal $X \in \mathbb{R}$. For a subset $\mathcal{T}_\tau \subset \{0, \ldots, \tau\}$, we reconstruct the signal X from partial measurements $\boldsymbol{Y}_\tau \in \mathbb{R}^{1 \times |\mathcal{T}_\tau|}$ that consist of $\{Y_{t'} : t' \in \mathcal{T}_\tau\}$, given by

$$\boldsymbol{Y}_\tau = X\boldsymbol{1}^{\text{T}} + \boldsymbol{W}_\tau, \tag{2}$$

where $\boldsymbol{W}_\tau = \{W_{t'} : t' \in \mathcal{T}_\tau\} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\tau)$ denotes a zero-mean Gaussian row vector with covariance $\mathbb{E}[\boldsymbol{W}_\tau^{\text{T}} \boldsymbol{W}_\tau] = \boldsymbol{\Sigma}_\tau \in \mathbb{R}^{|\mathcal{T}_\tau| \times |\mathcal{T}_\tau|}$. The covariance matrix $\boldsymbol{\Sigma}_\tau$ is assumed to be positive definite. When all measurements are used, i.e. $\mathcal{T}_\tau = \{0, \ldots, \tau\}$, $\boldsymbol{\Sigma}_\tau$ is the $(\tau + 1) \times (\tau + 1)$ upper-left submatrix in $\boldsymbol{\Sigma}_t$ for all $\tau < t$. The goal is to estimate the signal X from the knowledge on the measurement vector $\boldsymbol{Y}_\tau$.

We know that the posterior mean estimator $\mathbb{E}[X | \boldsymbol{Y}_\tau]$ is the best among all possible estimators $f(\boldsymbol{Y}_\tau)$ in terms of the mean-square error (MSE) $\mathbb{E}[\{X - f(\boldsymbol{Y}_\tau)\}^2]$. In computing the posterior mean estimator, we use a two-step approach: A first step is the derivation of a sufficient statistic S_τ for estimation of X based on the measurements $\boldsymbol{Y}_\tau$. The second step is to compute the conditional mean of X given the sufficient statistic S_τ, instead of the original measurements $\boldsymbol{Y}_\tau$. This two-step approach is useful in proving technical lemmas to establish the main results while it is equivalent to direct computation of the posterior mean estimator, i.e. $\mathbb{E}[X | \boldsymbol{Y}_\tau] = \mathbb{E}[X | S_\tau]$.

B. Sufficient Statistic

We evaluate a sufficient statistic for estimation of X based on the measurements $\boldsymbol{Y}_\tau$. By definition, the probability density function (pdf) of $\boldsymbol{Y}_\tau$ given X is defined as

$$p(\boldsymbol{Y}_\tau | X) = \frac{\exp\{-(\boldsymbol{Y}_\tau - X\boldsymbol{1}^{\text{T}})\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{Y}_\tau - X\boldsymbol{1}^{\text{T}})^{\text{T}}/2\}}{\{(2\pi)^{|\mathcal{T}_\tau|} \det \boldsymbol{\Sigma}_\tau\}^{1/2}}. \tag{3}$$

Evaluating the exponent in the pdf yields

$$\begin{aligned}&(\boldsymbol{Y}_\tau - X\boldsymbol{1}^{\text{T}})\boldsymbol{\Sigma}_\tau^{-1}(\boldsymbol{Y}_\tau - X\boldsymbol{1}^{\text{T}})^{\text{T}}\\&=\boldsymbol{Y}_\tau\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{Y}_\tau^{\text{T}} - 2X\boldsymbol{Y}_\tau\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1} + X^2\boldsymbol{1}^{\text{T}}\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1},\end{aligned} \tag{4}$$

which implies that $\boldsymbol{Y}_\tau\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}$ is a sufficient statistic for estimation of X given $\boldsymbol{Y}_\tau$. Normalizing this sufficient statistic, we arrive at

$$S_\tau = \frac{\boldsymbol{Y}_\tau\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\text{T}}\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}} = X + \tilde{W}_\tau, \tag{5}$$

with

$$\tilde{W}_\tau = \frac{\boldsymbol{W}_\tau\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\text{T}}\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}}. \tag{6}$$

It is straightforward to confirm that $\tilde{W}_\tau$ is a zero-mean Gaussian random variable with variance $\mathbb{E}[\tilde{W}_\tau^2]$, given by

$$\mathbb{E}[\tilde{W}_\tau^2] = \frac{1}{\boldsymbol{1}^{\text{T}}\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}}. \tag{7}$$

Furthermore, the correlation $\mathbb{E}[\tilde{W}_{\tau'}\tilde{W}_\tau]$ is given by

$$\mathbb{E}[\tilde{W}_{\tau'}\tilde{W}_\tau] = \frac{\boldsymbol{1}^{\text{T}}\boldsymbol{\Sigma}_{\tau'}^{-1}\mathbb{E}[\boldsymbol{W}_{\tau'}^{\text{T}}\boldsymbol{W}_\tau]\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\text{T}}\boldsymbol{\Sigma}_{\tau'}^{-1}\boldsymbol{1}\boldsymbol{1}^{\text{T}}\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}}. \tag{8}$$

When all measurements are used, i.e. $\mathcal{T}_\tau = \{0, \ldots, \tau\}$, the correlation (8) reduces to

$$\mathbb{E}[\tilde{W}_{\tau'}\tilde{W}_\tau] = \frac{1}{\boldsymbol{1}^{\text{T}}\boldsymbol{\Sigma}_\tau^{-1}\boldsymbol{1}} = \mathbb{E}[\tilde{W}_\tau^2] \tag{9}$$

for all $\tau' \leq \tau$, since $\mathbb{E}[\boldsymbol{W}_{\tau'}^{\text{T}}\boldsymbol{W}_\tau] = (\boldsymbol{I}_{\tau'+1}, \boldsymbol{O})\boldsymbol{\Sigma}_\tau$ holds for $\boldsymbol{W}_\tau = (W_0, \ldots, W_\tau)$. This property is used to prove the convergence of the error covariance in Bayes-optimal LM-OAMP via that of its diagonal elements, which are monotonically decreasing MSEs when we use the posterior mean estimator of the signal vector given all preceding messages.

Finally, we investigate the impact of small eigenvalues for $\boldsymbol{\Sigma}_\tau$ on the variance $\mathbb{E}[\tilde{W}_\tau^2]$.

Proposition 1: Suppose that $\mathbb{E}[\tilde{W}_\tau^2] > \sigma_0^2$ holds for some $\sigma_0^2 > 0$. For the eigen-decomposition $\boldsymbol{\Sigma}_\tau = \sum_{\tau'} \lambda_{\tau'} \boldsymbol{u}_{\tau'} \boldsymbol{u}_{\tau'}^{\text{T}}$, we have

$$\boldsymbol{1}^{\text{T}}\boldsymbol{u}_{\tau'} = \mathcal{O}(\sqrt{\lambda_{\tau'}}) \quad \text{as } \lambda_{\tau'} \downarrow 0. \tag{10}$$

Proof: Applying the eigen-decomposition of $\boldsymbol{\Sigma}_\tau^{-1}$ to the definition of $\mathbb{E}[\tilde{W}_\tau^2]$ in (7), we obtain

$$\mathbb{E}[\tilde{W}_\tau^2] = \left\{\sum_{\tau'} \frac{(\boldsymbol{1}^{\text{T}}\boldsymbol{u}_{\tau'})^2}{\lambda_{\tau'}}\right\}^{-1} > \sigma_0^2, \tag{11}$$

which implies $(\boldsymbol{1}^{\text{T}}\boldsymbol{u}_{\tau'})^2/\lambda_{\tau'} < \sigma_0^{-2}$ for all τ'. Thus, we arrive at (10). ∎

The quantity σ_0^2 in Proposition 1 may be obtained via the Bayes-optimal performance for the original compressed sensing problem (1). There exists $\sigma_0^2 > 0$ as long as noisy measurements are considered.

Proposition 1 implies that the inner product $\boldsymbol{1}^{\text{T}}\boldsymbol{u}_{\tau'}$ must tend to zero as some eigenvalue $\lambda_{\tau'} > 0$ goes to zero. In other

words, all eigenvectors associated with small eigenvalues have a tendency to be orthogonal to the vector $\mathbf{1}$. Unless this asymptotic orthogonality holds, error-free estimation of X is possible via the matched filter (MF) $\mathbf{Y}_\tau \mathbf{u}_{\tau'}/\mathbf{1}^{\mathrm{T}}\mathbf{u}_{\tau'} \to X$ in probability as $\lambda_{\tau'} \downarrow 0$.

C. Extrinsic Denoiser

The Bayes-optimal estimator is defined as the posterior mean estimator $f_{\mathrm{opt}}(S_t) = \mathbb{E}[X|S_t]$ of X given the sufficient statistic S_t. By definition, we have $\mathbb{E}[X|S_t] = \mathbb{E}[X|\mathbf{Y}_t]$. To realize asymptotic Gaussianity in LM-OAMP, we define the extrinsic denoiser $f_t^{\mathrm{ext}} : \mathbb{R}^{|\mathcal{T}_t|} \to \mathbb{R}$ as

$$f_t^{\mathrm{ext}}(\mathbf{Y}_t) = \frac{f_{\mathrm{opt}}(S_t) - \sum_{\tau \in \mathcal{T}_t} \xi_{\tau,t} Y_\tau}{1 - \sum_{\tau \in \mathcal{T}_t} \xi_{\tau,t}}, \tag{12}$$

where $\xi_{\tau,t}$ for $\tau \in \mathcal{T}_t$ denotes the averaged partial derivative of $f_{\mathrm{opt}}(S_t)$ with respect to Y_τ, given by

$$\xi_{\tau,t} = \mathbb{E}[f'_{\mathrm{opt}}(S_t)] \frac{\mathbf{e}_{I_{\mathcal{T}_t}(\tau)}^{\mathrm{T}} \mathbf{\Sigma}_t^{-1} \mathbf{1}}{\mathbf{1}^{\mathrm{T}} \mathbf{\Sigma}_t^{-1} \mathbf{1}}. \tag{13}$$

See Section I-D for the definition of the mapping $I_{\mathcal{T}_t}(\cdot)$. By definition, we have $\sum_{\tau \in \mathcal{T}_t} \xi_{\tau,t} = \mathbb{E}[f'_{\mathrm{opt}}(S_t)]$.

The numerator in the extrinsic denoiser (12) is the Onsager correction of the Bayes-optimal estimator $f_{\mathrm{opt}}(S_t)$, which originates from a general error model in [44]. The significance of the denominator is presented in the following proposition:

Proposition 2: Among all possible Lipschitz-continuous extrinsic denoisers $\psi_t : \mathbb{R}^{|\mathcal{T}_t|} \to \mathbb{R}$ satisfying $\mathbb{E}[\partial\psi(\mathbf{Y}_t)/\partial Y_\tau] = 0$ for all $\tau \in \mathcal{T}_t$, the extrinsic denoiser f_t^{ext} configured from the Bayes-optimal estimator minimizes $\mathbb{E}[\{\psi_t(\mathbf{Y}_t) - X\}^2]$.

Proof: See Appendix A-A. ∎

Proposition 2 implies that the denominator is the best option in terms of the MSE, as long as the Bayes-optimal denoiser is used.

D. Estimator for Error Covariance

For denoisers $f_{t'}$ and f_t, LM-OAMP needs a consistent estimator for the error covariance $\mathbb{E}[\{X - f_{t'}(S_{t'})\}\{X - f_t(S_t)\}]$ under the assumption of the known signal power $\mathbb{E}[X^2]$ and covariance $\mathbb{E}[\tilde{W}_{t'}\tilde{W}_t]$. When the Bayes-optimal estimators $f_{t'} = f_{\mathrm{opt}}(S_{t'})$ and $f_t = f_{\mathrm{opt}}(S_t)$ are used, we can use the posterior covariance $C(S_{t'}, S_t)$ as a consistent estimator, given by

$$C(S_{t'}, S_t) = \mathbb{E}[\{X - f_{\mathrm{opt}}(S_{t'})\}\{X - f_{\mathrm{opt}}(S_t)\}|S_{t'}, S_t]. \tag{14}$$

In particular, $C(S_t, S_t)$ reduces to the posterior variance $\mathbb{E}[\{X - f_{\mathrm{opt}}(S_t)\}^2|S_t]$. However, it is not straightforward to construct a consistent estimator for general denoisers.

We derive a consistent estimator of the error covariance for general $f_{t'}$ and f_t. For any t' and t, let $\{S_{t',n}, S_{t,n}\}_{n=1}^N$ denote independent samples that follow the same joint distribution as for the two sufficient statistics $(S_{t'}, S_t)$, i.e.

$$S_{t,n} = X_n + \tilde{W}_{t,n}, \tag{15}$$

where $\{X_n\}$ are independent random variables satisfying $X_n \sim X$, while $\{\tilde{W}_{t',n}, \tilde{W}_{t,n}\}_{n=1}^N$ are independent zero-mean

Gaussian random variables with covariance $\mathbb{E}[\tilde{W}_{t',n}\tilde{W}_{t,n}] = \mathbb{E}[\tilde{W}_{t'}\tilde{W}_t]$.

Lemma 1: Suppose that $f_{t'}$ and f_t are Lipschitz-continuous denoisers and let

$$\begin{aligned}\hat{C}(S_{t',n}, S_{t,n}) &= \mathbb{E}[X^2] + f_{t'}(S_{t',n})f_t(S_{t,n}) \\ &\quad + \mathbb{E}[\tilde{W}_{t'}\tilde{W}_t]f'_t(S_{t,n}) - S_{t',n}f_t(S_{t,n}) \\ &\quad + \mathbb{E}[\tilde{W}_{t'}\tilde{W}_t]f'_{t'}(S_{t',n}) - S_{t,n}f_{t'}(S_{t',n}).\end{aligned} \tag{16}$$

Then, the sample average of $\{\hat{C}(S_{t',n}, S_{t,n})\}_{n=1}^N$ is a consistent estimator: $N^{-1}\sum_{n=1}^N \hat{C}(S_{t',n}, S_{t,n})$ converges almost surely to the error covariance $\mathbb{E}[\{X - f_{t'}(S_{t'})\}\{X - f_t(S_t)\}]$ as $N \to \infty$.

Proof: See Appendix A-B. ∎

Corollary 1: Suppose that f_t is a Lipschitz-continuous denoiser and let

$$\hat{C}(S_{t,n}) = \mathbb{E}[X^2] + \mathbb{E}[\tilde{W}_t^2]f'_t(S_{t,n}) - S_{t,n}f_t(S_{t,n}). \tag{17}$$

Then, the sample average $N^{-1}\sum_{n=1}^N \hat{C}(S_{t,n})$ is a consistent estimator for $\mathbb{E}[X\{X - f_t(S_t)\}]$ as $N \to \infty$.

Proof: Use Lemma 1 for $S_{t'} = S_t$ and $f_{t'} = 0$. ∎

E. Properties of Bayes-Optimal Estimator

We investigate properties of the Bayes-optimal estimator $f_{\mathrm{opt}}(S_t) = \mathbb{E}[X|S_t]$. We have the following technical tools to analyze Bayes-optimal LM-OAMP:

Lemma 2: Let $\mathcal{T}_\tau = \{0, \ldots, \tau\}$ for all τ.
- The monotonicity $\mathbb{E}[\{X - f_{\mathrm{opt}}(S_{t'})\}^2] \geq \mathbb{E}[\{X - f_{\mathrm{opt}}(S_t)\}^2]$ and $\mathbb{E}[\tilde{W}_{t'}^2] \geq \mathbb{E}[\tilde{W}_t^2]$ holds for all $t' < t$.
- The identity $C(S_{t'}, S_t) = C(S_t, S_t)$ for the posterior covariance (14) holds for $t' \leq t$ if $\mathbb{E}[\tilde{W}_{t'}\tilde{W}_t] = \mathbb{E}[\tilde{W}_t^2]$ is satisfied.

Proof: See Appendix A-C. ∎

Lemma 3: Let $\mathcal{T}_\tau = \{0, \ldots, \tau\}$ for all τ and suppose that the covariance $\mathbf{\Sigma}_t$ for the noise vector $\mathbf{W}_t$ in (2) satisfies $[\mathbf{\Sigma}_t]_{\tau',\tau} = [\mathbf{\Sigma}_t]_{\tau,\tau'} = [\mathbf{\Sigma}_t]_{\tau,\tau}$ for all $\tau' < \tau$. If $\{\mathbf{\Sigma}_\tau\}_{\tau=1}^t$ are positive definite, the following properties hold:
- The monotonicity $[\mathbf{\Sigma}_t]_{\tau,\tau} > [\mathbf{\Sigma}_t]_{\tau+1,\tau+1}$ holds for all $\tau \in \{0, \ldots, t-1\}$.
- The identity $\mathbb{E}[\tilde{W}_t^2] = \mathbb{E}[W_t^2]$ holds for $\tilde{W}_t$ given in (6).
- The identities $S_t = Y_t$ and $\xi_{\tau,t} = 0$ hold for all $\tau \neq t$.

Proof: See Appendix A-D. ∎

The former property in Lemma 2 is used to prove the convergence of state evolution recursions for Bayes-optimal LM-OAMP. The latter property and Lemma 3 are useful in proving an exact reduction of Bayes-optimal LM-OAMP to conventional Bayes-optimal OAMP.

III. LONG-MEMORY OAMP

A. Outline

LM-OAMP is a generalization of conventional OAMP [30], [31]. OAMP consists of two modules—called modules A and B in this paper. Module A uses a linear filter to compute a posterior message of the signal vector $\mathbf{x}$ while module B utilizes a separable denoiser to refine the posterior message for each signal element. A crucial step in each
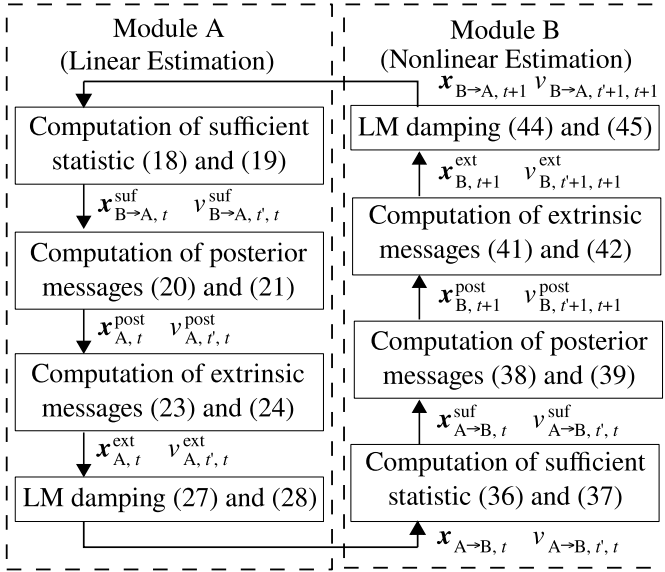
Fig. 1. Diagram of LM-OAMP.

module is the so-called Onsager correction to obtain extrinsic messages of $\boldsymbol{x}$. This step realizes asymptotic Gaussianity for the estimation errors in the two modules [31], [35]. The obtained messages may be damped in a heuristic manner to improve the convergence property of OAMP for finite-sized systems.

LM-OAMP consists of two modules in Fig. 1, of which each has four steps. In the first two steps, each module utilizes all preceding messages to compute a posterior message of $\boldsymbol{x}$ while OAMP only uses those in the latest iteration. A first step in computing the posterior message is computation of a sufficient statistic for estimation of $\boldsymbol{x}$ given all preceding messages, as presented in Section II. The second step computes a posterior message of $\boldsymbol{x}$ based on the sufficient statistic. This step corresponds to the computation of the posterior messages in conventional OAMP. While the first two steps are equivalent to direct computation of the posterior message given all preceding messages, the two-step approach simplifies the formulation of the Onsager correction in a third step.

To realize asymptotic Gaussianity in LM-OAMP, each module computes the Onsager correction of each posterior message. This third step is heavily based on a general framework of state evolution [44], which can evaluate asymptotic correlations between the current message and all preceding messages rigorously.

The last step is optional: LM damping [43] is employed to improve the convergence property of LM-OAMP for finite-sized systems. As long as the Onsager-corrected messages in the third step are damped, LM damping does not break asymptotic Gaussianity for the estimation errors while heuristic damping of posterior messages breaks asymptotic Gaussianity [44].

B. Notation

The notations used in LM-OAMP are summarized. Let $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t} \in \mathbb{R}^N$ and $\{v_{\mathrm{A}\to\mathrm{B},t',t} \in \mathbb{R}\}_{t'=0}^t$ denote mean and

covariance messages for the signal vector $\boldsymbol{x}$, respectively, passed from module A to module B in iteration t. The covariance $v_{\mathrm{A}\to\mathrm{B},\tau',\tau}$ corresponds to an estimate of the true error covariance $N^{-1}\mathbb{E}[(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},\tau'} - \boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},\tau} - \boldsymbol{x})]$ for all iterations $\tau', \tau \in \{0, \ldots, t\}$. For notational convenience, we define the infinite-dimensional real symmetric matrix $\boldsymbol{V}_{\mathrm{A}\to\mathrm{B}}$ satisfying $[\boldsymbol{V}_{\mathrm{A}\to\mathrm{B}}]_{\tau',\tau} = [\boldsymbol{V}_{\mathrm{A}\to\mathrm{B}}]_{\tau,\tau'} = v_{\mathrm{A}\to\mathrm{B},\tau',\tau}$. Messages passed in the opposite direction are written as $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t} \in \mathbb{R}^N$, $\{v_{\mathrm{B}\to\mathrm{A},t',t} \in \mathbb{R}\}_{t'=0}^t$, and $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A}}$.

The two modules utilize part of preceding messages in each iteration t. Let $\mathcal{T}_{\mathrm{A},t} \subset \{0, \ldots, t\}$ and $\mathcal{T}_{\mathrm{B},t} \subset \{0, \ldots, t\}$ denote subsets of indices that represent preceding messages used in iteration t for modules A and B, respectively. We assume that both subsets $\mathcal{T}_{\mathrm{A},t}$ and $\mathcal{T}_{\mathrm{B},t}$ contain the current index t.

We write the message matrix obtained by arranging the preceding messages $\{\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'} : t' \in \mathcal{T}_{\mathrm{B},t}\}$ as $\boldsymbol{X}_{\mathrm{A}\to\mathrm{B},t} \in \mathbb{R}^{N\times|\mathcal{T}_{\mathrm{B},t}|}$. More precisely, $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}$ for $t' \in \mathcal{T}_{\mathrm{B},t}$ is the ith column in $\boldsymbol{X}_{\mathrm{A}\to\mathrm{B},t}$ with $i = I_{\mathcal{T}_{\mathrm{B},t}}(t')$, defined in Section I-D. Furthermore, we define $\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t',t} \in \mathbb{R}^{|\mathcal{T}_{\mathrm{B},t'}|\times|\mathcal{T}_{\mathrm{B},t}|}$ as the matrix obtained by extracting the (τ', τ) elements from $\boldsymbol{V}_{\mathrm{A}\to\mathrm{B}}$ for all $(\tau', \tau) \in \mathcal{T}_{\mathrm{B},t'} \times \mathcal{T}_{\mathrm{B},t}$.

Similarly, we define the message matrices $\boldsymbol{X}_{\mathrm{B}\to\mathrm{A},t}$ and $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t',t} \in \mathbb{R}^{|\mathcal{T}_{\mathrm{A},t'}|\times|\mathcal{T}_{\mathrm{A},t}|}$ passed in the opposite direction via the preceding messages $\{\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t'} : t' \in \mathcal{T}_{\mathrm{A},t}\}$ and $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A}}$. In particular, LM-OAMP reduces to conventional OAMP [30] when the subsets $\mathcal{T}_{\mathrm{A},t} = \mathcal{T}_{\mathrm{B},t} = \{t\}$ are used, while $\mathcal{T}_{\mathrm{A},t} = \mathcal{T}_{\mathrm{B},t} = \{0, \ldots, t\}$ is the best in terms of estimation performance. Interestingly, this paper proves that Bayes-optimal LM-OAMP reduces to Bayes-optimal OAMP for $\mathcal{T}_{\mathrm{A},t} = \mathcal{T}_{\mathrm{B},t} = \{0, \ldots, t\}$.

Update rules in LM-OAMP are defined so as to realize asymptotic Gaussianity for $\{\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}\}$. While a mathematically rigorous definition for asymptotic Gaussianity is presented shortly, a rough intuition is to regard the estimation errors $\boldsymbol{h}_t = \boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t} - \boldsymbol{x}$ and $\boldsymbol{b}_t = \boldsymbol{V}^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t} - \boldsymbol{x})$ for the singular-value decomposition (SVD) $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ as zero-mean Gaussian vectors with covariance $\mathbb{E}[\boldsymbol{h}_\tau \boldsymbol{h}_{\tau'}^{\mathrm{T}}] = [\boldsymbol{V}_{\mathrm{A}\to\mathrm{B}}]_{\tau',\tau}\boldsymbol{I}_N$ and $\mathbb{E}[\boldsymbol{b}_\tau \boldsymbol{b}_{\tau'}^{\mathrm{T}}] = [\boldsymbol{V}_{\mathrm{B}\to\mathrm{A}}]_{\tau',\tau}\boldsymbol{I}_N$, respectively. These properties help us understand LM-OAMP intuitively while they are too strong to justify. Asymptotic Gaussianity proved in Theorem 1 is enough strong to prove the Bayes-optimality of LM-OAMP while it is weaker than the properties mentioned above.

C. Module A (Linear Estimation)

Module A consists of four steps in each iteration t. In the first step, module A uses the preceding message matrix $\boldsymbol{X}_{\mathrm{B}\to\mathrm{A},t}$ and the covariance messages $\{v_{\mathrm{B}\to\mathrm{A},\tau',\tau} : \tau', \tau \in \{0, \ldots, t\}\}$ to compute a sufficient statistic $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}^{\mathrm{suf}} \in \mathbb{R}^N$ for estimation of $\boldsymbol{x}$ and the corresponding covariance $\{v_{\mathrm{B}\to\mathrm{A},t',t}^{\mathrm{suf}}\}_{t'=0}^t$, given by

$$\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}^{\mathrm{suf}} = \frac{\boldsymbol{X}_{\mathrm{B}\to\mathrm{A},t}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}}, \tag{18}$$

$$v_{\mathrm{B}\to\mathrm{A},t',t}^{\mathrm{suf}} = \frac{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t',t'}^{-1}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t',t}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t',t'}^{-1}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}}. \tag{19}$$

For the initial iteration $t = 0$, we use $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},0} = \boldsymbol{0}$ and $v_{\mathrm{B}\to\mathrm{A},0,0} = \mathbb{E}[\|\boldsymbol{x}\|^2]/N$. These update rules are obtained by regarding the estimation errors $\{\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'} - \boldsymbol{x}\}$ as zero-mean Gaussian vectors with covariance $\mathbb{E}[(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},\tau} - \boldsymbol{x})(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},\tau'} - \boldsymbol{x})^{\mathrm{T}}] = [\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t}]_{\tau',\tau}\boldsymbol{I}_N$. See Section II for the derivation.

The covariance matrix $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}$ is guaranteed to be asymptotically positive definite via state evolution as long as t is finite. However, $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}$ has small positive eigenvalues that converge to zero as $t \to \infty$. Nonetheless, such small eigenvalues causes no numerical issues in computing $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}$ since the eigenvectors associated with such small eigenvalues should be orthogonal to the vector $\boldsymbol{1}$, as proved in Proposition 1.

Remark 1: The sufficient statistic (18) is a normalized linear combination of $\{\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t'}\}_{t'=0}^{t}$ while it might not be a convex combination. Thus, (18) can be regarded as a damped estimate in module B. In fact, (18) was originally obtained via the optimization of damping [43]. Nonetheless, the interpretation in terms of a sufficient statistic is important in analyzing the convergence of LM-OAMP as a technical tool.

The second step is computation of the posterior mean $\boldsymbol{x}_{\mathrm{A},t}^{\mathrm{post}} \in \mathbb{R}^N$ and covariance $\{v_{\mathrm{A},t',t}^{\mathrm{post}} \in \mathbb{R}\}_{t'=0}^{t}$. A linear filter $\boldsymbol{W}_t \in \mathbb{R}^{M \times N}$ is applied to the residual after interference subtraction based on the sufficient statistic (18).

$$\boldsymbol{x}_{\mathrm{A},t}^{\mathrm{post}} = \boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}^{\mathrm{suf}} + \boldsymbol{W}_t^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}^{\mathrm{suf}}), \qquad (20)$$

$$v_{\mathrm{A},t',t}^{\mathrm{post}} = \gamma_{t',t}v_{\mathrm{B}\to\mathrm{A},t',t}^{\mathrm{suf}} + \frac{\sigma^2}{N}\mathrm{Tr}\left(\boldsymbol{W}_{t'}\boldsymbol{W}_t^{\mathrm{T}}\right), \qquad (21)$$

with

$$\gamma_{t',t} = \frac{1}{N}\mathrm{Tr}\left\{\left(\boldsymbol{I}_N - \boldsymbol{W}_{t'}^{\mathrm{T}}\boldsymbol{A}\right)^{\mathrm{T}}\left(\boldsymbol{I}_N - \boldsymbol{W}_t^{\mathrm{T}}\boldsymbol{A}\right)\right\}. \qquad (22)$$

The formula (21) for the posterior covariance is justified in the large system limit via state evolution while the posterior mean (20) is equivalent to that in conventional OAMP [30].

The third step is the Onsager correction to realize asymptotic Gaussianity. Module A computes the extrinsic mean $\boldsymbol{x}_{\mathrm{A},t}^{\mathrm{ext}} \in \mathbb{R}^N$ and covariance $\{v_{\mathrm{A},t',t}^{\mathrm{ext}}\}_{t'=0}^{t}$, given by

$$\boldsymbol{x}_{\mathrm{A},t}^{\mathrm{ext}} = \frac{\boldsymbol{x}_{\mathrm{A},t}^{\mathrm{post}} - \sum_{t'\in\mathcal{T}_{\mathrm{A},t}}\xi_{\mathrm{A},t',t}\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t'}}{1 - \xi_{\mathrm{A},t}}, \qquad (23)$$

$$v_{\mathrm{A},t',t}^{\mathrm{ext}} = \frac{v_{\mathrm{A},t',t}^{\mathrm{post}} - \xi_{\mathrm{A},t'}\xi_{\mathrm{A},t}v_{\mathrm{B}\to\mathrm{A},t',t}^{\mathrm{suf}}}{(1 - \xi_{\mathrm{A},t'})(1 - \xi_{\mathrm{A},t})}, \qquad (24)$$

where $\xi_{\mathrm{A},t',t} \in \mathbb{R}$ is given by

$$\xi_{\mathrm{A},t',t} = \xi_{\mathrm{A},t}\frac{\boldsymbol{e}_{I_{\mathcal{T}_{\mathrm{A},t}}(t')}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}} \qquad (25)$$

for $t' \in \mathcal{T}_{\mathrm{A},t}$, with

$$\xi_{\mathrm{A},t} = \frac{1}{N}\mathrm{Tr}\left(\boldsymbol{I}_N - \boldsymbol{W}_t^{\mathrm{T}}\boldsymbol{A}\right). \qquad (26)$$

The factor $\xi_{\mathrm{A},t',t}$ is the empirical average for the partial derivatives of the posterior mean $\boldsymbol{x}_{\mathrm{A},t}^{\mathrm{post}}$ with respect to the elements in $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t'}$. The formula (24) for the covariance is justified via state evolution.

The numerator in (23) is the Onsager correction of the posterior message $\boldsymbol{x}_{\mathrm{A},t}^{\mathrm{post}}$ to realize asymptotic Gaussianity, and

originates from a general error model in [44]. The denominator in (23) has been selected to simplify state evolution, as considered in VAMP [31]. In particular, this selection is the best when the true posterior mean estimator is used. See Section II-C for the details.

The last step is LM damping [43] of the extrinsic messages. Module A feeds the damped messages $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}$ and $\{v_{\mathrm{A}\to\mathrm{B},t',t}\}_{t'=0}^{t}$ forward to module B.

$$\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t} = \sum_{\tau=0}^{t}\theta_{\mathrm{A},\tau,t}\boldsymbol{x}_{\mathrm{A},\tau}^{\mathrm{ext}}, \qquad (27)$$

$$v_{\mathrm{A}\to\mathrm{B},t',t} = \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\theta_{\mathrm{A},\tau',t'}\theta_{\mathrm{A},\tau,t}v_{\mathrm{A},\tau',\tau}^{\mathrm{ext}}, \qquad (28)$$

where the damping factors $\{\theta_{\mathrm{A},\tau,t}\}$ satisfy the following normalization condition:

$$\sum_{\tau=0}^{t}\theta_{\mathrm{A},\tau,t} = 1, \quad \theta_{\mathrm{A},t,t} \neq 0 \qquad (29)$$

for all t, where the condition $\theta_{\mathrm{A},t,t} \neq 0$ circumvents rank-deficient $\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}$. Note that the positivity of the damping factors is not assumed since it is not required in state evolution.

The main novelty for module A in LM-OAMP is in the sufficient statistic (18), the correction (23) for the posterior message to realize asymptotic Gaussianity, and the update rules (19), (21), (24), and (28) for the covariance messages. In particular, the sufficient statistic is a key technical tool to analyze the convergence of state evolution recursions for LM-OAMP rigorously.

Example 1 (LMMSE): Consider the linear minimum mean-square error (LMMSE) filter

$$\boldsymbol{W}_t = v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}\left(\sigma^2\boldsymbol{I}_M + v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}. \qquad (30)$$

We know that the LMMSE filter minimizes the posterior variance $v_{\mathrm{A},t,t}^{\mathrm{post}}$ in (21). Let $\eta(x)$ denote the η-transform [52] of the empirical eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, given by

$$\eta(x) = \frac{1}{N}\mathrm{Tr}\left\{\left(\boldsymbol{I}_N + x\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\right\}. \qquad (31)$$

It is straightforward to confirm that the variance $v_{\mathrm{A},t,t}^{\mathrm{post}}$ in (21) reduces to

$$v_{\mathrm{A},t,t}^{\mathrm{post}} = \xi_{\mathrm{A},t}v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}, \qquad (32)$$

where $\xi_{\mathrm{A},t}$ in (26) is given by

$$\xi_{\mathrm{A},t} = \eta\left(\frac{v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}}{\sigma^2}\right). \qquad (33)$$

For $t' \neq t$, on the other hand, $v_{\mathrm{A},t',t}^{\mathrm{post}}$ reduces to

$$v_{\mathrm{A},t',t}^{\mathrm{post}} = \gamma_{t',t}v_{\mathrm{B}\to\mathrm{A},t',t}^{\mathrm{suf}} + \frac{v_{\mathrm{B}\to\mathrm{A},t',t'}^{\mathrm{suf}}v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}}{v_{\mathrm{B}\to\mathrm{A},t',t'}^{\mathrm{suf}} - v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}}$$
$$\cdot\left\{\eta\left(\frac{v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}}{\sigma^2}\right) - \eta\left(\frac{v_{\mathrm{B}\to\mathrm{A},t',t'}^{\mathrm{suf}}}{\sigma^2}\right)\right\}, \qquad (34)$$

where $\gamma_{t',t}$ in (22) is given by

$$\gamma_{t',t} = \frac{v_{\mathrm{B}\to\mathrm{A},t',t'}^{\mathrm{suf}}}{v_{\mathrm{B}\to\mathrm{A},t',t'}^{\mathrm{suf}} - v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}} \eta\left(\frac{v_{\mathrm{B}\to\mathrm{A},t',t'}^{\mathrm{suf}}}{\sigma^2}\right)$$
$$- \frac{v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}}{v_{\mathrm{B}\to\mathrm{A},t',t'}^{\mathrm{suf}} - v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}} \eta\left(\frac{v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}}{\sigma^2}\right). \quad (35)$$

D. Module B (Nonlinear Estimation)

Module B is formulated in the same manner as in module A. The only exception is element-wise nonlinear denoising while module A has used the linear filter to mitigate the interference. Thus, the same explanation is omitted as in module A.

In the first step, module B uses the preceding message matrix $\boldsymbol{X}_{\mathrm{A}\to\mathrm{B},t}$ and covariance $\{v_{\mathrm{A}\to\mathrm{B},\tau',\tau} : \tau',\tau \in \{0,\ldots,t\}\}$ in iteration t to compute a sufficient statistic $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}}$ and the corresponding covariance $\{v_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}}\}_{t'=0}^{t}$, given by

$$\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}} = \frac{\boldsymbol{X}_{\mathrm{A}\to\mathrm{B},t}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\boldsymbol{1}}, \quad (36)$$

$$v_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}} = \frac{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t',t'}^{-1}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t',t}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t',t'}^{-1}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\boldsymbol{1}}. \quad (37)$$

The second step is computation of the posterior mean $\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}} \in \mathbb{R}^N$ and covariance[2] $\{v_{\mathrm{B},t'+1,t+1}^{\mathrm{post}}\}_{t'=0}^{t}$, which are defined with a scalar denoiser $f_t : \mathbb{R} \to \mathbb{R}$ as

$$\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}} = f_t(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}}), \quad (38)$$

$$v_{\mathrm{B},t'+1,t+1}^{\mathrm{post}} = 1 + \frac{1}{N}f_{t'}^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}^{\mathrm{suf}})f_t(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}})$$
$$+ v_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}}\xi_{\mathrm{B},t} - \frac{1}{N}f_t^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}})\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}^{\mathrm{suf}}$$
$$+ v_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}}\xi_{\mathrm{B},t'} - \frac{1}{N}f_{t'}^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}^{\mathrm{suf}})\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}}, \quad (39)$$

with

$$\xi_{\mathrm{B},t} = \langle f_t'(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}})\rangle. \quad (40)$$

The posterior mean message $\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}}$ is used as an estimator of the signal vector $\boldsymbol{x}$. Since the sufficient statistic $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}}$ depends on the preceding messages $\boldsymbol{X}_{\mathrm{A}\to\mathrm{B},t}$, the posterior message $\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}}$ is a function of the preceding message matrix $\boldsymbol{X}_{\mathrm{B}\to\mathrm{A},t}$. In this sense, module B utilizes LM denoising.

The posterior covariance (39) is a consistent estimator of the covariance $N^{-1}\mathbb{E}[\{\boldsymbol{x} - f_{t'}(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}^{\mathrm{suf}})\}^{\mathrm{T}}\{\boldsymbol{x} - f_t(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}})\}]$ for the estimation errors. See Section II-D for the details.

The third step is the Onsager correction to realize asymptotic Gaussianity. Module B computes the extrinsic mean $\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{ext}} \in \mathbb{R}^N$ and covariance $\{v_{\mathrm{B},t'+1,t+1}^{\mathrm{ext}}\}_{t'=0}^{t}$,

$$\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{ext}} = \frac{\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}} - \sum_{t'\in\mathcal{T}_{\mathrm{B},t}}\xi_{\mathrm{B},t',t}\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}}{1 - \xi_{\mathrm{B},t}}, \quad (41)$$

$$v_{\mathrm{B},t'+1,t+1}^{\mathrm{ext}} = \frac{v_{\mathrm{B},t'+1,t+1}^{\mathrm{post}} - \xi_{\mathrm{B},t'}\xi_{\mathrm{B},t}v_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}}}{(1 - \xi_{\mathrm{B},t'})(1 - \xi_{\mathrm{B},t})}, \quad (42)$$

[2] If the true prior distribution of $\boldsymbol{x}$ is available, (39) should be replaced with the posterior covariance $\langle C(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}^{\mathrm{suf}}, \boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}})\rangle$ in (14) while $v_{\mathrm{B},0,t+1}^{\mathrm{post}}$ should be set to the posterior variance $\langle C(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}}, \boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}})\rangle$.

with

$$\xi_{\mathrm{B},t',t} = \xi_{\mathrm{B},t}\frac{\boldsymbol{e}_{\mathcal{I}_{\mathrm{B},t}(t')}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\boldsymbol{1}} \quad (43)$$

for $t' \in \mathcal{T}_{\mathrm{B},t}$, where $\xi_{\mathrm{B},t}$ is given by (40).

The last step is LM damping of the extrinsic messages. Module B feeds the obtained messages $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t+1} \in \mathbb{R}^N$ and $\{v_{\mathrm{B}\to\mathrm{A},t',t+1}\}_{t'=0}^{t+1}$ back to module A. For $t',t \geq 0$,

$$\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t+1} = \sum_{\tau=0}^{t}\theta_{\mathrm{B},\tau,t}\boldsymbol{x}_{\mathrm{B},\tau+1}^{\mathrm{ext}}, \quad (44)$$

$$v_{\mathrm{B}\to\mathrm{A},t'+1,t+1} = \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\theta_{\mathrm{B},\tau',t'}\theta_{\mathrm{B},\tau,t}v_{\mathrm{B},\tau'+1,\tau+1}^{\mathrm{ext}}, \quad (45)$$

with

$$\sum_{\tau=0}^{t}\theta_{\mathrm{B},\tau,t} = 1, \quad \theta_{\mathrm{B},t,t} \neq 0. \quad (46)$$

Otherwise, for $t \geq 0$ we use

$$v_{\mathrm{B}\to\mathrm{A},0,t+1} = \sum_{\tau=0}^{t}\frac{\theta_{\mathrm{B},\tau,t}v_{\mathrm{B},0,\tau+1}^{\mathrm{post}}}{1 - \xi_{\mathrm{B},\tau}}, \quad (47)$$

with

$$v_{\mathrm{B},0,t+1}^{\mathrm{post}} = 1 + v_{\mathrm{A}\to\mathrm{B},t,t}^{\mathrm{suf}}\xi_{\mathrm{B},t} - \frac{1}{N}f_t^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}})\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}}. \quad (48)$$

IV. THEORETICAL ANALYSIS

A. State Evolution

We analyze the dynamics of LM-OAMP in the large system limit via state evolution [44]. A first step is a proof of asymptotic Gaussianity for the estimation errors in LM-OAMP. Asymptotic Gaussianity is proved by confirming that the error model for LM-OAMP is included into a general error model proposed in [44], which guarantees asymptotic Gaussianity. A second step is a derivation of state evolution recursions via asymptotic Gaussianity.

We first postulate technical assumptions to justify asymptotic Gaussianity.

Assumption 1: The signal vector $\boldsymbol{x}$ has i.i.d. elements with zero mean, unit variance, and bounded $(2 + \epsilon)$th moment for some $\epsilon > 0$.

Assumption 1 is an assumption to simplify state evolution analysis. We would need non-separable denoising if we postulated dependent signal elements. See [53]–[55] for non-separable denoising.

Assumption 2: The sensing matrix $\boldsymbol{A}$ is right-orthogonally invariant: For any orthogonal matrix $\boldsymbol{\Phi}$ independent of $\boldsymbol{A}$, the equivalence in distribution $\boldsymbol{A}\boldsymbol{\Phi} \sim \boldsymbol{A}$ holds. More precisely, for the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ the orthogonal matrix $\boldsymbol{V}$ is independent of $\boldsymbol{U}\boldsymbol{\Sigma}$ and Haar-distributed [52], [56]. Furthermore, the empirical eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ converges almost surely to a compactly supported deterministic distribution with unit first moment in the large system limit.

The unit-first-moment assumption is equivalent to the almost sure convergence $N^{-1}\mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) \overset{\mathrm{a.s.}}{\to} 1$ in the large system limit. The right-orthogonal invariance is an important assumption to justify asymptotic Gaussianity, for which a

rough intuition is as follows: Let $\boldsymbol{m} \in \mathbb{R}^N$ denote a vector independent of $\boldsymbol{V}$. Since $\boldsymbol{V}$ is Haar-distributed, the orthogonal transform $\boldsymbol{Vm}$ is distributed as $(\|\boldsymbol{m}\|/\|\boldsymbol{z}\|)\boldsymbol{z}$ for a standard Gaussian vector $\boldsymbol{z}$. If the amplitude $\|\boldsymbol{m}\|/\|\boldsymbol{z}\|$ converges in probability to a constant $a > 0$ as $N \to \infty$, $\boldsymbol{Vm}$ is distributed as if it followed $\mathcal{N}(\mathbf{0}, a^2 \boldsymbol{I}_N)$.

It is a challenging issue in random matrix theory to weaken the right-orthogonal invariance to more practical assumptions, including discrete cosine transform (DCT) or Hadamard matrices with random permutation [57]. See [58]–[60] for theoretical progress in this direction.

Assumption 3: The noise vector $\boldsymbol{w}$ is orthogonally invariant: For any orthogonal matrix $\boldsymbol{\Phi}$ independent of $\boldsymbol{w}$, the equivalence in distribution $\boldsymbol{\Phi w} \sim \boldsymbol{w}$ holds. Furthermore, the almost sure convergence $M^{-1}\|\boldsymbol{w}\|^2 \overset{\text{a.s.}}{\to} \sigma^2$ holds for the variance $\sigma^2 > 0$. The vector $\boldsymbol{w}$ has bounded $(2+\epsilon)$th moments for some $\epsilon > 0$.

The AWGN vector $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_M)$ with variance σ^2 satisfies Assumption 3. When the sensing matrix is left-orthogonally invariant, the orthogonal invariance in $\boldsymbol{w}$ can be induced from $\boldsymbol{A}$.

Assumption 4: The linear filter matrix $\boldsymbol{W}_t$ in module A has the same SVD structure as the sensing matrix $\boldsymbol{A}$, i.e. $\boldsymbol{W}_t = \boldsymbol{U}\tilde{\boldsymbol{W}}_t\boldsymbol{V}^{\text{T}}$ for $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\text{T}}$. Furthermore, the diagonal matrix $\tilde{\boldsymbol{W}}^{\text{T}}\tilde{\boldsymbol{W}}$ is in the space spanned by $\{\boldsymbol{\Lambda}^j\}_{j=0}^{\infty}$ with $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{\text{T}}\boldsymbol{\Sigma}$.

Assumption 4 contains practically important linear filters, such as the MF $\boldsymbol{W}_t = \boldsymbol{A}$ and the LMMSE filter (30).

Assumption 5: The scalar denoiser f_t in module B is Lipschitz-continuous and nonlinear.

The Lipschitz-continuity in Assumption 5 is a standard assumption in state evolution analysis [17], [31], [35], [44] while the nonlinearity is required for the positive definiteness of $\boldsymbol{V}_{\text{B}\to\text{A},t,t}$. In module A, Assumption 4 and the compact support in Assumption 2 are substituted for Assumption 5.

We are ready to present asymptotic Gaussianity for the estimation errors in LM-OAMP.

Theorem 1 (Asymptotic Gaussianity): Suppose that Assumptions 1–5 are satisfied and consider any iteration t. Then, the following results hold for all iterations $\tau', \tau \in \{0, \ldots, t\}$.

- The covariance $N^{-1}(\boldsymbol{x}_{\text{B}\to\text{A},\tau'} - \boldsymbol{x})^{\text{T}}(\boldsymbol{x}_{\text{B}\to\text{A},\tau} - \boldsymbol{x})$ converges almost surely to some constant $\bar{v}_{\text{B}\to\text{A},\tau',\tau}$ in the large system limit. Furthermore, the $(t+1) \times (t+1)$ upper-left block of the matrix $\bar{\boldsymbol{V}}_{\text{B}\to\text{A}}$ with $[\bar{\boldsymbol{V}}_{\text{B}\to\text{A}}]_{\tau',\tau} = \bar{v}_{\text{B}\to\text{A},\tau',\tau}$ is positive definite as long as t is finite.

- Let $\bar{\boldsymbol{V}}_{\text{B}\to\text{A},t',t} \in \mathbb{R}^{|\mathcal{T}_{\text{A},t'}| \times |\mathcal{T}_{\text{A},t}|}$ denote the matrix obtained by extracting the elements $\bar{v}_{\text{B}\to\text{A},\tau',\tau}$ for all $\tau' \in \mathcal{T}_{\text{A},t'}$ and $\tau \in \mathcal{T}_{\text{A},t}$ from the covariance matrix $\bar{\boldsymbol{V}}_{\text{B}\to\text{A}}$ defined in the first property. Then, the covariance $N^{-1}(\boldsymbol{x}_{\text{A},\tau'}^{\text{post}} - \boldsymbol{x})^{\text{T}}(\boldsymbol{x}_{\text{A},\tau}^{\text{post}} - \boldsymbol{x})$ converges almost surely to some constant $\bar{v}_{\text{A},\tau',\tau}^{\text{post}}$ in the large system limit, given by

$$\bar{v}_{\text{A},\tau',\tau}^{\text{post}} = \lim_{M=\delta N\to\infty} \left\{ \gamma_{\tau',\tau} \bar{v}_{\text{B}\to\text{A},\tau',\tau}^{\text{suf}} + \frac{\sigma^2}{N}\text{Tr}\left(\boldsymbol{W}_{\tau'}\boldsymbol{W}_\tau^{\text{T}}\right) \right\}, \quad (49)$$

where $\gamma_{\tau',\tau}$ is given by (22), with

$$\bar{v}_{\text{B}\to\text{A},\tau',\tau}^{\text{suf}} = \frac{\mathbf{1}^{\text{T}}\boldsymbol{V}_{\text{B}\to\text{A},\tau',\tau'}^{-1}\bar{\boldsymbol{V}}_{\text{B}\to\text{A},\tau',\tau}\boldsymbol{V}_{\text{B}\to\text{A},\tau,\tau}^{-1}\mathbf{1}}{\mathbf{1}^{\text{T}}\boldsymbol{V}_{\text{B}\to\text{A},\tau',\tau'}^{-1}\mathbf{11}^{\text{T}}\boldsymbol{V}_{\text{B}\to\text{A},\tau,\tau}^{-1}\mathbf{1}}. \quad (50)$$

- The covariance $N^{-1}(\boldsymbol{x}_{\text{A}\to\text{B},\tau'} - \boldsymbol{x})^{\text{T}}(\boldsymbol{x}_{\text{A}\to\text{B},\tau} - \boldsymbol{x})$ converges almost surely to some constant $\bar{v}_{\text{A}\to\text{B},\tau',\tau}$ in the large system limit. Furthermore, the $(t+1) \times (t+1)$ upper-left block of the matrix $\bar{\boldsymbol{V}}_{\text{A}\to\text{B}}$ with $[\bar{\boldsymbol{V}}_{\text{A}\to\text{B}}]_{\tau',\tau} = \bar{v}_{\text{A}\to\text{B},\tau',\tau}$ is positive definite as long as t is finite.

- Let $\bar{\boldsymbol{V}}_{\text{A}\to\text{B},t',t} \in \mathbb{R}^{|\mathcal{T}_{\text{B},t'}| \times |\mathcal{T}_{\text{B},t}|}$ denote the matrix defined by extracting the elements $\bar{v}_{\text{A}\to\text{B},\tau',\tau}$ for all $\tau' \in \mathcal{T}_{\text{B},t'}$ and $\tau \in \mathcal{T}_{\text{B},t}$ from the covariance matrix $\bar{\boldsymbol{V}}_{\text{A}\to\text{B}}$ defined in the third property. Then, the covariance $N^{-1}(\boldsymbol{x}_{\text{B},\tau'+1}^{\text{post}} - \boldsymbol{x})^{\text{T}}(\boldsymbol{x}_{\text{B},\tau+1}^{\text{post}} - \boldsymbol{x})$ converges almost surely to some constant $\bar{v}_{\text{B},\tau'+1,\tau+1}^{\text{post}}$ in the large system limit,

$$\bar{v}_{\text{B},\tau'+1,\tau+1}^{\text{post}} = \mathbb{E}[\{f_{\tau'}(x_1 + z_{\tau'}) - x_1\} \\ \cdot \{f_\tau(x_1 + z_\tau) - x_1\}], \quad (51)$$

where $\{z_{\tau'}, z_\tau\}$ are independent of the signal element x_1 and zero-mean Gaussian random variables with covariance $\mathbb{E}[z_{\tau'}z_\tau] = \bar{v}_{\text{A}\to\text{B},\tau',\tau}^{\text{suf}}$, given by

$$\bar{v}_{\text{A}\to\text{B},\tau',\tau}^{\text{suf}} = \frac{\mathbf{1}^{\text{T}}\boldsymbol{V}_{\text{A}\to\text{B},\tau',\tau'}^{-1}\bar{\boldsymbol{V}}_{\text{A}\to\text{B},\tau',\tau}\boldsymbol{V}_{\text{A}\to\text{B},\tau,\tau}^{-1}\mathbf{1}}{\mathbf{1}^{\text{T}}\boldsymbol{V}_{\text{A}\to\text{B},\tau',\tau'}^{-1}\mathbf{11}^{\text{T}}\boldsymbol{V}_{\text{A}\to\text{B},\tau,\tau}^{-1}\mathbf{1}}. \quad (52)$$

Furthermore, $\xi_{\text{B},\tau}$ in (40) converges almost surely to $\bar{\xi}_{\text{B},\tau}$ in the large system limit,

$$\bar{\xi}_{\text{B},\tau} = \mathbb{E}[f'_\tau(x_1 + z_\tau)]. \quad (53)$$

Proof: See Appendix B. ∎

The second and fourth results in Theorem 1 are the precise meaning of asymptotic Gaussianity for LM-OAMP while the asymptotic Gaussianity for the second result is not explicit, because of the linear filter in module A. We cannot claim the joint Gaussianity for the estimation errors: the convergence of the joint distribution of $\{\boldsymbol{x}_{\text{A}\to\text{B},\tau}^{\text{suf}} - \boldsymbol{x}\}_{\tau=0}^t$ toward a joint Gaussian distribution with covariance $\mathbb{E}[(\boldsymbol{x}_{\text{A}\to\text{B},\tau}^{\text{suf}} - \boldsymbol{x})(\boldsymbol{x}_{\text{A}\to\text{B},\tau'}^{\text{suf}} - \boldsymbol{x})^{\text{T}}] = \bar{v}_{\text{A}\to\text{B},\tau',\tau}^{\text{suf}}\boldsymbol{I}$. We only claim the almost sure convergence of $N^{-1}(\boldsymbol{x}_{\text{B},\tau'+1}^{\text{post}} - \boldsymbol{x})^{\text{T}}(\boldsymbol{x}_{\text{B},\tau+1}^{\text{post}} - \boldsymbol{x})$ toward (51) given via the zero-mean Gaussian random variables $z_{\tau'}$ and z_τ with covariance $\mathbb{E}[z_{\tau'}z_\tau] = \bar{v}_{\text{A}\to\text{B},\tau',\tau}^{\text{suf}}$.

We next derive state evolution recursions by evaluating the covariance $\bar{v}_{\text{A}\to\text{B},\tau',\tau}$ and $\bar{v}_{\text{B}\to\text{A},\tau',\tau}$ in Theorem 1.

Theorem 2 (State Evolution): Suppose that Assumptions 1–5 are satisfied. Then, the state evolution recursions for module A are given by

$$\bar{v}_{\text{B}\to\text{A},t',t}^{\text{suf}} = \frac{\mathbf{1}^{\text{T}}\bar{\boldsymbol{V}}_{\text{B}\to\text{A},t',t'}^{-1}\bar{\boldsymbol{V}}_{\text{B}\to\text{A},t',t}\bar{\boldsymbol{V}}_{\text{B}\to\text{A},t,t}^{-1}\mathbf{1}}{\mathbf{1}^{\text{T}}\bar{\boldsymbol{V}}_{\text{B}\to\text{A},t',t'}^{-1}\mathbf{11}^{\text{T}}\bar{\boldsymbol{V}}_{\text{B}\to\text{A},t,t}^{-1}\mathbf{1}}, \quad (54)$$

$$\bar{\xi}_{\text{A},t} = \lim_{M=\delta N\to\infty} \frac{1}{N}\text{Tr}\left(\boldsymbol{I}_N - \boldsymbol{W}_t^{\text{T}}\boldsymbol{A}\right), \quad (55)$$

$$\bar{v}_{\text{A},t',t}^{\text{ext}} = \frac{\bar{v}_{\text{A},t',t}^{\text{post}} - \bar{\xi}_{\text{A},t'}\bar{\xi}_{\text{A},t}\bar{v}_{\text{B}\to\text{A},t',t}^{\text{suf}}}{(1 - \bar{\xi}_{\text{A},t'})(1 - \bar{\xi}_{\text{A},t})}, \quad (56)$$

$$\bar{v}_{\text{A}\to\text{B},t',t} = \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\theta_{\text{A},\tau',t'}\theta_{\text{A},\tau,t}\bar{v}_{\text{A},\tau',\tau}^{\text{ext}}, \quad (57)$$

with $\bar{v}_{\mathrm{B}\to\mathrm{A},0,0} = 1$, where $\bar{v}_{\mathrm{A},t',t}^{\mathrm{post}}$ is defined in (49). In these expressions, $v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}$ in $\boldsymbol{W}_t$ is replaced with $\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}$. On the other hand, the state evolution recursions for module B are given by

$$\bar{v}_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}} = \frac{\mathbf{1}^{\mathrm{T}}\bar{\boldsymbol{V}}_{\mathrm{A}\to\mathrm{B},t',t'}^{-1}\bar{\boldsymbol{V}}_{\mathrm{A}\to\mathrm{B},t',t}\bar{\boldsymbol{V}}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\bar{\boldsymbol{V}}_{\mathrm{A}\to\mathrm{B},t',t'}^{-1}\mathbf{1}\mathbf{1}^{\mathrm{T}}\bar{\boldsymbol{V}}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\mathbf{1}}, \quad (58)$$

$$\bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{ext}} = \frac{\bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{post}} - \bar{\xi}_{\mathrm{B},t'}\bar{\xi}_{\mathrm{B},t}\bar{v}_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}}}{(1-\bar{\xi}_{\mathrm{B},t'})(1-\bar{\xi}_{\mathrm{B},t})}, \quad (59)$$

$$\bar{v}_{\mathrm{B}\to\mathrm{A},t'+1,t+1} = \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\theta_{\mathrm{B},\tau',t'}\theta_{\mathrm{B},\tau,t}\bar{v}_{\mathrm{B},\tau'+1,\tau+1}^{\mathrm{ext}} \quad (60)$$

for $t',t \geq 0$, where $\bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{post}}$ and $\bar{\xi}_{\mathrm{B},t}$ are defined as (51) and (53), respectively. For $t' = -1$, (60) is replaced with

$$\bar{v}_{\mathrm{B}\to\mathrm{A},0,t+1} = \sum_{\tau=0}^{t}\frac{\theta_{\mathrm{B},\tau,t}\bar{v}_{\mathrm{B},0,\tau+1}^{\mathrm{post}}}{1-\bar{\xi}_{\mathrm{B},\tau}} \quad (61)$$

for $t \geq 0$, with

$$\bar{v}_{\mathrm{B},0,\tau+1}^{\mathrm{post}} = \mathbb{E}[x_1\{x_1 - f_\tau(x_1 + z_\tau)\}], \quad (62)$$

where z_τ is independent of x_1 and a zero-mean Gaussian random variable with variance $\mathbb{E}[z_\tau^2] = \bar{v}_{\mathrm{A}\to\mathrm{B},\tau,\tau}^{\mathrm{suf}}$.

Proof: See Appendix C. ∎

Theorem 2 presents the state evolution recursions for LM-OAMP, which justify the update rules for the covariance messages in LM-OAMP. In particular, the MSE $N^{-1}\|\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}} - \boldsymbol{x}\|^2$ converges almost surely to $\bar{v}_{\mathrm{B},t+1,t+1}^{\mathrm{post}}$ in the large system limit.

Example 2: As an example of Theorem 2, we derive state evolution recursions for damped OAMP. Let $\mathcal{T}_{\mathrm{A},t} = \mathcal{T}_{\mathrm{B},t} = \{t\}$ and consider damping $\theta_{\mathrm{A},0,t} = (1-\theta_{\mathrm{A}})^t$, $\theta_{\mathrm{A},\tau,t} = \theta_{\mathrm{A}}(1-\theta_{\mathrm{A}})^{t-\tau}$ for all $\tau > 0$ in (27) with some $\theta_{\mathrm{A}} \in [0,1]$, and $\theta_{\mathrm{B},0,t} = (1-\theta_{\mathrm{B}})^t$, $\theta_{\mathrm{B},\tau,t} = \theta_{\mathrm{B}}(1-\theta_{\mathrm{B}})^{t-\tau}$ for all $\tau > 0$ in (44) with some $\theta_{\mathrm{B}} \in [0,1]$ [31]. From (27) we have $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},0} = \boldsymbol{x}_{\mathrm{A},0}^{\mathrm{ext}}$ and

$$\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t} = \theta_{\mathrm{A}}\sum_{\tau=1}^{t}(1-\theta_{\mathrm{A}})^{t-\tau}\boldsymbol{x}_{\mathrm{A},\tau}^{\mathrm{ext}} + (1-\theta_{\mathrm{A}})^t\boldsymbol{x}_{\mathrm{A},0}^{\mathrm{ext}}$$
$$= \theta_{\mathrm{A}}\boldsymbol{x}_{\mathrm{A},t}^{\mathrm{ext}} + (1-\theta_{\mathrm{A}})\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t-1} \quad (63)$$

for $t > 0$. Similarly, from (44) we have $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},1} = \boldsymbol{x}_{\mathrm{B}\to\mathrm{A},1}^{\mathrm{ext}}$ and for $t > 0$

$$\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t+1} = \theta_{\mathrm{B}}\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{ext}} + (1-\theta_{\mathrm{B}})\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}. \quad (64)$$

The state evolution recursion for module A in damped OAMP is given by

$$\bar{v}_{\mathrm{A},t',t}^{\mathrm{ext}} = \frac{\bar{v}_{\mathrm{A},t',t}^{\mathrm{post}} - \bar{\xi}_{\mathrm{A},t'}\bar{\xi}_{\mathrm{A},t}\bar{v}_{\mathrm{B}\to\mathrm{A},t',t}}{(1-\bar{\xi}_{\mathrm{A},t'})(1-\bar{\xi}_{\mathrm{A},t})}, \quad (65)$$

where $\bar{v}_{\mathrm{A},t',t}^{\mathrm{post}}$ is defined in (49) with $\bar{v}_{\mathrm{B}\to\mathrm{A},t',t}^{\mathrm{suf}} = \bar{v}_{\mathrm{B}\to\mathrm{A},t',t}$. Furthermore, we have $\bar{v}_{\mathrm{A}\to\mathrm{B},0,0} = \bar{v}_{\mathrm{A},0,0}^{\mathrm{ext}}$ and

$$\bar{v}_{\mathrm{A}\to\mathrm{B},0,t} = \theta_{\mathrm{A}}\bar{v}_{\mathrm{A},0,t}^{\mathrm{ext}} + (1-\theta_{\mathrm{A}})\bar{c}_{\mathrm{A},0,t-1}, \quad (66)$$

$$\bar{v}_{\mathrm{A}\to\mathrm{B},t',t} = \theta_{\mathrm{A}}\bar{c}_{\mathrm{A},t',t} + (1-\theta_{\mathrm{A}})\bar{v}_{\mathrm{A}\to\mathrm{B},t'-1,t}, \quad (67)$$

for all $t',t > 0$, with

$$l\bar{c}_{\mathrm{A},t',t} = \lim_{M=\delta N\to\infty}\frac{1}{N}(\boldsymbol{x}_{\mathrm{A},t'}^{\mathrm{ext}} - \boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t} - \boldsymbol{x})$$
$$\overset{\mathrm{a.s.}}{=} \theta_{\mathrm{A}}\bar{v}_{\mathrm{A},t',t}^{\mathrm{ext}} + (1-\theta_{\mathrm{A}})\bar{c}_{\mathrm{A},t',t-1}, \quad \bar{c}_{\mathrm{A},t',0} = \bar{v}_{\mathrm{A},t',0}^{\mathrm{ext}}. \quad (68)$$

The state evolution recursions for module B reduce to

$$\bar{v}_{\mathrm{B},0,t+1}^{\mathrm{ext}} = \frac{\bar{v}_{\mathrm{B},0,t+1}^{\mathrm{post}}}{1-\bar{\xi}_{\mathrm{B},t}}, \quad (69)$$

$$\bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{ext}} = \frac{\bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{post}} - \bar{\xi}_{\mathrm{B},t'}\bar{\xi}_{\mathrm{B},t}\bar{v}_{\mathrm{A}\to\mathrm{B},t',t}}{(1-\bar{\xi}_{\mathrm{B},t'})(1-\bar{\xi}_{\mathrm{B},t})} \quad (70)$$

for $t' \geq 0$, where $\bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{post}}$ is given by (51) with $\bar{v}_{\mathrm{A}\to\mathrm{B},\tau',\tau}^{\mathrm{suf}} = \bar{v}_{\mathrm{A}\to\mathrm{B},\tau',\tau}$. Furthermore, we have $\bar{v}_{\mathrm{B}\to\mathrm{A},0,0} = 1$, $\bar{v}_{\mathrm{B}\to\mathrm{A},0,1} = \bar{v}_{\mathrm{B},0,1}^{\mathrm{ext}}$, $\bar{v}_{\mathrm{B}\to\mathrm{A},1,1} = \bar{v}_{\mathrm{B},1,1}^{\mathrm{ext}}$, and

$$\bar{v}_{\mathrm{B}\to\mathrm{A},0,t+1} = \theta_{\mathrm{B}}\bar{v}_{\mathrm{B},0,t+1}^{\mathrm{ext}} + (1-\theta_{\mathrm{B}})\bar{v}_{\mathrm{B}\to\mathrm{A},0,t}, \quad (71)$$

$$\bar{v}_{\mathrm{B}\to\mathrm{A},1,t+1} = \theta_{\mathrm{B}}\bar{v}_{\mathrm{B},1,t+1}^{\mathrm{ext}} + (1-\theta_{\mathrm{B}})\bar{c}_{\mathrm{B},1,t}, \quad (72)$$

$$\bar{v}_{\mathrm{B}\to\mathrm{A},t'+1,t+1} = \theta_{\mathrm{B}}\bar{c}_{\mathrm{B},t'+1,t+1} + (1-\theta_{\mathrm{B}})\bar{v}_{\mathrm{B}\to\mathrm{A},t',t+1} \quad (73)$$

for all $t',t > 0$, with

$$\bar{c}_{\mathrm{B},t'+1,t+1} = \lim_{M=\delta N\to\infty}\frac{1}{N}(\boldsymbol{x}_{\mathrm{B},t'}^{\mathrm{ext}} - \boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t+1} - \boldsymbol{x})$$
$$\overset{\mathrm{a.s.}}{=} \theta_{\mathrm{B}}\bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{ext}} + (1-\theta_{\mathrm{B}})\bar{c}_{\mathrm{B},t',t}, \quad \bar{c}_{\mathrm{B},t',1} = \bar{v}_{\mathrm{B},t',1}^{\mathrm{ext}}. \quad (74)$$

These results imply that the correct state evolution recursions for damped OAMP are the two-dimensional discrete systems with respect to the covariance messages. Equivalent results were derived for the MF $\boldsymbol{W}_t = \boldsymbol{A}$ and zero-mean i.i.d. Gaussian sensing matrices [42].

The following heuristic damping for the variance messages was used:

$$\frac{1}{v_{\mathrm{A}\to\mathrm{B},t,t}} = \theta_{\mathrm{A}}\left(\frac{1}{v_{\mathrm{A},t,t}^{\mathrm{post}}} - \frac{1}{v_{\mathrm{B}\to\mathrm{A},t,t}}\right) + \frac{1-\theta_{\mathrm{A}}}{v_{\mathrm{A}\to\mathrm{B},t-1,t-1}}, \quad (75)$$

$$\frac{1}{v_{\mathrm{B}\to\mathrm{A},t+1,t+1}} = \theta_{\mathrm{B}}\left(\frac{1}{v_{\mathrm{B},t+1,t+1}^{\mathrm{post}}} - \frac{1}{v_{\mathrm{A}\to\mathrm{B},t,t}}\right) + \frac{1-\theta_{\mathrm{B}}}{v_{\mathrm{B}\to\mathrm{A},t,t}} \quad (76)$$

in precision domain [31] or

$$v_{\mathrm{A}\to\mathrm{B},t,t} = \theta_{\mathrm{A}}\left(\frac{1}{v_{\mathrm{A},t,t}^{\mathrm{post}}} - \frac{1}{v_{\mathrm{B}\to\mathrm{A},t,t}}\right)^{-1}$$
$$+ (1-\theta_{\mathrm{A}})v_{\mathrm{A}\to\mathrm{B},t-1,t-1}, \quad (77)$$

$$v_{\mathrm{B}\to\mathrm{A},t+1,t+1} = \theta_{\mathrm{B}}\left(\frac{1}{v_{\mathrm{B},t+1,t+1}^{\mathrm{post}}} - \frac{1}{v_{\mathrm{B}\to\mathrm{A},t,t}}\right)^{-1}$$
$$+ (1-\theta_{\mathrm{B}})v_{\mathrm{A}\to\mathrm{B},t,t} \quad (78)$$

in variance domain. The heuristic damping cannot provide accurate estimates for the MSE in each module while it results in a one-dimensional discrete system to describe the variance messages in OAMP.

V. OPTIMIZATION

A. Damping

LM-OAMP has several design parameters such as the subsets of indices $\mathcal{T}_{A,t}$, $\mathcal{T}_{B,t}$, the linear filter $\boldsymbol{W}_t$, damping factors $\{\theta_{A,\tau,t}\}$, $\{\theta_{B,\tau,t}\}$, and the denoiser f_t. Since this paper focuses on the Bayes-optimal LM-OAMP, the LMMSE filter (30) and the Bayes-optimal denoiser are used. Thus, the remaining design parameters are $\mathcal{T}_{A,t}$, $\mathcal{T}_{B,t}$, $\{\theta_{A,\tau,t}\}$, and $\{\theta_{B,\tau,t}\}$.

We use Theorem 2 to optimize the damping factors $\{\theta_{A,\tau,t}\}$ and $\{\theta_{B,\tau,t}\}$. A reasonable criterion is the minimization of the variance $\bar{v}^{\mathrm{suf}}_{B\to A,t,t}$ and $\bar{v}^{\mathrm{suf}}_{A\to B,t,t}$ for the sufficient statistics given in (54) and (58). This criterion results in greedy minimization of the MSEs $\bar{v}^{\mathrm{post}}_{A,t,t}$ and $\bar{v}^{\mathrm{post}}_{B,t+1,t+1}$ for the posterior estimators.

When all preceding messages are used, i.e. $\mathcal{T}_{A,t} = \mathcal{T}_{B,t} = \{0,\ldots,t\}$, we have

$$\bar{v}^{\mathrm{suf}}_{B\to A,t,t} = \frac{1}{\mathbf{1}^{\mathrm{T}}\bar{\boldsymbol{V}}^{-1}_{B\to A,t,t}\mathbf{1}}, \quad \bar{v}^{\mathrm{suf}}_{A\to B,t,t} = \frac{1}{\mathbf{1}^{\mathrm{T}}\bar{\boldsymbol{V}}^{-1}_{A\to B,t,t}\mathbf{1}}. \tag{79}$$

In this case, the damping factors can be optimized via the following lemma:

Lemma 4: For any T, let $\boldsymbol{C}_T \in \mathbb{R}^{(T+1)\times(T+1)}$ denote a positive-definite symmetric matrix with $[\boldsymbol{C}_T]_{\tau',\tau} = c_{\tau',\tau}$. Suppose that $\boldsymbol{V}_T \in \mathbb{R}^{(T+1)\times(T+1)}$ has the (t',t) element $v_{t',t}$ for $t',t \in \{0,\ldots,T\}$, given by

$$v_{t',t} = \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} c_{\tau',\tau}\theta_{\tau',t'}\theta_{\tau,t}, \tag{80}$$

where $\{\theta_{\tau,t}\}_{\tau=0}^{t}$ satisfy the normalization condition,

$$\sum_{\tau=0}^{t}\theta_{\tau,t} = 1. \tag{81}$$

Define the upper triangular matrix $\boldsymbol{\Theta}_T \in \mathbb{R}^{(T+1)\times(T+1)}$ as

$$\boldsymbol{\Theta}_T = \begin{bmatrix} \theta_{0,0} & \cdots & \theta_{0,T} \\ & \ddots & \vdots \\ \boldsymbol{O} & & \theta_{T,T} \end{bmatrix}. \tag{82}$$

If $\boldsymbol{\Theta}_T$ has full rank, then the following identity holds:

$$\mathbf{1}^{\mathrm{T}}\boldsymbol{V}_T^{-1}\mathbf{1} = \mathbf{1}^{\mathrm{T}}\boldsymbol{C}_T^{-1}\mathbf{1}. \tag{83}$$

Proof: By definition, $\boldsymbol{V}_T$ is represented as $\boldsymbol{V}_T = \boldsymbol{\Theta}_T^{\mathrm{T}}\boldsymbol{C}_T\boldsymbol{\Theta}_T$. Since both $\boldsymbol{V}_T$ and $\boldsymbol{\Theta}_T$ are invertible, we have $\mathbf{1}^{\mathrm{T}}\boldsymbol{V}_T^{-1}\mathbf{1} = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Theta}_T^{-1}\boldsymbol{C}_T^{-1}(\boldsymbol{\Theta}_T^{-1})^{\mathrm{T}}\mathbf{1}$. The normalization (81) implies $\mathbf{1}^{\mathrm{T}}\boldsymbol{\Theta}_T = \mathbf{1}^{\mathrm{T}}$, so that $\mathbf{1}^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Theta}_T\boldsymbol{\Theta}_T^{-1} = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Theta}_T^{-1}$ holds. Combining these results, we arrive at Lemma 4. ∎

Using Lemma 4 for (57) and (60), we find that the damping factors $\{\theta_{A,\tau,t}\}$ and $\{\theta_{B,\tau,t}\}$ are independent of the variance (79) for the sufficient statistics if $\mathcal{T}_{A,t} = \mathcal{T}_{B,t} = \{0,\ldots,t\}$ are used. This observation is intuitively trivial since $\boldsymbol{x}^{\mathrm{suf}}_{B\to A,t}$ and $\boldsymbol{x}^{\mathrm{suf}}_{A\to B,t}$ in (18) and (36) are sufficient statistics. Thus, we assume no damping $\theta_{A,\tau,t} = \theta_{B,\tau,t} = \delta_{\tau,t}$ when all preceding information $\mathcal{T}_{A,t} = \mathcal{T}_{B,t} = \{0,\ldots,t\}$ are used.

B. Bayes-Optimal LM-OAMP

We consider Bayes-optimal LM-OAMP. In general, the state evolution recursions in Theorem 2 are two-dimensional discrete systems. Interestingly, the systems reduce to one-dimensional systems when all preceding information $\mathcal{T}_{A,t} = \mathcal{T}_{B,t} = \{0,\ldots,t\}$, the LMMSE filter (30), and the Bayes-optimal denoiser are used in each iteration.

The Bayes-optimal denoiser is defined as the denoiser f_τ that minimizes the MSE $\bar{v}^{\mathrm{post}}_{B,\tau+1,\tau+1}$ in (51). By definition, the Bayes-optimal denoiser is equal to the posterior mean estimator $f_\tau = \mathbb{E}[x_1|x_1 + z_\tau]$ with z_τ defined in Theorem 1.

To present the state evolution recursions for the Bayes-optimal LM-OAMP, we review state evolution recursions for conventional Bayes-optimal OAMP [31], [35], which are one-dimensional systems with respect to two variance parameters $\bar{v}_{A\to B,t}$ and $\bar{v}_{B\to A,t}$,

$$\bar{v}_{A\to B,t} = \left(\frac{1}{\bar{\xi}_{A,t}\bar{v}_{B\to A,t}} - \frac{1}{\bar{v}_{B\to A,t}}\right)^{-1}, \tag{84}$$

$$\bar{v}_{B\to A,t+1} = \left(\frac{1}{\bar{v}^{\mathrm{post}}_{B,t+1}} - \frac{1}{\bar{v}_{A\to B,t}}\right)^{-1}, \tag{85}$$

with $\bar{v}_{B\to A,0} = 1$. In (84), $\bar{\xi}_{A,t}$ is given by

$$\bar{\xi}_{A,t} = 1 - \lim_{M=\delta N\to\infty}\frac{\bar{v}_{B\to A,t}}{N}\mathrm{Tr}\left\{\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right.$$
$$\left.\cdot\left(\sigma^2\boldsymbol{I}_M + \bar{v}_{B\to A,t}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\right\}. \tag{86}$$

Furthermore, $\bar{v}^{\mathrm{post}}_{B,t+1} = \mathbb{E}[\{\mathbb{E}[x_1|x_1 + z_t] - x_1\}^2]$ in (85) corresponds to the MSE for Bayes-optimal OAMP in the large system limit, with z_t denoting a zero-mean independent Gaussian random variable with variance $\bar{v}_{A\to B,t}$.

Theorem 3 (Bayes-Optimal LM-OAMP): Suppose that Assumptions 1–5 are satisfied. Consider $\mathcal{T}_{A,t} = \mathcal{T}_{B,t} = \{0,\ldots,t\}$, $\theta_{A,\tau,t} = \theta_{B,\tau,t} = \delta_{\tau,t}$, the LMMSE filter (30), and the Bayes-optimal denoiser. Then,

- The state evolution recursions for the Bayes-optimal LM-OAMP with the initialization $\bar{v}_{B\to A,0,0} = 1$ satisfy $\bar{v}_{A\to B,t',t} = \bar{v}_{A\to B,t}$ and $\bar{v}_{B\to A,t'+1,t+1} = \bar{v}_{B\to A,t+1}$ for all $t' \leq t$, which are defined via the state evolution recursions (84) and (85) for Bayes-optimal OAMP.
- The state $(\bar{v}_{A\to B,t',t}, \bar{v}_{B\to A,t'+1,t+1})$ converges to a FP as t tends to infinity for all $t' \leq t$.

Proof: See Appendix D. ∎

Theorem 3 implies that the Bayes-optimal LM-OAMP converges for general signal prior and asymptotic eigenvalue distributions of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, while conventional MP algorithms require convergence analysis for individual problems [39]–[41].

The state evolution recursions for the Bayes-optimal LM-OAMP are essentially one-dimensional: $\bar{v}_{A\to B,t',t}$ and $\bar{v}_{B\to A,t'+1,t+1}$ are well defined via the one-dimensional recursions (84) and (85) for Bayes-optimal OAMP [31], [35]. As a by-product, we arrive at the convergence of conventional Bayes-optimal OAMP.

Corollary 2 (Bayes-Optimal OAMP): Suppose that Assumptions 1–5 are satisfied. Then, the state evolution

recursions (84) and (85) for conventional Bayes-optimal OAMP converge to a FP. Furthermore, the FP characterizes the Bayes-optimal performance if it is unique.

Proof: Since the Bayes-optimality of the unique FP is due to [30], we only prove the convergence of the state evolution recursions for Bayes-optimal OAMP.

The latter property in Theorem 3 implies the convergence of $\bar{v}_{\mathrm{A}\to\mathrm{B},t',t}$ and $\bar{v}_{\mathrm{B}\to\mathrm{A},t',t}$ for the Bayes-optimal LM-OAMP to a FP. Furthermore, the former property in Theorem 3 claims that the diagonal elements $\bar{v}_{\mathrm{A}\to\mathrm{B},t,t}$ and $\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}$ are respectively equal to $\bar{v}_{\mathrm{A}\to\mathrm{B},t}$ and $\bar{v}_{\mathrm{B}\to\mathrm{A},t}$ in the state evolution recursions (84) and (85) for Bayes-optimal OAMP [31], [35]. These observations imply the convergence of the state evolution recursions for Bayes-optimal OAMP to a FP. ∎

Theorem 3 implies that the dynamics of the MSE performance for the Bayes-optimal LM-OAMP is equal to that for Bayes-optimal OAMP. Interestingly, we can confirm an equivalence between the Bayes-optimal LM-OAMP and Bayes-optimal OAMP.

Proposition 3: Consider $\mathcal{T}_{\mathrm{A},t} = \mathcal{T}_{\mathrm{B},t} = \{0,\ldots,t\}, \theta_{\mathrm{A},\tau,t} = \theta_{\mathrm{B},\tau,t} = \delta_{\tau,t}$, the LMMSE filter (30), and the Bayes-optimal denoiser. If the covariance matrices $\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}$ and $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}$ given via (28) and (45) are positive definite for all t, then the Bayes-optimal LM-OAMP is equivalent to Bayes-optimal OAMP.

Proof: Repeating the proof of the former property in Theorem 3, we can prove the identities $v_{\mathrm{A}\to\mathrm{B},t',t} = v_{\mathrm{A}\to\mathrm{B},t,t}$ and $v_{\mathrm{B}\to\mathrm{A},t'+1,t+1} = v_{\mathrm{B}\to\mathrm{A},t+1,t+1}$ for all $t' \leq t$. From the positive definiteness of $\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},t,t}$ and $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},t,t}$, we can use the last property in Lemma 3 to find that (18) and (36) reduce to $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}^{\mathrm{suf}} = \boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}$ and $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}} = \boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}$, respectively, and that $\xi_{\mathrm{A},t',t} = \xi_{\mathrm{B},t',t} = 0$ in (25) and (43) holds for all $t' \neq t$.

The former observation implies that one can skip the computation of the sufficient statistics in the first step of the Bayes-optimal LM-OAMP. The latter observation indicates that the extrinsic mean messages in (23) and (41) are equivalent to those in Bayes-optimal OAMP [30]. Thus, the Bayes-optimal LM-OAMP is equivalent to Bayes-optimal OAMP. ∎

Theorem 1 justifies the positive-definiteness assumption in Proposition 3 in the large system limit. Intuitively, Proposition 3 holds since $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}$ and $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}$ are sufficient statistics for estimation of the signal vector given all preceding messages in each module. The Bayes-optimal LM-OAMP may be regarded as a theoretical tool to prove that $\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}$ and $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}$ are sufficient statistics. These observations imply that conventional Bayes-optimal OAMP is the best option among all possible LM-MP algorithms in terms of the reconstruction performance.

VI. Numerical Results

A. Simulation Conditions

In all numerical results, we assume an i.i.d. Bernoulli-Gaussian (BG) signal vector $\boldsymbol{x}$ and artificial sensing matrices $\boldsymbol{A}$. Each signal element x_n is independently sampled from the Gaussian distribution $\mathcal{N}(0, \rho^{-1})$ with probability $\rho \in [0,1]$.

Otherwise, x_n takes zero. See [44, Appendix F] for properties of the BG prior.

In artificial sensing matrices with condition number $\kappa > 1$, $\boldsymbol{A}$ is assumed to have non-zero singular values $\sigma_0 \geq \cdots \geq \sigma_{M-1} > 0$ satisfying the condition number $\kappa = \sigma_0/\sigma_{M-1}$, $\sigma_m/\sigma_{m-1} = \kappa^{-1/(M-1)}$, and $\sigma_0^2 = N(1 - \kappa^{-2/(M-1)})/(1 - \kappa^{-2M/(M-1)})$. As shown in [44, Eq. (65)], the η-transform η in (31) converges almost surely to

$$\eta(x) \overset{\text{a.s.}}{\to} 1 - \frac{1}{C}\ln\left(\frac{\kappa^2 - 1 + \kappa^2 Cx}{\kappa^2 - 1 + Cx}\right) \tag{87}$$

in the large system limit, with $C = 2\delta^{-1}\ln\kappa$. The η-transform (87) is used in solving state evolution recursions.

To reduce the computational complexity, we assume the SVD structure $\boldsymbol{A} = \boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$, in which $\boldsymbol{V}$ denotes a Hadamard matrix with random row permutation. Such orthogonal matrices $\boldsymbol{V}$ may be regarded as a practical alternative of Haar orthogonal matrices [58]–[60].

B. Implementation

We consider the LMMSE filter (30) in module A and the Bayes-optimal denoiser $f_t(s_t) = \mathbb{E}[x_1|s_t]$ with $s_t = x_1 + z_t$ in module B, in which $\{z_\tau\}$ are defined in Theorem 1. In computing the covariance message $v_{\mathrm{B},t'+1,t+1}^{\mathrm{post}}$, we use the posterior covariance $v_{\mathrm{B},t'+1,t+1}^{\mathrm{post}} = \langle C(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t'}^{\mathrm{suf}}, \boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}^{\mathrm{suf}})\rangle$ with $C(s_{t'}, s_t) = \mathbb{E}[\{x_1 - f_{t'}(s_{t'})\}\{x_1 - f_t(s_t)\}|s_{t'}, s_t]$, instead of (39). See [44, Appendix F] for the details.

Let $\boldsymbol{V}_{\mathrm{B},t,t}^{\mathrm{ext}} \in \mathbb{R}^{|\mathcal{T}_{\mathrm{B},t}| \times |\mathcal{T}_{\mathrm{B},t}|}$ denote the covariance matrix with $[\boldsymbol{V}_{\mathrm{B},t,t}^{\mathrm{ext}}]_{\tau',\tau} = v_{\mathrm{B},\tau',\tau}^{\mathrm{ext}}$ given in (42). The matrix $\boldsymbol{V}_{\mathrm{B},t,t}^{\mathrm{ext}}$ is guaranteed to be positive definite in the large system limit. For finite-sized systems, however, it might not be positive definite for some t. To circumvent this issue, it is recommended to replace $v_{\mathrm{B},\tau',\tau}^{\mathrm{ext}}$ and $v_{\mathrm{B},\tau,\tau'}^{\mathrm{ext}}$ with $v_{\mathrm{B},\tau,\tau}^{\mathrm{ext}}$ if $v_{\mathrm{B},\tau',\tau'}^{\mathrm{ext}} v_{\mathrm{B},\tau,\tau}^{\mathrm{ext}} - (v_{\mathrm{B},\tau',\tau}^{\mathrm{ext}})^2 < \epsilon$ holds for small $\epsilon > 0$, which was set to $\epsilon = 10^{-6}$ in numerical simulations.

As discussed in Section II, this replacement implies that the AWGN observation $s_\tau = x_1 + z_\tau$ in Theorem 1 becomes a sufficient statistic for estimation of x_1 based on both $s_{\tau'}$ and s_τ for $\tau' < \tau$. In other words, it means that $s_{\tau'}$ is not used since it is correlated with s_τ strongly.

C. Numerical Simulations

Damped OAMP with the correct state evolution (67) and (73)—called LM-OAMP—is compared to damped OAMP with the heuristic damping (77) and (78) [31]—called OAMP. See Example 2 for the details.

We first verify the correctness of the state evolution results via numerical simulations for large systems $N = 2^{13}$. Figure 2 shows the MSEs versus the number of iterations for LM-OAMP with damping only in module B, i.e. $\theta_{\mathrm{A}} = 1$ and $\theta_{\mathrm{B}} \in (0,1)$. The state evolution results predict the MSEs of LM-OAMP accurately while OAMP is not in agreement with the state evolution results for small θ_{B}.

We next compare LM-OAMP with OAMP for smaller systems than in Fig. 2. As shown in Fig. 3, LM-OAMP is slightly inferior to OAMP especially for small systems. This is
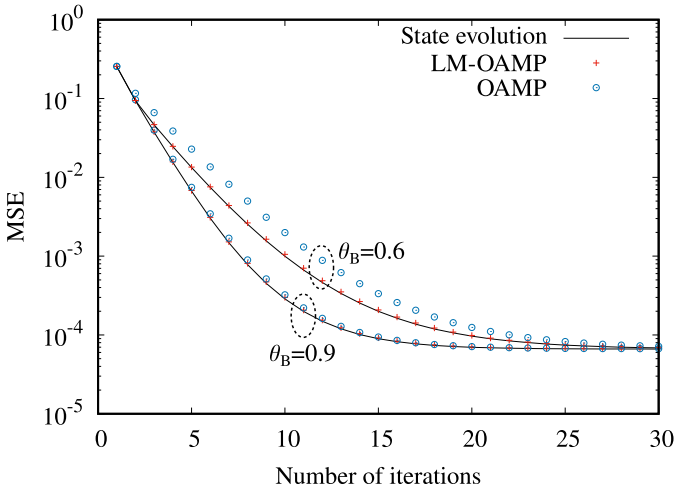
Fig. 2. MSE versus the number of iterations for $M = 2^{12}$, $N = 2^{13}$, BG signals with signal density $\rho = 0.1$, condition number $\kappa = 10^3$, $\theta_A = 1$, and SNR $1/\sigma^2 = 40$ dB.
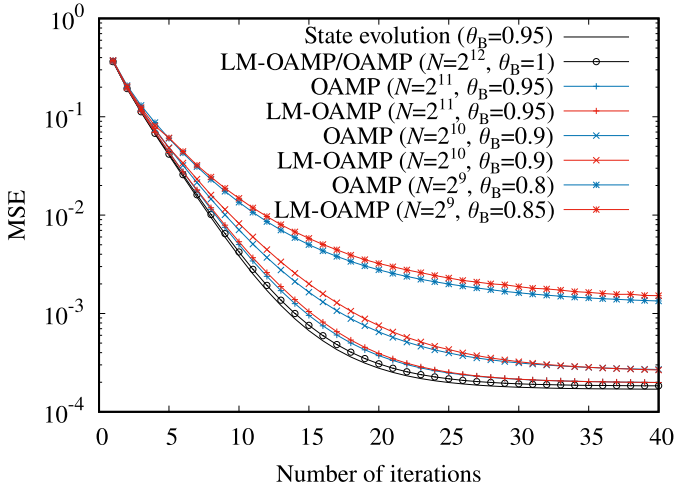

Fig. 3. MSE versus the number of iterations for compression ratio $\delta = 0.5$, BG signals with signal density $\rho = 0.1$, condition number $\kappa = 10^4$, $\theta_A = 1$, and SNR $1/\sigma^2 = 40$ dB.

surprising since LM-OAMP uses the consistent state evolution in the large system limit.

This result is intuitively understood as follows: LM-OAMP uses the covariance messages in all preceding iterations to compute the state evolution prediction in the current iteration. As a result, all prediction errors in the preceding iterations are accumulated for finite-sized systems, so that LM-OAMP should have larger prediction errors than conventional OAMP without LM processing. Figure 3 shows that this prediction errors due to finite size effects are more dominant than the inconsistency of the heuristic damping in the large system limit.

VII. CONCLUSION

This paper has proposed LM-MP, constructed by adding computation of sufficient statistics for estimation of the signal vector given all preceding messages to the beginning of each module in conventional MP. The proposed LM-MP is a modification of conventional MP in two points.

In one point, the Bayes-optimal LM-MP provides a proof strategy to guarantee the convergence of state evolution recursions for conventional Bayes-optimal MP systematically. In the proof strategy, one-dimensional state evolution recursions for the Bayes-optimal MP are embedded into two-dimensional state evolution recursions for the Bayes-optimal LM-MP, which are systematically guaranteed to converge. As a result, one can prove the convergence of the state evolution recursions for the Bayes-optimal MP. While it has been applied to Bayes-optimal OAMP [30], [31] in this paper, the proposed LM-MP strategy is general and applicable to the convergence analysis of the other Bayes-optimal MP algorithms.

The other point is that LM-MP may improve the convergence properties of conventional MP. If state evolution recursions for MP with Bayes-optimal denoiser are not embedded into those for the corresponding LM-MP, the LM-MP strategy indicates that the conventional MP has room for improvement: The constructed LM-MP improves the convergence properties of the conventional MP for large systems. In this sense, MAMP [43] can be regarded as a modification of an MP algorithm without LM damping.

Possible directions for future work are summarized to conclude this paper. A first issue is another criterion for optimality such as the maximum a posteriori (MAP) scenario while this paper has focused on the minimum MSE (MMSE) scenario. It is an interesting direction whether the LM-MP strategy can be generalized to the other criteria for optimality.

The other issue is an improvement of the state evolution prediction for LM-MP in finite-sized systems. State evolution recursions for LM-MP are more sensitive to finite size effects than for conventional MP while they are consistent in the large system limit. It is important to construct covariance estimators that are consistent in the large system limit and accurate for finite-sized systems. Such covariance estimators should increase practical value of LM-MP for finite-sized systems while the numerical simulations in this paper have shown the inferiority of damped LM-OAMP to OAMP with heuristic damping for finite-sized systems.

APPENDIX A
PROOFS

A. Proof of Proposition 2

We follow [30, Appendix B] to prove Proposition 2. We use the definition $f_{\mathrm{opt}}(S_t) = \mathbb{E}[X|S_t]$ to obtain

$$\mathbb{E}[(\psi_t - X)^2] = \mathbb{E}[(\psi_t - f_{\mathrm{opt}} + f_{\mathrm{opt}} - X)^2]$$
$$= \mathbb{E}[(\psi_t - f_{\mathrm{opt}})^2] + \mathbb{E}[(f_{\mathrm{opt}} - X)^2]. \quad (88)$$

Thus, the minimization of $\mathbb{E}[(\psi_t - X)^2]$ is equivalent to that of $\mathbb{E}[(\psi_t - f_{\mathrm{opt}})^2]$.

To evaluate $\mathbb{E}[(\psi_t - f_{\mathrm{opt}})^2]$, we repeat the same argument for the extrinsic denoiser f_t^{ext} to obtain

$$\mathbb{E}[(\psi_t - f_{\mathrm{opt}})^2] = \mathbb{E}[(\psi_t - f_t^{\mathrm{ext}} + f_t^{\mathrm{ext}} - f_{\mathrm{opt}})^2]$$
$$= \mathbb{E}[(\psi_t - f_t^{\mathrm{ext}})^2] + 2\mathbb{E}[(\psi_t - f_t^{\mathrm{ext}})(f_t^{\mathrm{ext}} - f_{\mathrm{opt}})]$$
$$+ \mathbb{E}[(f_t^{\mathrm{ext}} - f_{\mathrm{opt}})^2], \quad (89)$$

which implies that $\psi_t = f_t^{\text{ext}}$ minimizes $\mathbb{E}[(\psi_t - f_{\text{opt}})^2]$ if the second term is equal to zero for all extrinsic denoisers ψ_t.

To complete the proof of Proposition 2, it is sufficient to prove $\mathbb{E}[\psi_t(f_t^{\text{ext}} - f_{\text{opt}})] = 0$ for any Lipschitz-continuous extrinsic denoiser ψ_t with $\mathbb{E}[\partial \psi_t / \partial Y_\tau] = 0$. From the definition of f_t^{ext} in (12), we have

$$f_t^{\text{ext}} - f_{\text{opt}} = \frac{\sum_{\tau \in \mathcal{T}_t} \xi_{\tau,t}(f_{\text{opt}} - X - W_\tau)}{1 - \sum_{\tau \in \mathcal{T}_t} \xi_{\tau,t}}. \quad (90)$$

Since $\{W_\tau : \tau \in \mathcal{T}_t\}$ are zero-mean Gaussian, we can use [44, Lemma 2] for Lipschitz-continuous ψ_t to obtain

$$\mathbb{E}[W_\tau \psi_t] = \sum_{t' \in \mathcal{T}_t} \mathbb{E}[W_\tau W_{t'}] \mathbb{E}\left[\frac{\partial \psi_t}{\partial Y_{t'}}\right] = 0, \quad (91)$$

where the last follows from the assumption $\mathbb{E}[\partial \psi_t / \partial Y_{t'}] = 0$. Thus, we have

$$\mathbb{E}[\psi_t(f_t^{\text{ext}} - f_{\text{opt}})] = \frac{\sum_{\tau \in \mathcal{T}_t} \xi_{\tau,t} \mathbb{E}[\psi_t(f_{\text{opt}} - X)]}{1 - \sum_{\tau \in \mathcal{T}_t} \xi_{\tau,t}} = 0, \quad (92)$$

where the last equality follows from $\mathbb{E}[\psi_t X] = \mathbb{E}[\psi_t f_{\text{opt}}]$ due to the definition $f_{\text{opt}}(S_t) = \mathbb{E}[X|S_t] = \mathbb{E}[X|\boldsymbol{Y}_t]$. Thus, Proposition 2 holds.

B. Proof of Lemma 1

Since $\{S_{t',n}, S_{t,n}\}_{n=1}^N$ are independent samples, we use the strong law of large numbers to obtain

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^N \hat{C}(S_{t',n}, S_{t,n}) \stackrel{\text{a.s.}}{=} \mathbb{E}[\hat{C}(S_{t'}, S_t)]. \quad (93)$$

Thus, it is sufficient to prove the unbiasedness of $\hat{C}$, i.e.

$$\mathbb{E}[\hat{C}(S_{t'}, S_t)] = \mathbb{E}[\{X - f_{t'}(S_{t'})\}\{X - f_t(S_t)\}]. \quad (94)$$

We first evaluate the expectation $\mathbb{E}[S_{t'} f_t(S_t)]$. Substituting $S_{t'} = X + \tilde{W}_{t'}$ and using Stein's lemma [61] for Lipschitz-continuous f_t, we have

$$\mathbb{E}[S_{t'} f_t(S_t)] = \mathbb{E}[X f_t(S_t)] + \mathbb{E}[\tilde{W}_{t'} \tilde{W}_t] \mathbb{E}[f_t'(S_t)]. \quad (95)$$

Applying this result to (16), we arrive at

$$\begin{aligned} \mathbb{E}[\hat{C}(S_{t'}, S_t)] &= \mathbb{E}[X^2] + \mathbb{E}[f_{t'}(S_{t'}) f_t(S_t)] \\ &\quad - \mathbb{E}[X f_t(S_t)] - \mathbb{E}[X f_{t'}(S_{t'})] \\ &= \mathbb{E}[\{X - f_{t'}(S_{t'})\}\{X - f_t(S_t)\}]. \end{aligned} \quad (96)$$

Thus, Lemma 1 holds.

C. Proof of Lemma 2

We prove the former property. Since $\{Y_\tau\}_{\tau=0}^t$ contains $\{Y_\tau\}_{\tau=0}^{t'}$ for $t' < t$, the optimality of the posterior mean estimator $f_{\text{opt}}(S_t)$ implies $\mathbb{E}[\{X - f_{\text{opt}}(S_t)\}^2] \leq \mathbb{E}[\{X - f_{\text{opt}}(S_{t'})\}^2]$. The monotonicity $\mathbb{E}[\tilde{W}_{t'}^2] \geq \mathbb{E}[\tilde{W}_t^2]$ follows from the monotonicity of the MSE $\mathbb{E}[\{X - f_{\text{opt}}(S_t)\}^2]$ with respect to the variance $\mathbb{E}[\tilde{W}_t^2]$.

We next prove the latter property. When $\mathbb{E}[\tilde{W}_{t'} \tilde{W}_t] = \mathbb{E}[\tilde{W}_t^2]$ holds, we have the following representation for the sufficient statistic:

$$S_{t'} = S_t + Z_{t'}, \quad S_t = X + \tilde{W}_t, \quad (97)$$

where $Z_{t'}$ is a zero-mean Gaussian random variable with variance $\mathbb{E}[\tilde{W}_{t'}^2] - \mathbb{E}[\tilde{W}_t^2] \geq 0$ and independent of X and $\tilde{W}_t$. The representation (97) can be verified from

$$\mathbb{E}[(\tilde{W}_t + Z_{t'})^2] = \mathbb{E}[\tilde{W}_{t'}^2], \quad \mathbb{E}[(\tilde{W}_t + Z_{t'})\tilde{W}_t] = \mathbb{E}[\tilde{W}_t^2]. \quad (98)$$

The representation (97) implies that S_t is a sufficient statistic for estimation of X given both $S_{t'}$ and S_t. Thus, we have $\mathbb{E}[X|S_{t'}, S_t] = \mathbb{E}[X|S_t] = f_{\text{opt}}(S_t)$. Applying this identity to the definition of $C(S_{t'}, S_t)$ in (14), we arrive at

$$\begin{aligned} &C(S_{t'}, S_t) - C(S_t, S_t) \\ &= \{f_{\text{opt}}(S_t) - f_{\text{opt}}(S_{t'})\}\{\mathbb{E}[X|S_{t'}, S_t] - f_{\text{opt}}(S_t)\} \\ &= 0. \end{aligned} \quad (99)$$

Thus, the latter property holds.

D. Proof of Lemma 3

We first prove the first property. It is straightforward to confirm

$$\det \boldsymbol{\Sigma}_t = [\boldsymbol{\Sigma}_t]_{t,t} \prod_{\tau=0}^{t-1} \Delta\Sigma_{\tau,t}, \quad (100)$$

with $\Delta\Sigma_{\tau,t} = [\boldsymbol{\Sigma}_t]_{\tau,\tau} - [\boldsymbol{\Sigma}_t]_{\tau+1,\tau+1}$. Since $\{\boldsymbol{\Sigma}_\tau\}_{\tau=1}^t$ have been assumed to be positive definite, we have the positivity $\{\det \boldsymbol{\Sigma}_\tau > 0\}_{\tau=1}^t$. Using (100) yields the first property $\Delta\Sigma_{\tau,t} > 0$ for all τ.

Let us prove (100). Since $[\boldsymbol{\Sigma}_t]_{\tau',\tau} = [\boldsymbol{\Sigma}_t]_{\tau,\tau'} = [\boldsymbol{\Sigma}_t]_{\tau,\tau}$ holds for all $\tau' < \tau$, we subtract the $(\tau+1)$th column in $\boldsymbol{\Sigma}_t$ from the τth column for all $\tau \in \{0, \ldots, t-1\}$ to obtain the upper triangular expression for the determinant,

$$\det \boldsymbol{\Sigma}_t = \begin{vmatrix} \Delta\Sigma_{0,t} & \cdots & \Delta\Sigma_{t-1,t} & [\boldsymbol{\Sigma}_t]_{t,t} \\ \boldsymbol{O} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \Delta\Sigma_{t-1,t} & \vdots \\ \boldsymbol{O} & \cdots & \boldsymbol{O} & [\boldsymbol{\Sigma}_t]_{t,t} \end{vmatrix}, \quad (101)$$

which implies (100). Thus, the first property holds.

We next prove the second property. Using the assumption $[\boldsymbol{\Sigma}_t]_{\tau',\tau} = [\boldsymbol{\Sigma}_t]_{\tau,\tau}$ and the first property $[\boldsymbol{\Sigma}_t]_{\tau',\tau'} > [\boldsymbol{\Sigma}_t]_{\tau,\tau}$ for all $\tau' < \tau$, we have the following representation for $\{Y_\tau\}$:

$$Y_t = X + W_t, \quad Y_{\tau-1} = Y_\tau + V_{\tau-1} \quad (102)$$

for $\tau \in \{1, \ldots, t\}$, where $\{V_{\tau-1}\}$ are independent zero-mean Gaussian random variables with $\mathbb{E}[V_{\tau-1}^2] = \Delta\Sigma_{\tau-1,t} > 0$. This representation implies that Y_t is a sufficient statistic for estimation of X based on $\{Y_\tau\}_{\tau=0}^t$. Thus, we arrive at $\mathbb{E}[\{X - f_{\text{opt}}(S_t)\}^2] = \mathbb{E}[\{X - f_{\text{opt}}(Y_t)\}^2]$, which implies $\mathbb{E}[\tilde{W}_t^2] = \mathbb{E}[W_t^2]$.

Finally, the last property follows immediately from the definitions of S_t and $\xi_{\tau,t}$ in (5) and (13), and the identity $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{1} = ([\boldsymbol{\Sigma}_t]_{t,t})^{-1}\boldsymbol{e}_t$. The identity $S_t = Y_t$ provides an alternative proof of the second property: $\mathbb{E}[\tilde{W}_t^2] = \mathbb{E}[(S_t - X)^2] = \mathbb{E}[(Y_t - X)^2] = \mathbb{E}[W_t^2]$.

The identity $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{1} = ([\boldsymbol{\Sigma}_t]_{t,t})^{-1}\boldsymbol{e}_t$ can be confirmed as follows: From the assumption in Lemma 3, $\boldsymbol{\Sigma}_t$ has the last column $[\boldsymbol{\Sigma}_t]_{t,t}\boldsymbol{1}$. Thus, we have $\boldsymbol{\Sigma}_t\boldsymbol{e}_t = [\boldsymbol{\Sigma}_t]_{t,t}\boldsymbol{1}$, which is equivalent to the identity $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{1} = ([\boldsymbol{\Sigma}_t]_{t,t})^{-1}\boldsymbol{e}_t$.

APPENDIX B
PROOF OF THEOREM 1

Asymptotic Gaussianity has been proved for a general error model proposed in [44]. Thus, we first prove that the error model for LM-OAMP is included into the general error model.

Lemma 5: Suppose that Assumption 4 holds. Let $h_t = x_{A \to B,t} - x$ denote the estimation error before denoising in iteration t. Define an error vector q_{t+1} associated with the estimation error after denoising by

$$q_{t+1} = \sum_{\tau=0}^{t} \frac{\theta_{B,\tau,t}}{1 - \xi_{B,\tau}} (x_{B,\tau+1}^{\mathrm{post}} - x). \qquad (103)$$

Then, the error model for LM-OAMP is included into the general error model in [44] and given by

$$b_t = V^{\mathrm{T}} \tilde{q}_t, \qquad (104)$$

$$m_t = \sum_{\tau=0}^{t} \frac{\theta_{A,\tau,t}}{1 - \xi_{A,\tau}} V^{\mathrm{T}} (x_{A,\tau}^{\mathrm{post}} - x), \qquad (105)$$

$$\tilde{m}_t = \sum_{\tau=0}^{t} \frac{\theta_{A,\tau,t} \{ V^{\mathrm{T}} (x_{A,\tau}^{\mathrm{post}} - x) - \sum_{t' \in \mathcal{T}_{A,\tau}} \xi_{A,t',\tau} b_{t'} \}}{1 - \xi_{A,\tau}}, \qquad (106)$$

$$V^{\mathrm{T}} (x_{A,t}^{\mathrm{post}} - x) = \left(I_N - \tilde{W}_t^{\mathrm{T}} \Sigma \right) \frac{B_t V_{B \to A,t,t}^{-1} 1}{1^{\mathrm{T}} V_{B \to A,t,t}^{-1} 1}$$
$$+ \tilde{W}_t^{\mathrm{T}} U^{\mathrm{T}} w, \qquad (107)$$

$$h_t = V \tilde{m}_t, \qquad (108)$$

$$\tilde{q}_{t+1} = \sum_{\tau=0}^{t} \frac{\theta_{B,\tau,t} (x_{B,\tau+1}^{\mathrm{post}} - x - \sum_{t' \in \mathcal{T}_{B,\tau}} \xi_{B,t',\tau} h_{t'})}{1 - \xi_{B,\tau}}, \qquad (109)$$

with $\tilde{q}_0 = -x$, where $B_t \in \mathbb{R}^{N \times |\mathcal{T}_{A,t}|}$ consists of $\{ b_\tau : \tau \in \mathcal{T}_{A,t} \}$. In particular, $\tilde{q}_t = x_{B \to A,t} - x$ holds.

Proof: We first prove

$$rl\tilde{q}_{t+1} = q_{t+1} - \sum_{t'=0}^{t} \left(\frac{1}{N} \sum_{n=1}^{N} \frac{\partial q_{n,t+1}}{\partial h_{n,t'}} \right) h_{t'}, \qquad (110)$$

which is equivalent to the definition in [44, Eq. (9)]. From the definition of q_{t+1} in (103), we have

$$\frac{1}{N} \sum_{n=1}^{N} \frac{\partial q_{n,t+1}}{\partial h_{n,t'}} = \frac{1}{N} \sum_{n=1}^{N} \sum_{\tau=0}^{t} \frac{\theta_{B,\tau,t}}{1 - \xi_{B,\tau}} \frac{\partial x_{B,n,\tau+1}^{\mathrm{post}}}{\partial h_{n,t'}}. \qquad (111)$$

We define the error matrix $H_t = X_{A \to B,t} - 1^{\mathrm{T}} \otimes x$ for the preceding messages before denoising, so that H_t consists of the columns $\{ h_{t'} : t' \in \mathcal{T}_{B,t} \}$. Using the definition of $x_{B,t+1}^{\mathrm{post}}$ in (38) with (36), we find that $x_{B,t+1}^{\mathrm{post}}$ is a function of $\{ h_{t'} : t' \in \mathcal{T}_{B,t} \}$ since the sufficient statistic (36) is a function

of $X_{A \to B,t} = 1^{\mathrm{T}} \otimes x + H_t$. Thus, we obtain

$$\frac{1}{N} \sum_{n=1}^{N} \frac{\partial x_{B,n,\tau+1}^{\mathrm{post}}}{\partial h_{n,t'}} = \frac{1}{N} \sum_{n=1}^{N} f_\tau' (x_{A \to B,n,\tau}^{\mathrm{suf}})$$
$$\cdot \frac{\partial}{\partial h_{n,t'}} \frac{([1^{\mathrm{T}} \otimes x + H_\tau]_{,n})^{\mathrm{T}} V_{B \to A,\tau,\tau}^{-1} 1}{1^{\mathrm{T}} V_{B \to A,\tau,\tau}^{-1} 1}$$
$$= \langle f_\tau' (x_{A \to B,\tau}^{\mathrm{suf}}) \rangle \frac{e_{I_{\mathcal{T}_{B,\tau}}(t')}^{\mathrm{T}} V_{B \to A,\tau,\tau}^{-1} 1}{1^{\mathrm{T}} V_{B \to A,\tau,\tau}^{-1} 1} = \xi_{B,t',\tau} \qquad (112)$$

for $t' \in \mathcal{T}_{B,\tau}$, with $\xi_{B,t',\tau}$ defined in (43), where $[\cdots]_{,n}$ denotes the nth row of $\cdots$. Substituting these results and the definition of q_{t+1} in (103) into (110), we arrive at the equivalence between (109) and (110).

We next prove $\tilde{q}_{t+1} = x_{B \to A,t+1} - x$. Using the definition of $x_{B,t}^{\mathrm{ext}}$ in (41) and $h_t = x_{A \to B,t} - x$ yields

$$x_{B,t+1}^{\mathrm{ext}} - x = \frac{x_{B,t+1}^{\mathrm{post}} - x - \sum_{t' \in \mathcal{T}_{B,t}} \xi_{B,t',t} h_{t'}}{1 - \xi_{B,t}}, \qquad (113)$$

where we have used $\sum_{t' \in \mathcal{T}_{B,t}} \xi_{B,t',t} = \xi_{B,t}$. Applying this expression to (109), we have

$$\tilde{q}_{t+1} = \sum_{\tau=0}^{t} \theta_{B,\tau,t} (x_{B,\tau+1}^{\mathrm{ext}} - x). \qquad (114)$$

Using the definition of $x_{B \to A,t+1}$ in (44) and the normalization (46), we arrive at $\tilde{q}_{t+1} = x_{B \to A,t+1} - x$.

We shall derive (107). Applying $\tilde{q}_t = x_{B \to A,t} - x$ to the sufficient statistic (18) yields

$$x_{B \to A,t}^{\mathrm{suf}} = x + \frac{\tilde{Q}_t V_{B \to A,t,t}^{-1} 1}{1^{\mathrm{T}} V_{B \to A,t,t}^{-1} 1}, \qquad (115)$$

with $\tilde{Q}_t = X_{B \to A,t} - (1^{\mathrm{T}} \otimes x)$. Substituting this result and the measurement model (1) into the definition of $x_{A,t}^{\mathrm{post}}$ in (20), we have

$$x_{A,t}^{\mathrm{post}} - x = \left(I_N - W_t^{\mathrm{T}} A \right) \frac{\tilde{Q}_t V_{B \to A,t,t}^{-1} 1}{1^{\mathrm{T}} V_{B \to A,t,t}^{-1} 1} + W_t^{\mathrm{T}} w. \qquad (116)$$

For the SVD $A = U \Sigma V^{\mathrm{T}}$, applying the SVD $W_t = U \tilde{W}_t V^{\mathrm{T}}$ in Assumption 4 to (116) and using the definition of b_t in (104), we arrive at (107) with $B_t = V^{\mathrm{T}} \tilde{Q}_t$.

Let us prove

$$\tilde{m}_t = m_t - \sum_{t'=0}^{t} \left(\frac{1}{N} \sum_{n=1}^{N} \frac{\partial m_{n,t}}{\partial b_{n,t'}} \right) b_{t'}, \qquad (117)$$

which is equivalent to the definition in [44, Eq. (11)]. Using the expression for m_t in (105) with (107), we repeat the derivation of (112) to have

$$\frac{1}{N} \sum_{n=1}^{N} \frac{\partial m_{n,t}}{\partial b_{n,t'}} = \sum_{\tau=0}^{t} \frac{\theta_{A,\tau,t} \xi_{A,t',\tau}}{1 - \xi_{A,\tau}} \qquad (118)$$

for $t' \in \mathcal{T}_{A,t}$, with $\xi_{A,t',t}$ defined in (25), where we have used the identity $\mathrm{Tr}(I_N - \tilde{W}_t^{\mathrm{T}} \Sigma) = \mathrm{Tr}(I_N - W_t^{\mathrm{T}} A)$. Substituting this expression and the definition of m_t in (105) into (117), we arrive at the equivalence between (106) and (117).

We have so far proved that the error model (104)–(109) is included into the general error model in [44]. To complete the proof of Lemma 5, we prove $\boldsymbol{h}_t = \boldsymbol{V}\tilde{\boldsymbol{m}}_t$. From the definition of $\tilde{\boldsymbol{m}}_t$ in (106), we have

$$\boldsymbol{V}\tilde{\boldsymbol{m}}_t = \sum_{\tau=0}^{t} \theta_{A,\tau,t} \frac{\boldsymbol{x}_{A,\tau}^{\text{post}} - \boldsymbol{x} - \sum_{t' \in \mathcal{T}_{A,\tau}} \xi_{A,t',\tau}\tilde{\boldsymbol{q}}_{t'}}{1 - \xi_{A,\tau}}, \quad (119)$$

with $\tilde{\boldsymbol{q}}_t = \boldsymbol{V}\boldsymbol{b}_t$. Using the definition of $\boldsymbol{x}_{A,t}^{\text{ext}}$ in (23) yields

$$\boldsymbol{x}_{A,t}^{\text{ext}} - \boldsymbol{x} = \frac{\boldsymbol{x}_{A,t}^{\text{post}} - \boldsymbol{x} - \sum_{t' \in \mathcal{T}_{A,t}} \xi_{A,t',t}\tilde{\boldsymbol{q}}_{t'}}{1 - \xi_{A,t}}, \quad (120)$$

with $\tilde{\boldsymbol{q}}_t = \boldsymbol{x}_{B \to A,t} - \boldsymbol{x}$, where we have used the identity $\sum_{t' \in \mathcal{T}_{A,t}} \xi_{A,t',t} = \xi_{A,t}$. Thus, we have

$$\boldsymbol{V}\tilde{\boldsymbol{m}}_t = \sum_{\tau=0}^{t} \theta_{A,\tau,t}(\boldsymbol{x}_{A,\tau}^{\text{ext}} - \boldsymbol{x}) = \boldsymbol{x}_{A \to B,t} - \boldsymbol{x} = \boldsymbol{h}_t, \quad (121)$$

where the second equality follows from the definition of $\boldsymbol{x}_{A \to B,t}$ in (27) and the normalization (29). Thus, Lemma 5 holds. ∎

Lemma 5 and Assumptions 1–5 allow us to use [44, Theorem 6] to prove Theorem 1. The first property in Theorem 1 follows from [44, Theorem 6 (B4) and Eq. (107)]. Similarly, the third property follows from [44, Theorem 6 (A4) and Eq. (98)].

We next prove the second property. Using (107) and [44, Eq. (94)] yields

$$\frac{1}{N}(\boldsymbol{x}_{A,\tau'}^{\text{post}} - \boldsymbol{x})^{\text{T}}(\boldsymbol{x}_{A,\tau}^{\text{post}} - \boldsymbol{x})$$

$$\stackrel{\text{a.s.}}{=} \frac{1}{N} \frac{\boldsymbol{1}^{\text{T}}\boldsymbol{V}_{B \to A,\tau',\tau'}^{-1}}{\boldsymbol{1}^{\text{T}}\boldsymbol{V}_{B \to A,\tau',\tau'}^{-1}\boldsymbol{1}} \mathbb{E}\left[\boldsymbol{B}_{\tau'}^{\text{T}}\boldsymbol{D}_{\tau',\tau}\boldsymbol{B}_\tau\right] \frac{\boldsymbol{V}_{B \to A,\tau,\tau}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\text{T}}\boldsymbol{V}_{B \to A,\tau,\tau}^{-1}\boldsymbol{1}}$$

$$+ N^{-1}\mathbb{E}\left[\boldsymbol{w}^{\text{T}}\boldsymbol{U}\tilde{\boldsymbol{W}}_{\tau'}\tilde{\boldsymbol{W}}_\tau^{\text{T}}\boldsymbol{U}^{\text{T}}\boldsymbol{w}\right] + o(1) \quad (122)$$

in the large system limit, with

$$\boldsymbol{D}_{\tau',\tau} = \left(\boldsymbol{I}_N - \tilde{\boldsymbol{W}}_{\tau'}^{\text{T}}\boldsymbol{\Sigma}\right)^{\text{T}} \left(\boldsymbol{I}_N - \tilde{\boldsymbol{W}}_\tau^{\text{T}}\boldsymbol{\Sigma}\right), \quad (123)$$

where $\{\boldsymbol{B}_\tau\}$ follow zero-mean Gaussian random matrices with covariance $\mathbb{E}[\boldsymbol{b}_\tau\boldsymbol{b}_{\tau'}^{\text{T}}] = \bar{v}_{B \to A,\tau',\tau}\boldsymbol{I}_N$ while $\boldsymbol{U}^{\text{T}}\boldsymbol{w}$ follows $\mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_M)$. Evaluating the expectation in (122) with the identities $\text{Tr}(\boldsymbol{D}_{\tau',\tau}) = \text{Tr}\{(\boldsymbol{I}_N - \boldsymbol{W}_{\tau'}^{\text{T}}\boldsymbol{A})^{\text{T}}(\boldsymbol{I}_N - \boldsymbol{W}_\tau^{\text{T}}\boldsymbol{A})\}$ and $\text{Tr}(\tilde{\boldsymbol{W}}_{\tau'}\tilde{\boldsymbol{W}}_\tau^{\text{T}}) = \text{Tr}(\boldsymbol{W}_{\tau'}\boldsymbol{W}_\tau^{\text{T}})$, we find that the right-hand side (RHS) is equal to (49) in the large system limit. Thus, the second property in Theorem 1 holds.

Finally, we prove the last property. Repeating the derivation of (115) for the sufficient statistic (36) yields

$$\boldsymbol{x}_{A \to B,t}^{\text{suf}} = \boldsymbol{x} + \boldsymbol{z}_t, \quad \boldsymbol{z}_t = \frac{\boldsymbol{H}_t\boldsymbol{V}_{A \to B,t,t}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\text{T}}\boldsymbol{V}_{A \to B,t,t}^{-1}\boldsymbol{1}}. \quad (124)$$

We use the definition of $\boldsymbol{x}_{B,\tau+1}$ in (38) and [44, Eq. (103)] to obtain

$$\frac{1}{N}(\boldsymbol{x}_{B,\tau'+1}^{\text{post}} - \boldsymbol{x})^{\text{T}}(\boldsymbol{x}_{B,\tau+1}^{\text{post}} - \boldsymbol{x})$$

$$\stackrel{\text{a.s.}}{=} \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}\left[\{f_{\tau'}(x_n + z_{n,\tau'}) - x_n\}\{f_\tau(x_n + z_{n,\tau}) - x_n\}\right]$$

$$+ o(1), \quad (125)$$

where $\{\boldsymbol{H}_\tau\}$ in (124) follow zero-mean Gaussian random matrices with covariance $\mathbb{E}[\boldsymbol{h}_\tau\boldsymbol{h}_{\tau'}^{\text{T}}] = \bar{v}_{A \to B,\tau',\tau}\boldsymbol{I}_N$. Evaluating the correlation $\mathbb{E}[z_{n,\tau'}z_{n,\tau}]$ with (8), we arrive at (51).

The almost sure convergence of $\xi_{B,\tau}$ in (43) to $\bar{\xi}_{B,\tau}$ follows from Assumption 5 and [44, Eq. (104)]. Thus, Theorem 1 holds.

APPENDIX C
PROOF OF THEOREM 2

We first evaluate the covariance $\bar{v}_{A \to B,t',t}$. Lemma 5 implies

$$\bar{v}_{A \to B,t',t} = \lim_{M = \delta N \to \infty} \frac{1}{N}\tilde{\boldsymbol{m}}_{t'}^{\text{T}}\tilde{\boldsymbol{m}}_t, \quad (126)$$

where $\tilde{\boldsymbol{m}}_t$ is given by (106). By definition, we have

$$\bar{v}_{A \to B,t',t} \stackrel{\text{a.s.}}{=} \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} \theta_{A,\tau',t'}\theta_{A,\tau,t}\bar{v}_{A,\tau',\tau}^{\text{ext}} + o(1), \quad (127)$$

which is equal to (57), with

$$(1 - \xi_{A,t'})(1 - \xi_{A,t})\bar{v}_{A,t',t}^{\text{ext}}$$

$$= \bar{v}_{A,t',t}^{\text{post}} - \sum_{\tau' \in \mathcal{T}_{A,t'}} \xi_{A,\tau',t'} \frac{\boldsymbol{b}_{\tau'}^{\text{T}}\boldsymbol{V}^{\text{T}}(\boldsymbol{x}_{A,t}^{\text{post}} - \boldsymbol{x})}{N}$$

$$- \sum_{\tau \in \mathcal{T}_{A,t}} \xi_{A,\tau,t} \frac{(\boldsymbol{x}_{A,t'}^{\text{post}} - \boldsymbol{x})^{\text{T}}\boldsymbol{V}\boldsymbol{b}_\tau}{N}$$

$$+ \sum_{\tau' \in \mathcal{T}_{A,t'}}\sum_{\tau \in \mathcal{T}_{A,t}} \xi_{A,\tau',t'}\xi_{A,\tau,t}\bar{v}_{B \to A,\tau',\tau}, \quad (128)$$

where $N^{-1}\boldsymbol{b}_{\tau'}^{\text{T}}\boldsymbol{b}_\tau = N^{-1}\tilde{\boldsymbol{q}}_{\tau'}^{\text{T}}\tilde{\boldsymbol{q}}_\tau \stackrel{\text{a.s.}}{\to} \bar{v}_{B \to A,\tau',\tau}$ has been used. We apply (26), (107), and [44, Eq. (31)] to obtain

$$\frac{1}{N}\boldsymbol{b}_{\tau'}^{\text{T}}\boldsymbol{V}^{\text{T}}(\boldsymbol{x}_{A,t}^{\text{post}} - \boldsymbol{x}) \stackrel{\text{a.s.}}{=} \sum_{\tau \in \mathcal{T}_{A,t}} \xi_{A,\tau,t}\bar{v}_{B \to A,\tau',\tau} + o(1), \quad (129)$$

where we have used (25). Applying this result to (128) yields

$$(1 - \xi_{A,t'})(1 - \xi_{A,t})\bar{v}_{A,t',t}^{\text{ext}} \stackrel{\text{a.s.}}{=} \bar{v}_{A,t',t}^{\text{post}}$$

$$- \sum_{\tau' \in \mathcal{T}_{A,t'}}\sum_{\tau \in \mathcal{T}_{A,t}} \xi_{A,\tau',t'}\xi_{A,\tau,t}\bar{v}_{B \to A,\tau',\tau} + o(1)$$

$$= \bar{v}_{A,t',t}^{\text{post}} - \bar{\xi}_{A,t'}\bar{\xi}_{A,t}\bar{v}_{B \to A,t',t}^{\text{suf}} + o(1), \quad (130)$$

where the second equality follows from (25), (50), and (55). Thus, we have arrived at (56) with $\bar{v}_{B \to A,t',t}^{\text{suf}}$ given by (50), instead of (54).

We next evaluate $\bar{v}_{B \to A,t'+1,t+1}$ for $t', t \geq 0$. Lemma 5 implies

$$\bar{v}_{B \to A,t'+1,t+1} = \frac{1}{N}\tilde{\boldsymbol{q}}_{t'+1}^{\text{T}}\tilde{\boldsymbol{q}}_{t+1}, \quad (131)$$

where $\tilde{\boldsymbol{q}}_{t+1}$ is given by (109). We use the definition of $\xi_{B,t',t}$ in (43) and Theorem 1 to obtain

$$\bar{v}_{B \to A,t'+1,t+1} \stackrel{\text{a.s.}}{=} \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} \theta_{B,\tau',t'}\theta_{B,\tau,t}\bar{v}_{B,\tau'+1,\tau+1}^{\text{ext}} + o(1), \quad (132)$$

which is equal to (60), with

$$
\begin{aligned}
&(1 - \bar{\xi}_{\mathrm{B},t'})(1 - \bar{\xi}_{\mathrm{B},t})\bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{ext}}\\
&= \bar{v}_{\mathrm{B},t'+1,t+1}^{\mathrm{post}} - \frac{1}{N}\sum_{\tau \in \mathcal{T}_{\mathrm{B},t}}\xi_{\mathrm{B},\tau,t}(\boldsymbol{x}_{\mathrm{B},t'+1}^{\mathrm{post}} - \boldsymbol{x})^{\mathrm{T}}\boldsymbol{h}_\tau\\
&\quad - \frac{1}{N}\sum_{\tau' \in \mathcal{T}_{\mathrm{B},t'}}\xi_{\mathrm{B},\tau',t'}\boldsymbol{h}_{\tau'}^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}} - \boldsymbol{x})\\
&\quad + \frac{1}{N}\sum_{\tau' \in \mathcal{T}_{\mathrm{B},t'}}\sum_{\tau \in \mathcal{T}_{\mathrm{B},t}}\xi_{\mathrm{B},\tau',t'}\xi_{\mathrm{B},\tau,t}\boldsymbol{h}_{\tau'}^{\mathrm{T}}\boldsymbol{h}_\tau.
\end{aligned} \tag{133}
$$

Utilizing [44, Eq. (30)] for $\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}}$ given in (38) with (36), and applying (43) and Theorem 1, we have

$$
\frac{1}{N}\boldsymbol{h}_{\tau'}^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{B},t+1}^{\mathrm{post}} - \boldsymbol{x}) \overset{\mathrm{a.s.}}{=} \sum_{\tau \in \mathcal{T}_{\mathrm{B},t}}\xi_{\mathrm{B},t',t}\bar{v}_{\mathrm{A}\to\mathrm{B},\tau',\tau} + o(1). \tag{134}
$$

Substituting this result into (133) and repeating the derivation of the second equality in (130), we arrive at (59) with $\bar{v}_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}}$ given by (52), instead of (58).

Finally, we evaluate $\bar{v}_{\mathrm{B}\to\mathrm{A},0,t+1} = N^{-1}\tilde{\boldsymbol{q}}_0^{\mathrm{T}}\tilde{\boldsymbol{q}}_{t+1}$. By definition, $\bar{v}_{\mathrm{B}\to\mathrm{A},0,0} = 1$ is trivial. Using $\tilde{\boldsymbol{q}}_0 = -\boldsymbol{x}$, the definition of $\tilde{\boldsymbol{q}}_{t+1}$ in (109), and Theorem 1, we arrive at (61) for $t \geq 0$.

To complete the proof of Theorem 2, we derive (54) and (58). Since we have defined the covariance messages that are consistent to the state evolution recursions in the large system limit, we have $\boldsymbol{V}_{\mathrm{A}\to\mathrm{B},\tau,\tau} \overset{\mathrm{a.s.}}{\to} \bar{\boldsymbol{V}}_{\mathrm{A}\to\mathrm{B},\tau,\tau}$ and $\boldsymbol{V}_{\mathrm{B}\to\mathrm{A},\tau,\tau} \overset{\mathrm{a.s.}}{\to} \bar{\boldsymbol{V}}_{\mathrm{B}\to\mathrm{A},\tau,\tau}$. Thus, (50) and (52) reduce to (54) and (58), respectively. This justifies the replacement of $v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}$ in $\boldsymbol{W}_t$ with $\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}$.

APPENDIX D
POOF OF THEOREM 3

We first prove the identity $\bar{v}_{\mathrm{A}\to\mathrm{B},t',t} = \bar{v}_{\mathrm{A}\to\mathrm{B},t}$ for all $t' \leq t$. When all preceding information $\mathcal{T}_{\mathrm{A},t} = \mathcal{T}_{\mathrm{B},t} = \{0,\ldots,t\}$ is used for all t, $\bar{v}_{\mathrm{B}\to\mathrm{A},t',t}^{\mathrm{suf}}$ in (54) for $t' \leq t$ reduces to

$$
\begin{aligned}
\bar{v}_{\mathrm{B}\to\mathrm{A},t',t}^{\mathrm{suf}} &= \frac{\boldsymbol{1}^{\mathrm{T}}\bar{\boldsymbol{V}}_{\mathrm{B}\to\mathrm{A},t',t'}^{-1}(\boldsymbol{I}_{t'},\boldsymbol{O})\bar{\boldsymbol{V}}_{\mathrm{B}\to\mathrm{A},t,t}\bar{\boldsymbol{V}}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}}{\boldsymbol{1}^{\mathrm{T}}\bar{\boldsymbol{V}}_{\mathrm{B}\to\mathrm{A},t',t'}^{-1}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\bar{\boldsymbol{V}}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}}\\
&= \frac{1}{\boldsymbol{1}^{\mathrm{T}}\bar{\boldsymbol{V}}_{\mathrm{B}\to\mathrm{A},t,t}^{-1}\boldsymbol{1}},
\end{aligned} \tag{135}
$$

as shown in (9). Using this result, the definition of $\gamma_{t',t}$ in (22), the LMMSE filter (30) with $v_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}$ replaced by $\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}$, we find that $\bar{v}_{\mathrm{A},t',t}^{\mathrm{post}}$ in (49) for $t' \leq t$ reduces to

$$
\begin{aligned}
\bar{v}_{\mathrm{A},t',t}^{\mathrm{post}} &\overset{\mathrm{a.s.}}{=} o(1) + \frac{\sigma^2}{N}\mathrm{Tr}\left(\boldsymbol{\Xi}_{t'}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Xi}_t\right)\\
&\quad + \frac{\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}}{N}\mathrm{Tr}\left\{\left(\boldsymbol{I}_N - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Xi}_{t'}\boldsymbol{A}\right)\left(\boldsymbol{I}_N - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Xi}_t\boldsymbol{A}\right)\right\}\\
&= \frac{\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}}{N}\mathrm{Tr}\left(\boldsymbol{I}_N - \boldsymbol{A}^T\boldsymbol{\Xi}_t\boldsymbol{A}\right) + o(1)\\
&= \bar{\xi}_{\mathrm{A},t}\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}} + o(1),
\end{aligned} \tag{136}
$$

with

$$
\boldsymbol{\Xi}_t = \bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}\left(\sigma^2\boldsymbol{I}_M + \bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}, \tag{137}
$$

where the last equality follows from (55). This expression is independent of t'. Thus, we have $\bar{v}_{\mathrm{A},t',t}^{\mathrm{post}} = \bar{v}_{\mathrm{A},t,t}^{\mathrm{post}}$. Substituting $\bar{\xi}_{\mathrm{A},t} = \bar{v}_{\mathrm{A},t,t}^{\mathrm{post}}/\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}$ into (56), we find that, for all $t' \leq t$, $\bar{v}_{\mathrm{A}\to\mathrm{B},t',t}$ in (57) is equal to $\bar{v}_{\mathrm{A}\to\mathrm{B},t}$ given in (84) with $\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}$ replaced by $\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}}$.

We next prove the identity $\bar{v}_{\mathrm{B}\to\mathrm{A},t'+1,t+1} = \bar{v}_{\mathrm{B}\to\mathrm{A},t+1}$ for all $t' \leq t$. Repeating the same derivation for (58) as in (135), we find

$$
\bar{v}_{\mathrm{A}\to\mathrm{B},t',t}^{\mathrm{suf}} = \frac{1}{\boldsymbol{1}^{\mathrm{T}}\bar{\boldsymbol{V}}_{\mathrm{A}\to\mathrm{B},t,t}^{-1}\boldsymbol{1}} = \bar{v}_{\mathrm{A}\to\mathrm{B},t,t}^{\mathrm{suf}} \tag{138}
$$

for all $t' \leq t$. Using the second property in Lemma 2 yields $\bar{v}_{\mathrm{B},t',t}^{\mathrm{post}} = \bar{v}_{\mathrm{B},t,t}^{\mathrm{post}}$ for all $t' \leq t$. Furthermore, we have the general relationship $\bar{\xi}_{\mathrm{B},t} = \bar{v}_{\mathrm{B},t+1,t+1}^{\mathrm{post}}/\bar{v}_{\mathrm{A}\to\mathrm{B},t,t}^{\mathrm{suf}}$ [35, Lemma 2] between the Bayes-optimal denoiser and the MMSE. Applying these results to (59), we find that, for all $t' \leq t$, $\bar{v}_{\mathrm{B}\to\mathrm{A},t'+1,t+1}$ in (60) is equal to $\bar{v}_{\mathrm{B}\to\mathrm{A},t+1}$ given in (85) with $\bar{v}_{\mathrm{A}\to\mathrm{B},t,t}$ replaced by $\bar{v}_{\mathrm{A}\to\mathrm{B},t,t}^{\mathrm{suf}}$.

To complete the proof of the former property in Theorem 3, we need to prove the identities $\bar{v}_{\mathrm{A}\to\mathrm{B},t,t}^{\mathrm{suf}} = \bar{v}_{\mathrm{A}\to\mathrm{B},t,t}$ and $\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}} = \bar{v}_{\mathrm{B}\to\mathrm{A},t,t}$. Without loss of generality, we focus on the former $\bar{v}_{\mathrm{A}\to\mathrm{B},t,t}^{\mathrm{suf}} = \bar{v}_{\mathrm{A}\to\mathrm{B},t,t}$.

We have already proved $\bar{v}_{\mathrm{A}\to\mathrm{B},t',t} = \bar{v}_{\mathrm{A}\to\mathrm{B},t,t}$ for all $t' \leq t$. Furthermore, Theorem 1 implies that $\{\bar{\boldsymbol{V}}_{\mathrm{A}\to\mathrm{B},\tau,\tau}\}_{\tau=1}^{t}$ are positive definite. Thus, we can use the second property in Lemma 3 to arrive at $\bar{v}_{\mathrm{A}\to\mathrm{B},t,t}^{\mathrm{suf}} = \bar{v}_{\mathrm{A}\to\mathrm{B},t,t}$, so that the former property in Theorem 3 holds.

Let us prove the latter property in Theorem 3. Since $\bar{v}_{\mathrm{A}\to\mathrm{B},t,t}^{\mathrm{suf}} = \bar{v}_{\mathrm{A}\to\mathrm{B},t,t}$ and $\bar{v}_{\mathrm{B}\to\mathrm{A},t,t}^{\mathrm{suf}} = \bar{v}_{\mathrm{B}\to\mathrm{A},t,t}$ hold, the first property in Lemma 2 implies that $\{\bar{v}_{\mathrm{A}\to\mathrm{B},t,t} \geq 0\}$ and $\{\bar{v}_{\mathrm{B}\to\mathrm{A},t,t} \geq 0\}$ are monotonically non-increasing sequences with respect to t. Also, it is possible to prove that $\{\bar{v}_{\mathrm{A}\to\mathrm{B},t,t} \geq 0\}$ and $\{\bar{v}_{\mathrm{B}\to\mathrm{A},t,t} \geq 0\}$ are strictly deceasing sequences, by using the first property in Lemma 3. Thus, there are the limits $\lim_{t\to\infty}\bar{v}_{\mathrm{A}\to\mathrm{B},t,t} = \bar{v}_{\mathrm{A}\to\mathrm{B}}$ and $\lim_{t\to\infty}\bar{v}_{\mathrm{B}\to\mathrm{A},t,t} = \bar{v}_{\mathrm{B}\to\mathrm{A}}$, which imply the convergence of $\bar{v}_{\mathrm{B},t+1,t+1}^{\mathrm{post}}$ in (51) to some constant $\bar{v}_{\mathrm{B}}^{\mathrm{post}}$. From (53) and (86), $\bar{\xi}_{\mathrm{B},t}$ and $\bar{\xi}_{\mathrm{A},t}$ converge to some constants $\bar{\xi}_{\mathrm{B}}$ and $\bar{\xi}_{\mathrm{A}}$ as $t \to \infty$. These observations imply that $\bar{v}_{\mathrm{A}\to\mathrm{B},t',t} = \bar{v}_{\mathrm{A}\to\mathrm{B},t}$ and $\bar{v}_{\mathrm{B}\to\mathrm{A},t',t} = \bar{v}_{\mathrm{B}\to\mathrm{A},t}$ given in (84) and (85) converge to a FP as t tends to infinity for all $t' \leq t$.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candès, 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.

[3] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[4] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.

[5] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[6] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.

[7] Z. Fanzi, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 37–48, Feb. 2011.

[8] A. Joseph and A. R. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2541–2557, May 2012.

[9] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 972–974, Jul. 2012.

[10] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

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

[12] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A, Math. General*, vol. 36, no. 43, pp. 11111–11121, Oct. 2003.

[13] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[14] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

[15] 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.

[16] J. Barbier and N. Macris, "The adaptive interpolation method: A simple scheme to prove replica formulas in Bayesian inference," *Probab. Theory Rel. Fields*, vol. 174, nos. 3–4, pp. 1133–1185, Aug. 2019.

[17] 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.

[18] 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.

[19] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[20] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, no. 2, May 2012, Art. no. 021005.

[21] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.

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

[23] K. Takeuchi, T. Tanaka, and T. Kawabata, "Performance improvement of iterative multiuser detection for large sparsely spread CDMA systems by spatial coupling," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1768–1794, Apr. 2015.

[24] F. Caltagirone, L. Zdeborová, and F. Krzakala, "On convergence of approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014, pp. 1812–1816.

[25] S. Rangan, P. Schniter, A. K. 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.

[26] Y. Kabashima and M. Vehkaperä, "Signal recovery using expectation consistent approximation for linear observations," in *Proc. IEEE ISIT*, Honolulu, HI, USA, Jun. 2014, pp. 226–230.

[27] 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. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brisbane, QLD, Australia, Apr. 2015, pp. 2021–2025.

[28] A. Manoel, F. Krzakala, E. Tramel, and L. Zdeborov, "Swept approximate message passing for sparse estimation," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 1123–1132.

[29] S. Rangan, A. K. Fletcher, P. Schniter, and U. S. Kamilov, "Inference for generalized linear models via alternating directions and Bethe free energy minimization," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 676–697, Jan. 2017.

[30] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[31] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.

[32] M. Opper and O. Winther, "Expectation consistent approximate inference," *J. Mach. Learn. Res.*, vol. 6, pp. 2177–2204, Dec. 2005.

[33] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell.*, Seattle, WA, USA, Aug. 2001, pp. 362–369.

[34] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, Aug. 2014.

[35] K. Takeuchi, "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.

[36] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," *Europhys. Lett.*, vol. 76, no. 6, pp. 1193–1199, 2006.

[37] A. M. Tulino, G. Caire, S. Verdú, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, Jul. 2013.

[38] J. Barbier, N. Macris, A. Maillard, and F. Krzakala, "The mutual information in random linear estimation beyond i.i.d. matrices," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 1390–1394.

[39] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and Y. Su, "Convergence analysis and assurance for Gaussian message passing iterative detector in massive MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6487–6501, Sep. 2016.

[40] C. Gerbelot, A. Abbara, and F. Krzakala, "Asymptotic errors for teacher-student convex generalized linear models (or : How to prove Kabashima's replica formula)," 2020, *arXiv:2006.06581*.

[41] M. Mondelli and R. Venkataramanan, "PCA initialization for approximate message passing in rotationally invariant models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, Dec. 2021, pp. 29616–29629.

[42] K. Mimura and J. Takeuchi, "Dynamics of damped approximate message passing algorithms," in *Proc. IEEE Inf. Theory Workshop*, Visby, Sweden, Aug. 2019, pp. 1–5.

[43] L. Liu, S. Huang, and B. M. Kurkoski, "Memory approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1379–1384.

[44] K. Takeuchi, "Bayes-optimal convolutional AMP," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4405–4428, Jul. 2021.

[45] M. Opper, B. Çakmak, and O. Winther, "A theory of solving TAP equations for Ising models with general invariant random matrices," *J. Phys. A, Math. Theor.*, vol. 49, no. 11, Feb. 2016, Art. no. 114002.

[46] Z. Fan, "Approximate message passing algorithms for rotationally invariant matrices," *Ann. Statist.*, vol. 50, no. 1, pp. 197–224, Feb. 2022.

[47] 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, pp. 22120–22144.

[48] K. Takeuchi, "Convolutional approximate message-passing," *IEEE Signal Process. Lett.*, vol. 27, pp. 416–420, 2020.

[49] 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.

[50] K. Takeuchi, "On the convergence of convolutional approximate message-passing for Gaussian signaling," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. 105, no. 2, pp. 100–108, Feb. 2022.

[51] K. Takeuchi, "On the convergence of orthogonal/vector AMP: Long-memory message-passing strategy," in *Proc. IEEE Int. Symp. Inf. Theory*, Espoo, Finland, Jun. 2022, pp. 1318–1323.

[52] A. M. Tulino and S. Verdù, *Random Matrix Theory and Wireless Communications*. Hanover, MA USA: Now Publishers, 2004.

[53] R. Berthier, A. Montanari, and P.-M. Nguyen, "State evolution for approximate message passing with non-separable functions," *Inf. Inference, A, J. IMA*, vol. 9, no. 1, pp. 33–79, Mar. 2020.

[54] Y. Ma, C. Rush, and D. Baron, "Analysis of approximate message passing with non-separable denoisers and Markov random field priors," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7367–7389, Nov. 2019.

[55] A. K. Fletcher, P. Pandit, S. Rangan, S. Sarkar, and P. Schniter, "Plug-in estimation in high-dimensional linear inverse problems a rigorous analysis," *J. Stat. Mech., Theory Exp.*, vol. 2019, Dec. 2019, Art. no. 124021.

[56] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. Providence, RI, USA: Amer. Math. Soc., 2000.

[57] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[58] G. W. Anderson and B. Farrell, "Asymptotically liberating sequences of random unitary matrices," *Adv. Math.*, vol. 255, pp. 381–413, Apr. 2014.

[59] C. Male, *Traffic Distributions and Independence: Permutation Invariant Random Matrices and the Three Notions of Independence*. Providence, RI, USA: Amer. Math. Soc., 2020.

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

[61] C. Stein, "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," in *Proc. 6th Berkeley Symp. Math. Statist. Probab.*, vol. 2, 1972, pp. 583–602.

Keigo Takeuchi (Member, IEEE) received the B.Eng., M.Inf., and Ph.D. degrees in informatics from Kyoto University, Kyoto, Japan, in 2004, 2006, and 2009, respectively.

From 2009 to 2016, he was an Assistant Professor at the Department of Communication Engineering and Informatics, The University of Electro-Communications, Tokyo, Japan. He held visiting appointments at the Norwegian University of Science and Technology (NTNU), Norway; and the National Sun Yat-sen University (NSYSU), Taiwan. He is currently an Associate Professor at the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Aichi, Japan. His research interests are in the field of wireless communications, statistical signal processing, and statistical-mechanical informatics.

Dr. Takeuchi has received the 2008 IEEE Kansai Section Student Paper Award and the IEEE Nagoya Section Young Researcher Awards 2017. He has served as an Associate Editor for *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* from 2013 to 2017.