Bayes-Optimal Convolutional AMP

Keigo Takeuchi, *Member, IEEE*

***Abstract*— This paper proposes Bayes-optimal convolutional approximate message-passing (CAMP) for signal recovery in compressed sensing. CAMP uses the same low-complexity matched filter (MF) for interference suppression as approximate message-passing (AMP). To improve the convergence property of AMP for ill-conditioned sensing matrices, the so-called Onsager correction term in AMP is replaced by a convolution of all preceding messages. The tap coefficients in the convolution are determined so as to realize asymptotic Gaussianity of estimation errors via state evolution (SE) under the assumption of orthogonally invariant sensing matrices. An SE equation is derived to optimize the sequence of denoisers in CAMP. The optimized CAMP is proved to be Bayes-optimal for all orthogonally invariant sensing matrices if the SE equation converges to a fixed-point and if the fixed-point is unique. For sensing matrices with low-to-moderate condition numbers, CAMP can achieve the same performance as high-complexity orthogonal/vector AMP that requires the linear minimum mean-square error (LMMSE) filter instead of the MF.**

***Index Terms*— Compressed sensing, approximate message-passing (AMP), orthogonal/vector AMP, convolutional AMP, large system limit, state evolution.**

I. Introduction

A. Compressed Sensing

COMPRESSED sensing (CS) [1], [2] is a powerful technique for recovering sparse signals from compressed measurements. Under the assumption of linear measurements, CS is formulated as estimation of a sparse signal vector $\boldsymbol{x} \in \mathbb{R}^N$ from a compressed measurement vector $\boldsymbol{y} \in \mathbb{R}^M$ ($M \leq N$) and a sensing matrix $\boldsymbol{A} \in \mathbb{R}^{M\times N}$, given by

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

where $\boldsymbol{w} \in \mathbb{R}^M$ is an unknown additive noise vector.

For simplicity in information-theoretic discussion [3], suppose that the signal vector $\boldsymbol{x}$ has independent and identically distributed (i.i.d.) elements. Sparsity of signals is measured with the Rényi information dimension [4] of each signal element. When each signal takes non-zero real values with probability $\rho \in [0, 1]$, the information dimension is equal to ρ.

Manuscript received March 27, 2020; revised April 1, 2021; accepted April 28, 2021. Date of publication May 4, 2021; date of current version June 16, 2021. This work was supported in part by the Grant-in-Aid for Scientific Research (B) (JSPS KAKENHI), Japan, under Grant 18H01441 and Grant 21H01326. This article was presented in part at the 2019 IEEE International Symposium on Information Theory. This article is to be presented at the 2021 IEEE International Symposium on Information Theory.

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

Communicated by R. Venkataramanan, Associate Editor for Machine Learning.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2021.3077471.

Digital Object Identifier 10.1109/TIT.2021.3077471

In the noiseless case $\boldsymbol{w} = \boldsymbol{0}$, Wu and Verdú [3] proved that, if and only if the compression rate $\delta = M/N$ is equal to or larger than the information dimension, there are some sensing matrix $\boldsymbol{A}$ and method for signal recovery such that the signal vector $\boldsymbol{x}$ can be recovered with negligibly small error probability in the large system limit, where M and N tend to infinity with the compression rate δ kept constant. Thus, an important issue in CS is a construction of practical sensing matrices and a low-complexity algorithm for signal recovery achieving the information-theoretic compression limit.

Important examples of sensing matrices are zero-mean i.i.d. sensing matrices [5] and random sensing matrices with orthogonal rows [6]. The information-theoretic compression limit of zero-mean i.i.d. sensing matrices was analyzed with the non-rigorous replica method [7], [8]—a tool developed in statistical mechanics [9], [10]. The compression limit is characterized via a potential function called free energy. The results themselves were rigorously justified in [11]–[14] while the justification of the replica method is still open. It is a simple exercise to prove that the compression limit for zero-mean i.i.d. sensing matrices is equal to the Rényi information dimension in the noiseless case, by using a relationship between the information dimension and mutual information [15, Theorem 6].

Random sensing matrices with orthogonal rows can be constructed efficiently in terms of both time and space complexity while zero-mean i.i.d. sensing matrices require $\mathcal{O}(MN)$ time and memory for matrix-vector multiplication. When the fast Fourier transform or fast Walsh-Hadamard transform is used, the matrix-vector multiplication needs $\mathcal{O}(N \log N)$ time and $\mathcal{O}(N)$ memory. Thus, random sensing matrices with orthogonal rows are preferable from a practical point of view.

The class of orthogonally invariant matrices includes zero-mean i.i.d. Gaussian matrices and Haar orthogonal matrices [16], [17], of which the latter is regarded as an idealized model of random matrices with orthogonal rows. The class allows us to analyze the information-theoretic compression limit in signal recovery. The replica method [18], [19] was used to analyze the compression limit for orthogonally invariant sensing matrices. The replica results themselves were justified in [20]. In particular, Haar orthogonal matrices achieve the Welch lower bound [21] and were proved to be optimal for Gaussian [22] and general [23] signals. In the noiseless case, of course, Haar orthogonal sensing matrices achieve the compression rate that is equal to the Rényi information dimension.

In practical systems, the measurement vector is subject not only to additive noise but also to multiplicative noise. A typical example is fading in wireless communication systems [24], [25]. The effective sensing matrix containing fading influence may be ill-conditioned even if a Haar orthogonal sensing matrix is used. Such effective sensing matrices can be modeled as orthogonally invariant matrices. Thus, an ultimate algorithm for signal recovery is required to be low complexity and Bayes-optimal for all orthogonally invariant sensing matrices.

B. Message-Passing

A promising solution to signal recovery is message-passing (MP). Approximate message-passing (AMP) [26] is a low-complexity and powerful algorithm for signal recovery from zero-mean i.i.d. sub-Gaussian measurements. Bayes-optimal AMP is regarded as an exact large-system approximation of loopy belief propagation (BP) [27]. The main feature of AMP is the so-called Onsager correction to realize asymptotic Gaussianity of the estimation errors before denoising. The Onsager correction originates from that in the Thouless-Anderson-Palmer (TAP) equation [28] for a solvable spin glass model with i.i.d. interaction between all spins [29]. The Onsager correction cancels intractable dependencies of the current estimation error on past estimation errors due to i.i.d. dense sensing matrices.

The convergence property of AMP was analyzed rigorously via state evolution (SE) [30], [31], inspired by Bolthausen's conditioning technique [32]. SE is a dense counterpart of density evolution [33] in sparse systems. SE tracks a few state variables to describe rigorous dynamics of MP in the large system limit. SE analysis in [30], [31] implies that AMP is Bayes-optimal for zero-mean i.i.d. sub-Gaussian sensing matrices when the compression rate δ is larger than a certain value called BP threshold [34]. Spatial coupling [34]–[37] is needed to realize the optimality of AMP for any compression rate. However, this paper does not consider spatial coupling since spatial coupling is a universal technique [34] to improve the performance of MP.

A disadvantage of AMP is that AMP fails to converge when the sensing matrix is non-zero mean [38] or ill-conditioned [39]. To solve this issue, orthogonal AMP (OAMP) [40] and vector AMP [41], [42] were proposed. The two MP algorithms are equivalent to each other. Bayes-optimal OAMP/VAMP can be regarded as an exact large-system approximation of expectation propagation (EP) [43]–[46]. Rigorous SE analysis [41], [42], [45], [46] proved that OAMP/VAMP is Bayes-optimal for orthogonally invariant sensing matrices when the compression rate is larger than BP threshold. While non-zero mean matrices are outside the class of orthogonally invariant matrices, numerical simulations in [42] indicated that OAMP/VAMP can treat the non-zero mean case.

A prototype of OAMP/VAMP was originally proposed by Opper and Winther [47, Appendix D]. Historically, they [48] generalized the Onsager correction in the TAP equation [28] from zero-mean i.i.d. spin interaction to orthogonally invariant interaction. Their method was formulated as the expectation-consistency (EC) approximation [47]. The EC approximation itself does not produce MP algorithms but a potential function of which a local minimum should be solved with some MP algorithm. OAMP/VAMP can be derived from an EP-type iteration–called a single loop algorithm [47]—to solve a local minimum of the EC potential. See [49, Appendix A] for the derivation of OAMP/VAMP via the EC approximation.

The main weakness of OAMP/VAMP is a per-iteration requirement of the linear minimum mean-square error (LMMSE) filter, of which the time complexity is $\mathcal{O}(M^3 + M^2N)$ per iteration. The singular-value decomposition (SVD) of the sensing matrix allows us to circumvent the use of the LMMSE filter [42]. However, the complexity of the SVD itself is high in general. The performance of OAMP/VAMP degrades significantly when the LMMSE filter is replaced by the low-complexity matched filter (MF) [40] used in AMP. Thus, OAMP/VAMP can be applied only to limited problems in which the SVD of the sensing matrix is computed efficiently.

In summary, it is still open to construct a low-complexity and Bayes-optimal MP algorithm for all orthogonally invariant sensing matrices. The purpose of this paper is to tackle the design issue of such ultimate MP algorithms.

C. Methodology

The main idea of this paper is to extend the class of MP algorithms. Conventional MP algorithms use update rules that depend only on messages in the latest iteration. Long-memory MP algorithms considered in this paper are allowed to depend on messages in all preceding iterations.

This class of long-memory MP algorithms was motivated by SE analysis of AMP for orthogonally invariant sensing matrices [50]. When the asymptotic singular-value distribution of the sensing matrix is equal to that of zero-mean i.i.d. Gaussian matrices, the error model of AMP was proved to be an instance of a general error model [50], in which each error depends on errors in all preceding iterations. This result implies that the Onsager correction in AMP uses messages in all preceding iterations to realize the asymptotic Gaussianity of the current estimation error while the representation itself of the correction term looks as if only messages in the latest iteration are utilized. Inspired by this observation, we consider long-memory MP algorithms as a starting point.

The idea of long-memory MP was originally proposed in Opper, Çakmak, and Winther's paper [51] to solve the TAP equations for spin glass models with orthogonally invariant interaction. Their methodology was based on non-rigorous dynamical functional theory. After the initial submission of this paper, their results were rigorously justified via SE in [52].

The proposed design of long-memory MP consists of three steps: A first step is an establishment of rigorous SE for analyzing the dynamics of long-memory MP algorithms for orthogonally invariant sensing matrices. This step has been already established in [50] by generalizing conventional SE analysis [42], [46] to the long-memory case. The SE analysis provides a sufficient condition for a long-memory MP algorithm to have Gaussian-distributed estimation errors in the large system limit. The main advantage in the SE analysis is to provide a systematic design of long-memory MP that satisfies the asymptotic Gaussianity in estimation errors while the class of long-memory MP is slightly smaller than in [51], [52].

A second step is to modify the Onsager correction in AMP so as to satisfy the sufficient condition for the asymptotic Gaussianity. A solvable class of long-memory MP was proposed in [53], where the Onsager correction was defined as a convolution of messages in all preceding iterations. The tap coefficients in the convolution were determined so as to satisfy the sufficient condition. Thus, long-memory MP proposed in [53] was called convolutional AMP (CAMP) and is the main object of this paper.

This paper generalizes CAMP in [53], motivated by an implementation of OAMP/VAMP based on conjugate gradient (CG) [54]. OAMP/VAMP applies the LMMSE filter to a message $\boldsymbol{z} \in \mathbb{R}^M$ after interference subtraction. The LMMSE filter is decomposed into a noise-whitening filter and MF. In principle, CG approximates the output of the noise-whitening filter with a vector in the Krylov subspace spanned by $\{\boldsymbol{z}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}, (\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^2\boldsymbol{z}, \ldots\}$ i.e. a finite weighted sum of $\{(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^j\boldsymbol{z}\}$. On the other hand, messages in the original CAMP [53] are in the 0th Krylov subspace $\{\alpha\boldsymbol{z} : \alpha \in \mathbb{R}\}$ since only the MF is used. To fill this gap, we generalize a convolution of all preceding messages in the original CAMP [53] to that of affine transforms of the preceding messages.

The last step is to optimize the sequence of denoisers in CAMP [55]. The optimization requires information on the distribution of the estimation errors before denoising in each iteration. Since the estimation errors are asymptotically Gaussian-distributed, we need to track the dynamics of the variance of the estimation errors. To analyze this dynamics, we utilize the SE analysis established in the first step.

D. Contributions

The contributions of this paper are sixfold: A first contribution (Theorem 1 in Section II) is to propose a general error model for long-memory MP and prove the asymptotic Gaussianity of estimation errors in the general error model via rigorous SE under the assumption of orthogonally invariant sensing matrices. The general error model contains both error models of AMP and OAMP/VAMP.

A second contribution (Section III-A) is the addition of a convolution proportional to $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ to the Onsager correction in [53], according to the above-mentioned argument on the Krylov subspace. This addition improves the convergence property of CAMP.

A third contribution (Theorem 2 in Section III-C) is to design tap coefficients in the convolution so as to guarantee the asymptotic Gaussianity of estimation errors for all orthogonally invariant sensing matrices. Part of the tap coefficients are used to realize the asymptotic Gaussianity. The remaining coefficients can be utilized to improve the convergence property of CAMP.

A fourth contribution (Theorem 3 in Section III-C) is to present the designed tap coefficients in closed-form. This closed-form representation circumvents numerical instability in solving the tap coefficients numerically. The third and fourth contributions are based on the same proof strategy as in [53].

A fifth contribution (Theorems 4 and 5 in Section III-D) is to optimize the sequence of denoisers in CAMP. An SE equation is derived to describe the dynamics of the variance of the estimation errors before denoising in CAMP. The SE equation is a two-dimensional nonlinear difference equation. By analyzing the fixed-point of the SE equation, we prove that optimized CAMP is Bayes-optimal for all orthogonally invariant sensing matrices if the SE equation converges to a fixed-point and if the fixed-point is unique.

The last contribution (Section IV) is numerical evaluation of CAMP. The remaining parameters in the Bayes-optimal CAMP are optimized numerically to improve the convergence property. Numerical simulations show that the CAMP can converge for sensing matrices with larger condition numbers than the original CAMP [53] when the design parameters are optimized. The CAMP can achieve the same performance as OAMP/VAMP for sensing matrices with low-to-moderate condition numbers while it is inferior to OAMP/VAMP for high condition numbers.

E. Organization

The remainder of this paper is organized as follows: After summarizing the notation used in this paper, we present a unified SE framework for analyzing long-memory MP under the assumption of orthogonally invariant sensing matrices in Section II. This section corresponds to the first step for proposing Bayes-optimal CAMP.

In Section III, we propose CAMP with design parameters. This section corresponds to the remaining two steps for establishing Bayes-optimal CAMP. The proposed CAMP is more general than in [53]. We utilize the SE framework established in Section II to determine the tap coefficients in CAMP that guarantee the asymptotic Gaussianity of estimation errors. To design the remaining design parameters, we derive an SE equation to optimize the performance of signal recovery.

Section IV presents numerical results. The remaining design parameters in CAMP are optimized via numerical simulations. The optimized CAMP is compared to conventional AMP and OAMP/VAMP via the SE equation and numerical simulations. Section V concludes this paper. The details for the proofs of the main theorems are presented in appendices.

F. Notation

For a matrix $\boldsymbol{M}$, the transpose of $\boldsymbol{M}$ is denoted by $\boldsymbol{M}^{\mathrm{T}}$. The notation $\mathrm{Tr}(\boldsymbol{A})$ represents the trace of a square matrix $\boldsymbol{A}$. For a symmetric matrix $\boldsymbol{A}$, the minimum eigenvalue of $\boldsymbol{A}$ is written as $\lambda_{\min}(\boldsymbol{A})$. The notation $\mathcal{O}_{M\times N}$ denotes the space of all possible $M \times N$ matrices with orthonormal columns for $M \geq N$ and orthonormal rows for $M < N$. In particular, $\mathcal{O}_{N\times N}$ reduces to the space $\mathcal{O}_N$ of all possible $N \times N$ orthogonal matrices.

For a vector $\boldsymbol{v}$, the notation $\mathrm{diag}(\boldsymbol{v})$ denotes the diagonal matrix of which the nth diagonal element is equal to $v_n = [\boldsymbol{v}]_n$. The norm $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}$ represents the Euclidean norm. For a matrix $\boldsymbol{M}_i$ with an index i, the tth column of $\boldsymbol{M}_i$

is denoted by $\boldsymbol{m}_{i,t}$. Furthermore, we write the nth element of $\boldsymbol{m}_{i,t}$ as $m_{i,t,n}$.

The Kronecker delta is denoted by $\delta_{\tau,t}$ while the Dirac delta function is represented as $\delta(\cdot)$. We write the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The notations $\overset{\text{a.s.}}{\to}$ and $\overset{\text{a.s.}}{=}$ denote almost sure convergence and equivalence, respectively.

We use the notational convention $\sum_{t=t_1}^{t_2} \cdots = 0$ and $\prod_{t=t_1}^{t_2} \cdots = 1$ for $t_1 > t_2$. For any multivariate function $\phi : \mathbb{R}^t \to \mathbb{R}$, the notation $\partial_{t'}\phi$ for $t' = 0, \ldots, t-1$ denotes the partial derivative of ϕ with respect to the t'th variable $x_{t'}$,

$$\partial_{t'}\phi = \frac{\partial \phi}{\partial x_{t'}}(x_0, \ldots, x_{t-1}). \tag{2}$$

For any vector $\boldsymbol{v} \in \mathbb{R}^N$, the notation $\langle \boldsymbol{v} \rangle = N^{-1}\sum_{n=1}^N v_n$ represents the arithmetic mean of the elements. For any scalar function $f :\in \mathbb{R} \to \mathbb{R}$, the notation $f(\boldsymbol{v})$ means the element-wise application of f to a vector $\boldsymbol{v}$, i.e. $[f(\boldsymbol{v})]_n = f(v_n)$.

For a sequence $\{p_t\}_{t=0}^{\infty}$, we define the Z-transform of $\{p_t\}$ as

$$P(z) = \sum_{t=0}^{\infty} p_t z^{-t}. \tag{3}$$

For two sequences $\{p_t, q_t\}_{t=0}^{\infty}$, we define the convolution operator $*$ as

$$p_{t+i} * q_{t+j} = \sum_{\tau=0}^{t} p_{\tau+i} q_{t-\tau+j}, \tag{4}$$

with $p_t = 0$ and $q_t = 0$ for $t < 0$. For finite-length sequences $\{p_t\}_{t=0}^{T}$ of length $T+1$, we transform them into infinite-length sequences by adding $p_t = 0$ and $q_t = 0$ for all $t > T$.

For two arrays $\{a_{t',t}, b_{t',t} : t', t = 0, \ldots, \infty\}$, we write the two-dimensional convolution as

$$a_{t'+i,t+j} * b_{t'+k,t+l} = \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} a_{\tau'+i,\tau+j} b_{t'-\tau'+k,t-\tau+l}, \tag{5}$$

where $a_{t',t} = 0$ and $b_{t',t} = 0$ are defined for $t' < 0$ or $t < 0$.

Whether a convolution is one-dimensional can be distinguished as follows: A convolution is one-dimensional, such as $a_{t+i} * b_{t+j}$, when both operands contain only one identical subscript. On the other hand, a convolution is two-dimensional, such as $(a_{t'}a_{t+i}) * b_{t'+j,t}$, when both operands include two identical subscripts.

II. Unified Framework

A. Definitions and Assumptions

We define the statistical properties of the random variables in the measurement model (1). The performance of MP is commonly measured in terms of the mean-square error (MSE). Nonetheless, we follow [30] to consider a general performance measure in terms of separable and pseudo-Lipschitz functions while we assume the separability and Lipschitz-continuity for denoisers.

Definition 1: A vector-valued function $\boldsymbol{f} = (f_1, \ldots, f_N)^{\mathrm{T}} : \mathbb{R}^{N\times t} \to \mathbb{R}^N$ is said to be separable if $[\boldsymbol{f}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)]_n = f_n(x_{1,n}, \ldots, x_{t,n})$ holds for all $\boldsymbol{x}_i \in \mathbb{R}^N$.

Definition 2: A function $f : \mathbb{R}^t \to \mathbb{R}$ is said to be pseudo-Lipschitz of order k [30] if there are some Lipschitz constant $L > 0$ and some order $k \in \mathbb{N}$ such that for all $\boldsymbol{x} \in \mathbb{R}^t$ and $\boldsymbol{y} \in \mathbb{R}^t$

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L(1 + \|\boldsymbol{x}\|^{k-1} + \|\boldsymbol{y}\|^{k-1})\|\boldsymbol{x} - \boldsymbol{y}\|. \tag{6}$$

By definition, any pseudo-Lipschitz function of order $k = 1$ is Lipschitz-continuous. A vector-valued function $\boldsymbol{f} = (f_1, \ldots, f_N)^{\mathrm{T}}$ is pseudo-Lipschitz if all element functions $\{f_n\}$ are pseudo-Lipschitz.

Definition 3: A separable pseudo-Lipschitz function $\boldsymbol{f} : \mathbb{R}^{N\times t} \to \mathbb{R}^N$ is said to be proper if the Lipschitz constant $L_n > 0$ of the nth function f_n satisfies

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} L_n^j < \infty \tag{7}$$

for any $j \in \mathbb{N}$.

A proper pseudo-Lipschitz function allows us apply a proof strategy for pseudo-Lipschitz functions with n-independent Lipschitz constant $L_n = L$ to the n-dependent case straightforwardly. The space of all possible separable and proper pseudo-Lipschitz functions of order k is denoted by $\mathcal{PL}(k)$. We have the inclusion relation $\mathcal{PL}(k) \subset \mathcal{PL}(k')$ for all $k < k'$ since $\|\boldsymbol{x}\|^k \leq \|\boldsymbol{x}\|^{k'}$ holds for $\|\boldsymbol{x}\| \gg 1$.

We assume statistical properties of the signal vector associated with separable and proper pseudo-Lipschitz functions of order $k \geq 2$. Note that the integer k in the following assumptions is an identical parameter that is equal to the order of separable and proper pseudo-Lipschitz functions used in SE to measure the performance of MP. If the MSE is considered, the integer k is set to 2.

Assumption 1: The signal vector $\boldsymbol{x}$ satisfies the following strong law of large numbers:

$$\langle \boldsymbol{f}(\boldsymbol{x}) \rangle - \mathbb{E}\left[\langle \boldsymbol{f}(\boldsymbol{x}) \rangle\right] \overset{\text{a.s.}}{\to} 0 \tag{8}$$

as $N \to \infty$ for any separable and proper pseudo-Lipschitz function $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^N$ of order $k \geq 2$. Furthermore, $\boldsymbol{x}$ has zero-mean and bounded $(2k - 2 + \epsilon)$th moments for some $\epsilon > 0$.

Assumption 1 follows from the classical strong law of large numbers when $\boldsymbol{x}$ has i.i.d. elements.

Definition 4: An orthogonal matrix $\boldsymbol{V} \in \mathcal{O}_N$ is said to be Haar-distributed [16] if $\boldsymbol{V}$ is orthogonally invariant, i.e. $\boldsymbol{V} \sim \boldsymbol{\Phi}\boldsymbol{V}\boldsymbol{\Psi}$ for all orthogonal matrices $\boldsymbol{\Phi}, \boldsymbol{\Psi} \in \mathcal{O}_N$ independent of $\boldsymbol{V}$.

Assumption 2: The sensing matrix $\boldsymbol{A}$ is right-orthogonally invariant, i.e. $\boldsymbol{A} \sim \boldsymbol{A}\boldsymbol{\Psi}$ for any orthogonal matrix $\boldsymbol{\Psi} \in \mathcal{O}_N$ independent of $\boldsymbol{A}$. More precisely, the orthogonal matrix $\boldsymbol{V} \in \mathcal{O}_N$ in the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ is Haar-distributed and independent of $\boldsymbol{U}\boldsymbol{\Sigma}$. 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 assumption of unit first moment implies the almost sure convergence $N^{-1}\mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) \overset{\text{a.s.}}{\to} 1$ in the large system limit. Assumption 2 holds when $\boldsymbol{A}$ has zero-mean i.i.d. Gaussian elements with variance M^{-1}. As shown in SE, the asymptotic Gaussianity of estimation errors in MP depends heavily on the

Haar assumption of $\boldsymbol{V}$. Intuitively, the orthogonal transform $\boldsymbol{Va}$ of a vector $\boldsymbol{a} \in \mathbb{R}^N$ is distributed as $N^{-1/2}\|\boldsymbol{a}\|\boldsymbol{z}$ in which $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ is a standard Gaussian vector and independent of $\|\boldsymbol{a}\|$. When the amplitude $N^{-1/2}\|\boldsymbol{a}\|$ tends to a constant as $N \to \infty$, the vector $\boldsymbol{Va}$ looks like a Gaussian vector. This is a rough intuition on the asymptotic Gaussianity of estimation errors.

Assumption 3: The noise vector $\boldsymbol{w}$ is orthogonally invariant, i.e. $\boldsymbol{w} \sim \boldsymbol{\Phi w}$ for any orthogonal matrix $\boldsymbol{\Phi} \in \mathcal{O}_M$ independent of $\boldsymbol{w}$. Furthermore, $\boldsymbol{w}$ has zero-mean, $\lim_{M\to\infty} M^{-1}\|\boldsymbol{w}\|^2 \overset{\text{a.s.}}{=} \sigma^2 > 0$, and bounded $(2k-2+\epsilon)$th moments for some $\epsilon > 0$.

Assumption 3 holds when $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_M)$ is an additive white Gaussian noise (AWGN) vector. It holds for $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{w}$ when the sensing matrix $\boldsymbol{A}$ is left-orthogonally invariant, i.e. $\boldsymbol{A} \sim \boldsymbol{\Phi A}$ for any orthogonal matrix $\boldsymbol{\Phi} \in \mathcal{O}_M$ independent of $\boldsymbol{A}$.

B. General Error Model

We propose a unified framework of SE for analyzing MP algorithms that have asymptotically Gaussian-distributed estimation errors for orthogonally invariant sensing matrices. Instead of starting with concrete MP algorithms, we consider a general class of error models. The proposed class does not necessarily contain the error models of all possible long-memory MP algorithms. However, it is a natural class of error models that allows us to prove the asymptotic Gaussianity of estimation errors for orthogonally invariant sensing matrices via a generalization of conventional SE [46].

Let $\boldsymbol{h}_t \in \mathbb{R}^N$ and $\boldsymbol{q}_{t+1} \in \mathbb{R}^N$ denote error vectors in iteration t before and after denoising, respectively. We assume that the error vectors are recursively given by

$$\boldsymbol{b}_t = \boldsymbol{V}^{\mathrm{T}}\tilde{\boldsymbol{q}}_t, \quad \tilde{\boldsymbol{q}}_t = \boldsymbol{q}_t - \sum_{t'=0}^{t-1}\langle \partial_{t'}\boldsymbol{\psi}_{t-1}\rangle \boldsymbol{h}_{t'}, \tag{9}$$

$$\boldsymbol{m}_t = \boldsymbol{\phi}_t(\boldsymbol{B}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda}), \tag{10}$$

$$\boldsymbol{h}_t = \boldsymbol{V}\tilde{\boldsymbol{m}}_t, \quad \tilde{\boldsymbol{m}}_t = \boldsymbol{m}_t - \sum_{t'=0}^{t}\langle \partial_{t'}\boldsymbol{\phi}_t\rangle \boldsymbol{b}_{t'}, \tag{11}$$

$$\boldsymbol{q}_{t+1} = \boldsymbol{\psi}_t(\boldsymbol{H}_{t+1}, \boldsymbol{x}), \tag{12}$$

with $\boldsymbol{q}_0 = -\boldsymbol{x}$. In (9), the orthogonal matrix $\boldsymbol{V} \in \mathcal{O}_N$ consists of the right-singular vectors in the SVD $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$, with $\boldsymbol{U} \in \mathcal{O}_M$. In (10) and (12), we have defined $\boldsymbol{B}_{t+1} = (\boldsymbol{b}_0, \ldots, \boldsymbol{b}_t)$ and $\boldsymbol{H}_{t+1} = (\boldsymbol{h}_0, \ldots, \boldsymbol{h}_t)$. Furthermore, $\boldsymbol{\lambda} \in \mathbb{R}^N$ is the vector of eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. The vector $\tilde{\boldsymbol{w}} \in \mathbb{R}^N$ is given by

$$\tilde{\boldsymbol{w}} = \begin{bmatrix} \boldsymbol{U}^{\mathrm{T}}\boldsymbol{w} \\ \boldsymbol{0} \end{bmatrix}, \tag{13}$$

where $\boldsymbol{w}$ is the additive noise vector in (1).

The vector-valued functions $\boldsymbol{\phi}_t : \mathbb{R}^{N\times(t+3)} \to \mathbb{R}^N$ and $\boldsymbol{\psi}_t : \mathbb{R}^{N\times(t+2)} \to \mathbb{R}^N$ are assumed to be separable, nonlinear, and proper Lipschitz-continuous.

Assumption 4: The functions $\boldsymbol{\phi}_t$ and $\boldsymbol{\psi}_t$ are separable. The nonlinearities $\boldsymbol{\phi}_t \neq \sum_{t'=0}^{t}\boldsymbol{D}_{t'}\boldsymbol{b}_{t'}$ and $\boldsymbol{\psi}_t \neq \sum_{t'=0}^{t}\tilde{\boldsymbol{D}}_{t'}\boldsymbol{h}_{t'}$ hold for all diagonal matrices $\{\boldsymbol{D}_{t'}, \tilde{\boldsymbol{D}}_{t'}\}$. The function $\boldsymbol{\phi}_t$ is Lipschitz-continuous with respect to the first $t+2$ variables and proper while $\boldsymbol{\psi}_t$ is proper Lipschitz-continuous with respect to all variables.

It might be possible to relax Assumption 4 to the non-separable case [56]–[58]. For simplicity, however, this paper postulates separable denoisers. The nonlinearity is a technical condition for circumventing the zero norm $N^{-1}\|\tilde{\boldsymbol{q}}_t\|^2 = 0$ or $N^{-1}\|\tilde{\boldsymbol{m}}_t\|^2 = 0$, which implies error-free estimation $N^{-1}\|\boldsymbol{b}_t\|^2 = 0$ or $N^{-1}\|\boldsymbol{h}_t\|^2 = 0$.

By definition, the nth function $\phi_{t,n}$ has a λ_n-dependent Lipschitz constant $L_n = L_n(\lambda_n)$. Thus, the proper assumption for $\boldsymbol{\phi}_t$ may be regarded as a condition on the asymptotic eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, as well as a condition on the denoiser $\boldsymbol{\phi}_t$. For example, $\boldsymbol{\phi}_t$ is proper when the asymptotic eigenvalue distribution has a compact support and when the Lipschitz constant $L_n(\lambda_n)$ itself is a pseudo-Lipschitz function of λ_n.

The main feature of the general error model is in the definitions of $\tilde{\boldsymbol{q}}_t$ and $\tilde{\boldsymbol{m}}_t$. The second terms on the right-hand sides (RHSs) of (9) and (11) are correction terms to realize the asymptotic Gaussianity of $\{\boldsymbol{b}_t\}$ and $\{\boldsymbol{h}_t\}$. The correction terms are a modification of conventional correction that allows us to prove the asymptotic Gaussianity via a natural generalization [59] of Stein's lemma used in conventional SE [46]. See Lemma 2 in Appendix A for the details.

The following examples imply that the general error model (9)–(12) contains those of OAMP/VAMP and AMP.

Example 1: Consider OAMP/VAMP [40], [42] with a sequence of scalar denoisers $f_t : \mathbb{R} \to \mathbb{R}$:

$$\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t} = \boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t} + \gamma_t\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}_t^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}), \tag{14}$$

$$v_{\mathrm{A}\to\mathrm{B},t} = \gamma_t - v_{\mathrm{B}\to\mathrm{A},t}, \tag{15}$$

$$\boldsymbol{W}_t = \sigma^2\boldsymbol{I}_M + v_{\mathrm{B}\to\mathrm{A},t}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}, \tag{16}$$

$$\gamma_t^{-1} = \frac{1}{N}\mathrm{Tr}\left(\boldsymbol{W}_t^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right), \tag{17}$$

$$\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t+1} = v_{\mathrm{B}\to\mathrm{A},t+1}\left(\frac{f_t(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t})}{\xi_t v_{\mathrm{A}\to\mathrm{B},t}} - \frac{\boldsymbol{x}_{\mathrm{A}\to\mathrm{B},t}}{v_{\mathrm{A}\to\mathrm{B},t}}\right), \tag{18}$$

$$\frac{1}{v_{\mathrm{B}\to\mathrm{A},t+1}} = \frac{1}{\xi_t v_{\mathrm{A}\to\mathrm{B},t}} - \frac{1}{v_{\mathrm{A}\to\mathrm{B},t}}, \tag{19}$$

with $\xi_t = \langle f_t'(\boldsymbol{x}_{\mathrm{A}\to\mathrm{B}}^t)\rangle$.

It is an exercise to prove that the error model of the OAMP/VAMP is an instance of the general error model with

$$[\boldsymbol{\phi}_t(\boldsymbol{b}_t, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})]_n = b_{t,n} - \frac{\gamma_t\lambda_n b_{t,n} - \gamma_t\sqrt{\lambda_n}\tilde{w}_n}{\sigma^2 + v_{\mathrm{B}\to\mathrm{A},t}\lambda_n}, \tag{20}$$

$$\boldsymbol{\psi}_t(\boldsymbol{h}_t, \boldsymbol{x}) = \frac{f_t(\boldsymbol{x} + \boldsymbol{h}_t) - \boldsymbol{x}}{1 - \xi_t}, \tag{21}$$

by using the fact that ξ_t converges almost surely to a constant in the large system limit [42], [46]. The two separable functions $\boldsymbol{\psi}_t$ and $\boldsymbol{\phi}_t$ for the OAMP/VAMP depend only on the vectors $\boldsymbol{b}_t$ and $\boldsymbol{h}_t$ in the latest iteration.

Example 2: Consider AMP [26] with a sequence of scalar denoisers $f_t : \mathbb{R} \to \mathbb{R}$:

$$\boldsymbol{x}_{t+1} = f_t(\boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t), \tag{22}$$

$$\boldsymbol{z}_t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_t + \frac{\xi_{t-1}}{\delta}\boldsymbol{z}_{t-1}. \tag{23}$$

Suppose that the empirical eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is equal to that for zero-mean i.i.d. Gaussian matrix $\boldsymbol{A}$ in the large system limit. Then, the error model of the AMP was proved in [50] to be an instance of the general error model with

$$\boldsymbol{\phi}_t = (\boldsymbol{I}_N - \boldsymbol{\Lambda})\boldsymbol{b}_t - \frac{\xi_{t-1}}{\delta}\boldsymbol{b}_{t-1} + \mathrm{diag}(\{\sqrt{\lambda_n}\})\tilde{\boldsymbol{w}} + \xi_{t-1}\left\{\left(1+\frac{1}{\delta}\right)\boldsymbol{I}_N - \boldsymbol{\Lambda}\right\}\boldsymbol{\phi}_{t-1} - \frac{\xi_{t-1}\xi_{t-2}}{\delta}\boldsymbol{\phi}_{t-2}, \tag{24}$$

$$\boldsymbol{\psi}_t(\boldsymbol{h}_t, \boldsymbol{x}) = f_t(\boldsymbol{x} + \boldsymbol{h}_t) - \boldsymbol{x}, \tag{25}$$

with $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda})$ and $\xi_t = \langle f_t'(\boldsymbol{x} + \boldsymbol{h}_t)\rangle$. Note that $\boldsymbol{\phi}_t$ is a function of $\boldsymbol{B}_{t+1}$ while $\boldsymbol{\psi}_t$ is a function of $\boldsymbol{h}_t$.

C. State Evolution

A rigorous SE result for the general error model (9)–(12) is presented in the large system limit.

Theorem 1: Suppose that Assumptions 1–4 hold. Then, the following properties hold for all $t = 0, \ldots$ and $t' = 0, \ldots, t$ in the large system limit:

1) The inner products $N^{-1}\tilde{\boldsymbol{m}}_t^{\mathrm{T}}\tilde{\boldsymbol{m}}_{t'}$ and $N^{-1}\tilde{\boldsymbol{q}}_t^{\mathrm{T}}\tilde{\boldsymbol{q}}_{t'}$ converge almost surely to some constants $\pi_{t,t'} \in \mathbb{R}$ and $\kappa_{t,t'} \in \mathbb{R}$, respectively.
2) Suppose that $\tilde{\boldsymbol{\psi}}_t(\boldsymbol{H}_{t+1}, \boldsymbol{x}) : \mathbb{R}^{N\times(t+2)} \to \mathbb{R}^N$ is a separable and proper pseudo-Lipschitz function of order k, that $\tilde{\boldsymbol{\phi}}_t(\boldsymbol{B}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda}) : \mathbb{R}^{N\times(t+3)} \to \mathbb{R}^N$ is separable, pseudo-Lipschitz of order k with respect to the first $t+2$ variables, and proper, and that $\boldsymbol{Z}_{t+1} = (\boldsymbol{z}_0, \ldots, \boldsymbol{z}_t) \in \mathbb{R}^{N\times(t+1)}$ denotes a zero-mean Gaussian random matrix with covariance $\mathbb{E}[\boldsymbol{z}_\tau \boldsymbol{z}_{\tau'}^{\mathrm{T}}] = \pi_{\tau,\tau'}\boldsymbol{I}_N$ for all $\tau, \tau' = 0, \ldots, t$, while a zero-mean Gaussian random matrix $\tilde{\boldsymbol{Z}}_{t+1} = (\tilde{\boldsymbol{z}}_0, \ldots, \tilde{\boldsymbol{z}}_t) \in \mathbb{R}^{N\times(t+1)}$ has covariance $\mathbb{E}[\tilde{\boldsymbol{z}}_\tau \tilde{\boldsymbol{z}}_{\tau'}^{\mathrm{T}}] = \kappa_{\tau,\tau'}\boldsymbol{I}_N$. Then,
$$\langle \tilde{\boldsymbol{\psi}}_t(\boldsymbol{H}_{t+1}, \boldsymbol{x})\rangle - \mathbb{E}\left[\langle \tilde{\boldsymbol{\psi}}_t(\boldsymbol{Z}_{t+1}, \boldsymbol{x})\rangle\right] \overset{\mathrm{a.s.}}{\to} 0, \tag{26}$$
$$\langle \tilde{\boldsymbol{\phi}}_t(\boldsymbol{B}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle - \mathbb{E}\left[\langle \tilde{\boldsymbol{\phi}}_t(\tilde{\boldsymbol{Z}}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle\right] \overset{\mathrm{a.s.}}{\to} 0. \tag{27}$$
In evaluating the expectation in (27), $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{w}$ in (13) follows the zero-mean Gaussian distribution with covariance $\sigma^2\boldsymbol{I}_M$. In particular, for $k = 1$
$$\langle \partial_{t'}\tilde{\boldsymbol{\psi}}_t(\boldsymbol{H}_{t+1}, \boldsymbol{x})\rangle - \mathbb{E}\left[\langle \partial_{t'}\tilde{\boldsymbol{\psi}}_t(\boldsymbol{Z}_{t+1}, \boldsymbol{x})\rangle\right] \overset{\mathrm{a.s.}}{\to} 0, \tag{28}$$
$$\langle \partial_{t'}\tilde{\boldsymbol{\phi}}_t(\boldsymbol{B}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle - \mathbb{E}\left[\langle \partial_{t'}\tilde{\boldsymbol{\phi}}_t(\tilde{\boldsymbol{Z}}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle\right] \overset{\mathrm{a.s.}}{\to} 0. \tag{29}$$
3) Suppose that $\tilde{\boldsymbol{\psi}}_t(\boldsymbol{H}_{t+1}, \boldsymbol{x}) : \mathbb{R}^{N\times(t+2)} \to \mathbb{R}^N$ is separable and proper Lipschitz-continuous, and that $\tilde{\boldsymbol{\phi}}_t(\boldsymbol{B}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda}) : \mathbb{R}^{N\times(t+3)} \to \mathbb{R}^N$ is separable, Lipschitz-continuous with respect to the first $t+2$ variables, and proper. Then,
$$\frac{1}{N}\boldsymbol{h}_{t'}^{\mathrm{T}}\left(\tilde{\boldsymbol{\psi}}_t - \sum_{\tau=0}^{t}\left\langle \partial_\tau \tilde{\boldsymbol{\psi}}_t\right\rangle \boldsymbol{h}_\tau\right) \overset{\mathrm{a.s.}}{\to} 0, \tag{30}$$
$$\frac{1}{N}\boldsymbol{b}_{t'}^{\mathrm{T}}\left(\tilde{\boldsymbol{\phi}}_t - \sum_{\tau=0}^{t}\left\langle \partial_\tau \tilde{\boldsymbol{\phi}}_t\right\rangle \boldsymbol{b}_\tau\right) \overset{\mathrm{a.s.}}{\to} 0. \tag{31}$$

Proof: See Appendix A. ■

Properties (26) and (27) are used to evaluate the performance of MP by specifying the functions $\tilde{\boldsymbol{\psi}}_t$ and $\tilde{\boldsymbol{\phi}}_t$ according to a performance measure. An important observation is the asymptotic Gaussianity of $\boldsymbol{H}_{t+1}$ and $\boldsymbol{B}_{t+1}$. In evaluating the performance of MP, we can replace them with tractable Gaussian random matrices $\boldsymbol{Z}_{t+1}$ and $\tilde{\boldsymbol{Z}}_{t+1}$.

The asymptotic Gaussianity originates from the definitions of $\tilde{\boldsymbol{q}}_t$ and $\tilde{\boldsymbol{m}}_t$ in (9) and (11). Properties (30) and (31) imply the asymptotic orthogonality $N^{-1}\boldsymbol{h}_{t'}^{\mathrm{T}}\tilde{\boldsymbol{q}}_{t+1} \overset{\mathrm{a.s.}}{\to} 0$ and $N^{-1}\boldsymbol{b}_{t'}^{\mathrm{T}}\tilde{\boldsymbol{m}}_t \overset{\mathrm{a.s.}}{\to} 0$. This orthogonality is used to prove that the distributions of $\boldsymbol{H}_{t+1}$ and $\boldsymbol{B}_{t+1}$ are asymptotically Gaussian.

Properties (30) and (31) can be regarded as computation formulas to evaluate $N^{-1}\boldsymbol{h}_{t'}^{\mathrm{T}}\tilde{\boldsymbol{\psi}}_t$ and $N^{-1}\boldsymbol{b}_{t'}^{\mathrm{T}}\tilde{\boldsymbol{\phi}}_t$. They can be computed via linear combinations of $\{N^{-1}\boldsymbol{h}_{t'}^{\mathrm{T}}\boldsymbol{h}_\tau\}_{\tau=0}^{t}$ and $\{N^{-1}\boldsymbol{b}_{t'}^{\mathrm{T}}\boldsymbol{b}_\tau\}_{\tau=0}^{t}$. In particular, (9), (11), and Property 1) in Theorem 1 imply $N^{-1}\boldsymbol{h}_{t'}^{\mathrm{T}}\boldsymbol{h}_\tau \overset{\mathrm{a.s.}}{\to} \pi_{t',\tau}$ and $N^{-1}\boldsymbol{b}_{t'}^{\mathrm{T}}\boldsymbol{b}_\tau \overset{\mathrm{a.s.}}{\to} \kappa_{t',\tau}$. Furthermore, the coefficients in the linear combinations can be computed with (28) and (29). From these observations, the SE equations of the general error model are given as dynamical systems with respect to $\{\pi_{t,t'}, \kappa_{t,t'}\}$ in general.

We do not derive SE equations with respect to $\{\pi_{t,t'}, \kappa_{t,t'}\}$ in a general form. Instead, we derive SE equations after specifying MP. The usefulness of Theorem 1 is clarified in deriving SE equations.

III. Signal Recovery

A. Convolutional Approximate Message-Passing

Let $\boldsymbol{x}_t \in \mathbb{R}^N$ denote an estimator of the signal vector $\boldsymbol{x}$ in iteration t. CAMP computes the estimator $\boldsymbol{x}_t$ recursively as

$$\boldsymbol{x}_{t+1} = f_t(\boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t), \tag{32}$$

$$\boldsymbol{z}_t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_t + \sum_{\tau=0}^{t-1}\xi_\tau^{(t-1)}(\theta_{t-\tau}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} - g_{t-\tau}\boldsymbol{I}_M)\boldsymbol{z}_\tau, \tag{33}$$

with the initial condition $\boldsymbol{x}_0 = \boldsymbol{0}$, where $\xi_\tau^{(t-1)} = \prod_{t'=\tau}^{(t-1)}\xi_{t'}$ is the product of $\{\xi_{t'}\}$ given by

$$\xi_t = \left\langle f_t'(\boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t)\right\rangle. \tag{34}$$

In (32) and (33), $\boldsymbol{A}$ and $\boldsymbol{y}$ are the sensing matrix and the measurement vector in (1), respectively. The functions $\{f_t : \mathbb{R} \to \mathbb{R}\}$ are a sequence of Lipschitz-continuous denoisers. The tap coefficients $\{g_\tau \in \mathbb{R}\}$ and $\{\theta_\tau \in \mathbb{R}\}$ in the convolution are design parameters. The parameters $\{\theta_\tau\}$ are optimized to improve the performance of the CAMP while $\{g_\tau\}$ are determined so as to realize the asymptotic Gaussianity of the estimation errors via Theorem 1.

To impose the initial condition $\boldsymbol{x}_0 = \boldsymbol{0}$, it is convenient to introduce the notational convention $f_{-1}(\cdot) = 0$, which is used throughout this paper.

The CAMP is a generalization of AMP [26] and reduces to AMP when $g_1 = -\delta^{-1}$, $g_\tau = 0$ for $\tau > 1$, and $\theta_\tau = 0$ hold. Also, as a generalization of CAMP in [53], the affine transform $(\theta_{t-\tau}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} - g_{t-\tau}\boldsymbol{I}_M)\boldsymbol{z}_\tau$ has been applied before the convolution. Nonetheless, the proposed MP is called CAMP simply. In particular, the MP algorithm reduces to the original CAMP [53] when $\theta_\tau = 0$ is assumed.

Remark 1: The design parameters $\{\theta_\tau\}$ are not required and can be set to zero for sensing matrices with identical non-zero singular values since $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ reduces to the identity matrix with the exception of a constant factor. Thus, non-zero parameters $\{\theta_\tau\}$ should be introduced only for the case of non-identical singular values.

B. Error Model

To design the parameters g_τ and θ_τ via Theorem 1, we derive an error model of the CAMP. Let $\boldsymbol{h}_t = \boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t - \boldsymbol{x}$ and $\boldsymbol{q}_{t+1} = \boldsymbol{x}_{t+1} - \boldsymbol{x}$ denote the error vectors before and after denoising f_t, respectively. Then, we have

$$\boldsymbol{q}_{t+1} = f_t(\boldsymbol{x}+\boldsymbol{h}_t) - \boldsymbol{x} \equiv \psi_t(\boldsymbol{h}_t, \boldsymbol{x}), \tag{35}$$

$$\tilde{\boldsymbol{q}}_{t+1} = \boldsymbol{q}_{t+1} - \xi_t \boldsymbol{h}_t. \tag{36}$$

Using the notational convention $f_{-1}(\cdot) = 0$, we obtain the initial condition $\boldsymbol{q}_0 = -\boldsymbol{x}$ imposed in the general error model.

We define $\boldsymbol{m}_t = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{h}_t$ and $\boldsymbol{b}_t = \boldsymbol{V}^{\mathrm{T}}\tilde{\boldsymbol{q}}_t$ to formulate the error model of the CAMP in a form corresponding to the general error model (9)–(12). Substituting the definition $\boldsymbol{h}_t = \boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t - \boldsymbol{x}$ into $\boldsymbol{m}_t = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{h}_t$ yields

$$\boldsymbol{m}_t = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{q}_t + \boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{z}_t, \tag{37}$$

where we have used the definition $\boldsymbol{q}_t = \boldsymbol{x}_t - \boldsymbol{x}$ and the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$. We utilize the definitions (36), $\boldsymbol{b}_t = \boldsymbol{V}^{\mathrm{T}}\tilde{\boldsymbol{q}}_t$, and $\boldsymbol{m}_t = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{h}_t$ to obtain

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{q}_t = \boldsymbol{b}_t + \xi_{t-1}\boldsymbol{m}_{t-1}. \tag{38}$$

Combining these two equations yields

$$\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{z}_t = \boldsymbol{m}_t - \boldsymbol{b}_t - \xi_{t-1}\boldsymbol{m}_{t-1}. \tag{39}$$

To obtain a closed-form equation with respect to $\boldsymbol{m}_t$, we left-multiply (33) by $\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}$ and use (1) to have

$$\begin{aligned}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{z}_t &= -\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{q}_t + \boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{w} \\ &\quad + \sum_{\tau=0}^{t-1}\xi_\tau^{(t-1)}(\theta_{t-\tau}\boldsymbol{\Lambda} - g_{t-\tau}\boldsymbol{I}_M)\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{z}_\tau,\end{aligned} \tag{40}$$

with $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma}$. Substituting (38) and (39) into this expression, we arrive at

$$\begin{aligned}\boldsymbol{m}_t = (\boldsymbol{I}_N - \boldsymbol{\Lambda}) &\quad (\boldsymbol{b}_t + \xi_{t-1}\boldsymbol{m}_{t-1}) + \boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{w} \\ + \sum_{\tau=0}^{t-1} &\quad \xi_\tau^{(t-1)}(\theta_{t-\tau}\boldsymbol{\Lambda} - g_{t-\tau}\boldsymbol{I}_M) \\ &\quad \cdot(\boldsymbol{m}_\tau - \boldsymbol{b}_\tau - \xi_{\tau-1}\boldsymbol{m}_{\tau-1}),\end{aligned} \tag{41}$$

where any vector with a negative index is set to zero. This expression implies that $\boldsymbol{\phi}_t$ for the CAMP depends on all messages $\boldsymbol{B}_{t+1}$.

We note that Assumption 4 holds under Assumption 2 since the denoiser f_t has been assumed to be Lipschitz-continuous.

C. Asymptotic Gaussianity

We compare the obtained error model with the general error model (9)–(12). The only difference is in (11): The correction $\tilde{\boldsymbol{m}}_t$ of $\boldsymbol{m}_t$ is used to define $\boldsymbol{h}_t$ in the general error model while no correction is performed in the error model of the CAMP. Thus, the general error model contains the error model of the CAMP when $\langle \partial_{t'}\boldsymbol{m}_t\rangle = 0$ holds for all $t' = 0, \ldots, t$. In the CAMP, the parameters $\{g_\tau\}$ are determined so as to guarantee $\langle \partial_{t'}\boldsymbol{m}_t\rangle = 0$ in the large system limit.

Let μ_j denote the jth moment of the asymptotic eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, given by

$$\mu_j = \lim_{M=\delta N\to\infty}\frac{1}{N}\mathrm{Tr}(\boldsymbol{\Lambda}^j). \tag{42}$$

Assumption 2 implies $\mu_1 = 1$. We define a coupled dynamical system $\{g_\tau^{(j)}\}$ determined via the tap coefficients $\{g_\tau\}$ and $\{\theta_\tau\}$ as

$$g_0^{(j)} = \mu_{j+1} - \mu_j, \tag{43}$$

$$\begin{aligned}g_1^{(j)} &= g_0^{(j)} - g_0^{(j+1)} - g_1(g_0^{(j)} + \mu_j) \\ &\quad + \theta_1(g_0^{(j+1)} + \mu_{j+1}),\end{aligned} \tag{44}$$

$$\begin{aligned}g_\tau^{(j)} &= g_{\tau-1}^{(j)} - g_{\tau-1}^{(j+1)} - g_\tau\mu_j + \theta_\tau\mu_{j+1} \\ &\quad + \sum_{\tau'=0}^{\tau-1}(\theta_{\tau-\tau'}g_{\tau'}^{(j+1)} - g_{\tau-\tau'}g_{\tau'}^{(j)}) \\ &\quad - \sum_{\tau'=1}^{\tau-1}(\theta_{\tau-\tau'}g_{\tau'-1}^{(j+1)} - g_{\tau-\tau'}g_{\tau'-1}^{(j)})\end{aligned} \tag{45}$$

for $\tau > 1$.

Theorem 2: Suppose that Assumptions 1–3 hold, that the denoiser f_t is Lipschitz-continuous, and that the tap coefficients $\{g_\tau\}$ and $\{\theta_\tau\}$ in the CAMP satisfy

$$g_1 = \theta_1(g_0^{(1)} + 1) - g_0^{(1)}, \tag{46}$$

$$g_\tau = \theta_\tau - g_{\tau-1}^{(1)} + \sum_{\tau'=0}^{\tau-1}\theta_{\tau-\tau'}g_{\tau'}^{(1)} - \sum_{\tau'=1}^{\tau-1}\theta_{\tau-\tau'}g_{\tau'-1}^{(1)}, \tag{47}$$

where $\{g_\tau^{(1)}\}$ is governed by the dynamical system (43)–(45). Then, $\langle \partial_{t'}\boldsymbol{m}_t\rangle \to 0$ holds in the large system limit, i.e. the error model of the CAMP is included into the general error model.

Proof: Let

$$g_{t',t}^{(j)} = -\lim_{M=\delta N\to\infty}\left\langle \boldsymbol{\Lambda}^j\partial_{t'}\boldsymbol{m}_t\right\rangle. \tag{48}$$

It is sufficient to prove $g_{t',t}^{(j)} \overset{\text{a.s.}}{=} \xi_{t'}^{(t-1)}g_{t-t'}^{(j)} + o(1)$ and $g_\tau^{(0)} = 0$ under the notational convention $\xi_{t'}^{(t)} = 1$ for $t' > t$. The latter property $g_\tau^{(0)} = 0$ follows from (43) for $\tau = 0$, (44) and (46) for $\tau = 1$, and from (45) and (47). See Appendix B for the proof of the former property. ∎

Throughout this paper, we assume that the tap coefficients $\{g_\tau\}$ and $\{\theta_\tau\}$ satisfy (46) and (47). Thus, Theorem 1 implies that the asymptotic Gaussianity is guaranteed for the CAMP. In principle, it is possible to compute the tap coefficients by solving the coupled dynamical system (43)–(47) numerically for a given moment sequence $\{\mu_j\}$. However, numerical evaluation indicated that the dynamical system is unstable against numerical errors when the moment sequence $\{\mu_j\}$ is a diverging sequence. Thus, we need a closed-form solution to the tap coefficients.

To present the closed-form solution, we define the η-transform of the asymptotic eigenvalue distribution of

$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ [17] as

$$\eta(x) = \lim_{M=\delta N\to\infty} \frac{1}{N}\mathrm{Tr}\left\{\left(\boldsymbol{I}_N + x\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\right\}. \tag{49}$$

By definition, we have the power-series expansion

$$\eta(x) = \lim_{M=\delta N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\frac{1}{1+x\lambda_n} = \sum_{j=0}^{\infty}\mu_j(-x)^j \tag{50}$$

for $|x| < 1/\max\{\lambda_n\}$. Let $G(z)$ denote the generating function of the tap coefficients $\{g_\tau\}$ given by

$$G(z) = \sum_{\tau=0}^{\infty} g_\tau z^{-\tau}, \quad g_0 = 1. \tag{51}$$

Similarly, we write the generating function of $\{\theta_\tau\}$ with $\theta_0 = 1$ as $\Theta(z)$.

Theorem 3: Suppose that the tap coefficients $\{g_\tau\}$ and $\{\theta_\tau\}$ satisfy (46) and (47). Then, the generating functions $G(z)$ and $\Theta(z)$ of $\{g_\tau\}$ and $\{\theta_\tau\}$ satisfy

$$\eta\left(\frac{1-(1-z^{-1})\Theta(z)}{(1-z^{-1})G(z)}\right) = (1-z^{-1})\Theta(z), \tag{52}$$

where η denotes the η-transform of the asymptotic eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

Proof: See Appendix C. ∎

Suppose that the η-transform is given. Since the η-transform has the inverse function, from Theorem 3 we have $(1-z^{-1})G(z) = [1-(1-z^{-1})\Theta(z)]/\eta^{-1}((1-z^{-1})\Theta(z))$ for a fixed generating function $\Theta(z)$. Each tap coefficient g_τ can be computed by evaluating the coefficient of the τth-order term in $G(z)$.

Corollary 1: Suppose that the sensing matrix $\boldsymbol{A}$ has independent Gaussian elements with mean $\sqrt{\gamma/M}$ and variance $(1-\gamma)/M$ for any $\gamma \in [0,1)$. Then, the tap coefficient g_t is given by

$$g_t = \left(1-\frac{1}{\delta}\right)\theta_t + \frac{1}{\delta}\sum_{\tau=0}^{t}(\theta_\tau - \theta_{\tau-1})\theta_{t-\tau} \tag{53}$$

for fixed tap coefficients $\{\theta_t\}$.

Proof: We shall evaluate the generating function $G(z)$. The R-transform $R(x)$ [17, Section 2.4.2] of the asymptotic eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is given by

$$R(x) = \frac{\delta}{\delta - x}. \tag{54}$$

Using Theorem 3 and the relationship between the R-transform and the η-transform [17, Eq. (2.74)]

$$\eta(x) = \frac{1}{1+xR(-x\eta(x))}, \tag{55}$$

we obtain

$$G(z) = \left[1-\frac{1}{\delta}+\frac{(1-z^{-1})}{\delta}\Theta(z)\right]\Theta(z), \tag{56}$$

which implies the time-domain expression (53). ∎

In particular, we consider the original CAMP $\theta_\tau = 0$ for $\tau > 0$. In this case, we have $g_1 = -\delta^{-1}$ and $g_\tau = 0$. As remarked in [53], the original CAMP reduces to the AMP for the i.i.d. Gaussian sensing matrix.

Corollary 2: Suppose that the sensing matrix $\boldsymbol{A}$ has M identical non-zero singular values for $M \leq N$, i.e. $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \delta^{-1}\boldsymbol{I}_M$. Then, the tap coefficient g_t in the original CAMP $\theta_t = 0$ for $t > 0$ is given by $g_\tau = 1-\delta^{-1}$ for all $\tau \geq 1$.

Proof: We evaluate the generating function $G(z)$. By definition, the η-transform is given by

$$\eta(x) = \frac{1}{N}\left(\frac{M}{1+x\delta^{-1}} + N - M\right) = 1-\delta+\frac{\delta^2}{\delta+x}. \tag{57}$$

Using Theorem 3 and $\Theta(z) = 1$ yields

$$G(z) = \frac{1-\delta^{-1}z^{-1}}{1-z^{-1}} = 1+\sum_{j=1}^{\infty}\left(1-\frac{1}{\delta}\right)z^{-j}, \tag{58}$$

which implies $g_\tau = 1-\delta^{-1}$ for all $\tau \geq 1$. ∎

Corollary 3: Suppose that the sensing matrix $\boldsymbol{A}$ has non-zero singular values $\sigma_0 \geq \cdots \geq \sigma_{M-1} > 0$ satisfying condition number $\kappa = \sigma_0/\sigma_{M-1} > 1$, $\sigma_m/\sigma_{m-1} = \kappa^{-1/(M-1)}$, and $\sigma_0^2 = N(1-\kappa^{-2/(M-1)})/(1-\kappa^{-2M/(M-1)})$. Assume $\theta_t = 0$ for all $t > t_1$ for some $t_1 \in \mathbb{N}$. Let $\alpha_0^{(j)} = 1$ and

$$\alpha_t^{(j)} = \begin{cases} \frac{C^{t/j}}{(t/j)!}\bar{\theta}_j^{t/j} & \text{if } t \text{ is divisible by } j, \\ 0 & \text{otherwise} \end{cases} \tag{59}$$

for $t \in \mathbb{N}$ and $j \in \{1,\ldots,t_1\}$, with $\bar{\theta}_t = \theta_{t-1} - \theta_t$ and $C = 2\delta^{-1}\ln\kappa$. Define $p_0 = \bar{q}_0 = 1$ and

$$p_t = -\frac{\beta_t^{(t_1)}}{\kappa^2-1}, \tag{60}$$

$$\bar{q}_t = \frac{1}{\bar{\theta}_1}\left(\frac{\beta_{t+1}^{(t_1)}}{C} - \sum_{\tau=1}^{t_1}\bar{\theta}_{\tau+1}\bar{q}_{t-\tau}\right) \tag{61}$$

for $t > 0$, with $\beta_t^{(t_1)} = \alpha_t^{(1)} * \alpha_t^{(2)} * \cdots * \alpha_t^{(t_1)}$. Then, the tap coefficient g_t is recursively given by

$$g_t = p_t - \sum_{\tau=1}^{t} q_\tau g_{t-\tau}, \tag{62}$$

with

$$q_t = \bar{q}_t - \bar{q}_{t-1}. \tag{63}$$

Proof: We first evaluate the inverse of the η-transform. By definition, $\sigma_m^2 = \kappa^{-2m/(M-1)}\sigma_0^2$ holds. Thus, we have

$$\mu_j = \frac{1}{N}\sum_{m=0}^{M-1}\sigma_m^{2j} = \sigma_0^{2j}\frac{1-\kappa^{-2jM/(M-1)}}{N(1-\kappa^{-2j/(M-1)})} \to \left(\frac{C}{1-\kappa^{-2}}\right)^j\frac{1-\kappa^{-2j}}{Cj} \tag{64}$$

in the large system limit, where we have used the convergence $N(1-\kappa^{-a/(M-1)}) \to \delta^{-1}a\ln\kappa$ for any $a \in \mathbb{R}$. We note the series-expansion $\ln(1+x) = \sum_{j=1}^{\infty}(-1)^{j-1}j^{-1}x^j$ for $|x| < 1$ to obtain

$$\eta(x) = 1+\sum_{j=1}^{\infty}(-x)^j\mu_j = 1-\frac{1}{C}\ln\left(\frac{\kappa^2-1+\kappa^2Cx}{\kappa^2-1+Cx}\right), \tag{65}$$

which implies the inverse function

$$\eta^{-1}(x) = \frac{(\kappa^2-1)\{e^{C(1-x)}-1\}}{C\{\kappa^2-e^{C(1-x)}\}}. \tag{66}$$

We next evaluate the generating function $G(z)$. Using Theorem 3 yields $G(z) = P(z)/Q(z)$, with

$$P(z) = \frac{\kappa^2 - e^{C\bar{\Theta}(z)}}{\kappa^2 - 1}, \tag{67}$$

$$Q(z) = (1 - z^{-1})\bar{Q}(z), \quad \bar{Q}(z) = \frac{e^{C\bar{\Theta}(z)} - 1}{C\bar{\Theta}(z)}, \tag{68}$$

$$\bar{\Theta}(z) = \sum_{t=1}^{\infty} \bar{\theta}_t z^{-t}. \tag{69}$$

Finally, we derive a time-domain expression of $G(z)$. It is an exercise to confirm that the series-expansions of $P(z)$ and $\bar{Q}(z)$ have the coefficients p_t and $\bar{q}_t$ for the tth-order terms, respectively. Then, the Z-transform of (62) is equal to $P(z)/Q(z)$. ∎

The sequences $\{p_\tau\}$ and $\{q_\tau\}$ in Corollary 3 define the generating functions $P(z)$ and $Q(z)$ with $p_0 = q_0 = 1$, respectively, which satisfy $G(z) = P(z)/Q(z)$. Thus, we derive an SE equation in time domain in terms of $\{p_\tau, q_\tau\}$, rather than $\{g_\tau\}$.

D. SE Equation

We design the tap coefficients $\{\theta_\tau\}$ so as to minimize the MSE $N^{-1}\|\boldsymbol{x}_t - \boldsymbol{x}\|^2$ for the CAMP estimator $\boldsymbol{x}_t$ in the large system limit. For that purpose, we derive an SE equation that describes the dynamics of the MSE. For simplicity, we assume i.i.d. signals.

The CAMP has no closed-form SE equation with respect to the MSEs $N^{-1}\|\boldsymbol{x}_t - \boldsymbol{x}\|^2$ in general. Instead, it has a closed-form SE equation with respect to the correlations

$$d_{t'+1,t+1} = \mathbb{E}\left[\{f_{t'}(x_1 + z_{t'}) - x_1\}\{f_t(x_1 + z_t) - x_1\}\right], \tag{70}$$

where $\{z_t\}$ denote zero-mean Gaussian random variables with covariance $a_{t',t} = \mathbb{E}[z_{t'} z_t]$. In particular, $d_{t+1,t+1}$ corresponds to the MSE of the CAMP estimator in iteration t.

As an asymptotic alternative to ξ_t, we use the following quantity:

$$\bar{\xi}_t = \mathbb{E}\left[f_t'(x_1 + z_t)\right], \tag{71}$$

which is a function of $a_{t,t}$. The notation $\bar{\xi}_{t'}^{(t)}$ is defined in the same manner as in $\xi_{t'}^{(t)}$.

Theorem 4: Assume that Assumptions 1–3 hold, that the denoiser f_t is Lipschitz-continuous, and that the signal vector $\boldsymbol{x}$ has i.i.d. elements. Suppose that the generating functions G and Θ for the tap coefficients $\{g_\tau\}$ and $\{\theta_\tau\}$—given in (51)—satisfy the condition (52) in Theorem 3.

- Define generating functions $A(y,z)$, $D(y,z)$, and $\Sigma(y,z)$ as

$$A(y,z) = \sum_{t',t=0}^{\infty} \frac{a_{t',t}}{\bar{\xi}_0^{(t'-1)} \bar{\xi}_0^{(t-1)}} y^{-t'} z^{-t}, \tag{72}$$

$$D(y,z) = \sum_{t',t=0}^{\infty} \frac{d_{t',t}}{\bar{\xi}_0^{(t'-1)} \bar{\xi}_0^{(t-1)}} y^{-t'} z^{-t}, \tag{73}$$

$$\Sigma(y,z) = \sum_{t',t=0}^{\infty} \frac{\sigma^2}{\bar{\xi}_0^{(t'-1)} \bar{\xi}_0^{(t-1)}} y^{-t'} z^{-t}. \tag{74}$$

  Then, the correlation $N^{-1}(\boldsymbol{x}_{t'} - \boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}_t - \boldsymbol{x})$ converges almost surely to $d_{t',t}$ in the large system limit, which satisfies the following SE equation in terms of the generating functions:

$$\begin{aligned} F_{G,\Theta}(y,z) A(y,z) = \; & \{G(z)\Delta_\Theta - \Theta(z)\Delta_G\} D(y,z) \\ & + (\Delta_{\Theta_1} - \Delta_\Theta) \Sigma(y,z), \end{aligned} \tag{75}$$

  with

$$\begin{aligned} F_{G,\Theta}(y,z) = \; & (y^{-1} + z^{-1} - 1)[G(z)\Delta_\Theta - \Theta(z)\Delta_G] \\ & + \Delta_{G_1} - \Delta_G, \end{aligned} \tag{76}$$

  where the notations $S_1(z) = z^{-1}S(z)$ and $\Delta_S = [S(y) - S(z)]/(y^{-1} - z^{-1})$ have been used for any generating function $S(z)$.
- Suppose that $G(z)$ is represented as $G(z) = P(z)/Q(z)$ for the generating functions $P(z)$ and $Q(z)$ of some sequences $\{p_\tau\}$ and $\{q_\tau\}$ with $p_0 = 1$ and $q_0 = 1$. Let $r_t = q_t * \theta_t$. Then, the SE equation (75) reduces to

$$\begin{aligned} \sum_{\tau'=0}^{t'} \sum_{\tau=0}^{t} \bar{\xi}_{t'-\tau'}^{(t'-1)} \bar{\xi}_{t-\tau}^{(t-1)} \Big\{ & \mathfrak{D}_{\tau',\tau} a_{t'-\tau',t-\tau} \\ & - (p_\tau * r_{\tau'+\tau+1} - r_\tau * p_{\tau'+\tau+1}) d_{t'-\tau',t-\tau} \\ & - \sigma^2 \left[(q_{\tau'} q_\tau) * (\theta_{\tau'+\tau} - \theta_{\tau'+\tau+1})\right] \Big\} = 0, \end{aligned} \tag{77}$$

  where all variables with negative indices are set to zero, with

$$\begin{aligned} \mathfrak{D}_{\tau',\tau} = \; & (p_{\tau'+\tau} - p_{\tau'+\tau+1}) * q_\tau + (p_\tau - p_{\tau-1}) * q_{\tau'+\tau+1} \\ + \; & (p_{\tau-1} - p_\tau) * r_{\tau'+\tau+1} + (r_\tau - r_{\tau-1}) * p_{\tau'+\tau+1} \\ + \; & p_\tau * (r_{\tau'+\tau} - \delta_{\tau',0} r_\tau) - r_\tau * (p_{\tau'+\tau} - \delta_{\tau',0} p_\tau). \end{aligned} \tag{78}$$

  In solving the SE equation (77), we impose the initial condition $d_{0,0} = 1$ and boundary conditions $d_{0,\tau+1} = d_{\tau+1,0} = -\mathbb{E}[x_1\{f_\tau(x_1 + z_\tau) - x_1\}]$ for any τ.

Proof: See Appendix D. ∎

The SE equation (77) in time domain is useful for numerical evaluation of $\{a_{t',t}\}$ while the generating-function representation (75) is used in fixed-point analysis. To apply Corollary 3, we have represented the generating function $G(z)$ as $G(z) = P(z)/Q(z)$. If $G(z)$ is given directly, the functions $P(z) = G(z)$ and $Q(z) = 1$ can be used. In this case, we have $p_\tau = g_\tau$, $q_\tau = \delta_{\tau,0}$, and $r_\tau = \theta_\tau$.

Note that $d_{t'+1,t+1}$ given in (70) is a function of $\{a_{t',t}, a_{t',t'}, a_{t,t}\}$, so that the SE equation (77) in time domain is a nonlinear difference equation with respect to $\{a_{t',t}\}$ for given tap coefficients $\{g_\tau\}$ and $\{\theta_\tau\}$. Theorem 4 allows us to compute the MSEs $a_{t,t}$ and $d_{t+1,t+1}$ before and after denoising.

The SE equation (77) in time domain can be solved recursively by extracting the term $\mathfrak{D}_{0,0} a_{t',t}$ for $\tau' = \tau = 0$ in the sum and moving the other terms to the RHS. More precisely, we can solve the SE equation (77) as follows:

1) Let $t = 0$ and solve $a_{0,0}$ with the SE equation (77) and the initial condition $d_{0,0} = 1$.
2) Suppose that $\{a_{\tau',\tau}, d_{\tau',\tau}\}$ have been obtained for all $\tau', \tau = 0, \dots, t$. Use the boundary condition $d_{0,t+1}$ in

Theorem 4 and compute $d_{\tau,t+1}$ with the definition (70) for all $\tau = 1, \dots, t+1$ while the symmetry $d_{t+1,\tau} = d_{\tau,t+1}$ is used in the lower triangular elements.

3) Compute $a_{\tau,t+1}$ with the SE equation (77) in the order $\tau = 0, \dots, t+1$ while the symmetry $a_{t+1,\tau} = a_{\tau,t+1}$ is used in the upper triangular elements.

4) If some termination conditions are satisfied, output $\{a_{\tau',\tau}, d_{\tau',\tau}\}$. Otherwise, update $t := t+1$ and go back to Step 2).

We can define the Bayes-optimal denoiser f_t via the MSE $d_{t+1,t+1}$ in the large system limit. A denoiser f_t is said to be Bayes-optimal if $f_t = \mathbb{E}[x_1|x_1+z_t]$ is the posterior mean of x_1 given an AWGN observation $x_1 + z_t$ with $z_t \sim \mathcal{N}(0, a_{t,t})$. We write the Bayes-optimal denoiser as $f_t(\cdot) = f_{\mathrm{opt}}(\cdot; a_{t,t})$.

The boundary condition $d_{0,\tau+1}$ in Theorem 4 has a simple representation for the Bayes-optimal denoiser f_{opt}. Since the posterior mean estimator $f_{\mathrm{opt}}(x_1 + z_\tau; a_{\tau,\tau})$ is uncorrelated with the estimation error $f_{\mathrm{opt}}(x_1 + z_\tau; a_{\tau,\tau}) - x_1$, we obtain

$$\begin{aligned} d_{0,\tau+1} = & \ \mathbb{E}[\{f_{\mathrm{opt}}(x_1+z_\tau; a_{\tau,\tau}) - x_1 - f_{\mathrm{opt}}(x_1+z_\tau; a_{\tau,\tau})\} \\ & \cdot \{f_{\mathrm{opt}}(x_1+z_\tau; a_{\tau,\tau}) - x_1\}] \\ = & \ \mathbb{E}[\{f_{\mathrm{opt}}(x_1+z_\tau; a_{\tau,\tau}) - x_1\}^2] = d_{\tau+1,\tau+1}. \end{aligned} \tag{79}$$

Theorem 5: Consider the Bayes-optimal denoiser under the same assumptions as in Theorem 4. Suppose that the SE equation (77) in time domain converges to a fixed-point $\{a_{\mathrm{s}}, d_{\mathrm{s}}\}$, i.e. $\lim_{t',t\to\infty} a_{t',t} = a_{\mathrm{s}}$ and $\lim_{t',t\to\infty} d_{t',t} = d_{\mathrm{s}}$. If $\Theta(\xi_{\mathrm{s}}^{-1}) = 1$ and $1 + (\xi_{\mathrm{s}} - 1)d\Theta(\xi_{\mathrm{s}}^{-1})/(dz^{-1}) \neq 0$ hold for $\xi_{\mathrm{s}} = d_{\mathrm{s}}/a_{\mathrm{s}}$, then the fixed-point $\{a_{\mathrm{s}}, d_{\mathrm{s}}\}$ of the SE equation (77) satisfies

$$a_{\mathrm{s}} = \frac{\sigma^2}{R(-d_{\mathrm{s}}/\sigma^2)}, \quad d_{\mathrm{s}} = \mathbb{E}\left[\{f_{\mathrm{opt}}(x_1 + z_{\mathrm{s}}; a_{\mathrm{s}}) - x_1\}^2\right], \tag{80}$$

with $z_{\mathrm{s}} \sim \mathcal{N}(0, a_{\mathrm{s}})$, where $R(x)$ denotes the R-transform of the asymptotic eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

Proof: See Appendix E. ■

The fixed-point equations given in (80) coincide with those for describing the asymptotic performance of the posterior mean estimator of the signal vector $\boldsymbol{x}$ [18]–[20]. This coincidence implies that the CAMP with Bayes-optimal denoisers is Bayes-optimal if the SE equation (77) converges toward a fixed-point and if the fixed-point is unique. Thus, we refer to CAMP with Bayes-optimal denoisers as Bayes-optimal CAMP.

E. Implementation

We summarize the implementation of the Bayes-optimal CAMP. We need to specify the sequence of denoisers $\{f_t\}$ and the tap coefficients $\{g_\tau, \theta_\tau\}$ in (32) and (33). For simplicity, assume $\theta_\tau = 0$ for all $\tau > 2$. To impose the condition $\Theta(a_{\mathrm{s}}/d_{\mathrm{s}}) = 1$ in Theorem 5, we use $\theta_0 = 1$, $\theta_1 = -\theta d_{\mathrm{s}}/a_{\mathrm{s}}$, and $\theta_2 = \theta \in \mathbb{R}$, in which a_{s} and d_{s} are a solution to the fixed-point equations (80). In particular, the CAMP reduces to the original one in [53] for $\theta = 0$.

For a given parameter θ, the tap coefficients $\{g_\tau\}$ are determined via Theorem 3. More precisely, we use the coefficients $\{p_\tau, q_\tau\}$ in the rational generating function $G(z) = P(z)/Q(z)$. See Corollaries 1–3 for examples of the coefficients.

TABLE I

COMPLEXITY IN $M \leq N$ AND THE NUMBER OF ITERATIONS t

	Time complexity	Space complexity
CAMP	$\mathcal{O}(tMN + t^2M + t^4)$	$\mathcal{O}(MN + tM + t^2)$
AMP	$\mathcal{O}(tMN)$	$\mathcal{O}(MN)$
OAMP/VAMP	$\mathcal{O}(M^2N + tMN)$	$\mathcal{O}(N^2 + MN)$

For given parameters $\{\theta, p_\tau, q_\tau\}$, we can solve the SE equation (77) numerically. The obtained parameter $a_{t,t}$ is used to determine the Bayes-optimal denoiser $f_t(\cdot) = f_{\mathrm{opt}}(\cdot; a_{t,t})$.

Damping [39] is a well-known technique to improve the convergence property in finite-sized systems. In damped CAMP, the update rule (32) is replaced by

$$\boldsymbol{x}_{t+1} = \zeta f_t(\boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t) + (1-\zeta)\boldsymbol{x}_t, \tag{81}$$

with damping factor $\zeta \in [0, 1]$. In solving the SE equation (77), the associated parameters $d_{t'+1,t+1}$ and $\bar{\xi}_t$ in (70) and (71) are damped as follows:

$$\begin{aligned} d_{t'+1,t+1} = & \zeta\mathbb{E}\left[\{f_{t'}(x_1 + z_{t'}) - x_1\}\{f_t(x_1 + z_t) - x_1\}\right] \\ & + (1-\zeta)d_{t',t}, \end{aligned} \tag{82}$$

$$\bar{\xi}_t = \zeta\mathbb{E}\left[f_t'(x_1 + z_t)\right] + (1-\zeta)\bar{\xi}_{t-1}. \tag{83}$$

In particular, no damping is applied for $\zeta = 1$.

Table I lists time and space complexity of the CAMP, AMP, and OAMP/VAMP. Let t denote the number of iterations. We assume that the scalar parameters in the CAMP can be computed in $\mathcal{O}(t^4)$ time. In particular, computation of $\{a_{t,t}\}$ via the SE equation (77) is dominant.

To compute the update rule (33) in the CAMP efficiently, the vectors $\boldsymbol{z}_t \in \mathbb{R}^M$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t \in \mathbb{R}^M$ are computed and stored in iteration t. We need $\mathcal{O}(MN)$ space complexity to store the sensing matrix $\boldsymbol{A}$, which is dominant for the case $t \ll N$. Furthermore, the time complexity is dominated by matrix-vector multiplications.

In the OAMP/VAMP, the SVD of $\boldsymbol{A}$ requires dominant complexity unless the sensing matrix has a special structure that enables efficient SVD computation. As a result, the OAMP/VAMP has higher complexity than the AMP and CAMP while the CAMP has comparable complexity to the AMP for $t \ll N$.

IV. Numerical Results

A. Simulation Conditions

The Bayes-optimal CAMP—called CAMP simply—is compared to the AMP and OAMP/VAMP. In all numerical results, 10^5 independent trials were simulated. We assumed the AWGN noise $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_M)$ and i.i.d. Bernoulli-Gaussian signals with signal density $\rho \in [0, 1]$ in the measurement model (1). The probability density function (pdf) of x_n is given by

$$p(x_n) = (1-\rho)\delta(x_n) + \frac{\rho}{\sqrt{2\pi/\rho}}e^{-\frac{x_n^2}{2/\rho}}. \tag{84}$$

Since x_n has zero mean and unit variance, the signal-to-noise ratio (SNR) is equal to $1/\sigma^2$. See Appendix F for evaluation of the correlation $d_{t'+1,t+1}$ given in (70).

Corollary 3 was used to simulate ill-conditioned sensing matrices $\boldsymbol{A}$. The non-zero singular values $\{\sigma_m\}$ of $\boldsymbol{A}$ are uniquely determined via the condition number κ. To reduce the complexity of the OAMP/VAMP, we assumed the SVD structure $\boldsymbol{A} = \mathrm{diag}\{\sigma_0, \ldots, \sigma_{M-1}, \boldsymbol{0}\}\boldsymbol{V}^{\mathrm{T}}$. Note that the CAMP does not require this SVD structure. The CAMP only needs the right-orthogonal invariance of $\boldsymbol{A}$. For a further reduction in the complexity, we used the Hadamard matrix $\boldsymbol{V}^{\mathrm{T}} \in \mathcal{O}_N$ with the rows permuted uniformly and randomly. This matrix $\boldsymbol{A}$ is a practical alternative of right-orthogonally invariant matrices.

We simulated damped AMP [39] with the same Bayes-optimal denoiser $f_t(\cdot) = f_{\mathrm{opt}}(\cdot; v_t)$ as in the CAMP. The variance parameter v_t was computed via the SE equation

$$v_t = \sigma^2 + \frac{1}{\delta}\mathrm{MMSE}(v_{t-1}), \quad \mathrm{MMSE}(v_{-1}) = 1, \tag{85}$$

with

$$\mathrm{MMSE}(v) = \mathbb{E}\left[\{f_{\mathrm{opt}}(x_1 + \sqrt{v}z; v) - x_1\}^2\right], \tag{86}$$

where $z \sim \mathcal{N}(0,1)$ denotes the standard Gaussian random variable independent of x_1. The SE equation (85) was derived in [30] under the assumption of zero-mean i.i.d. Gaussian sensing matrix with compression rate $\delta = M/N$. Furthermore, ξ_t in (23) was replaced by the asymptotic value $\bar{\xi}_t = \mathrm{MMSE}(v_t)/v_t$ [46, Lemma 2]. To improve the convergence property of the AMP, we replaced the update rule (22) with the damped rule

$$\boldsymbol{x}_{t+1} = \zeta f_t(\boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t) + (1-\zeta)\boldsymbol{x}_t. \tag{87}$$

Note that SE cannot describe the exact dynamics of AMP when damping is employed.

For the OAMP/VAMP [40], [42], we used the Bayes-optimal denoiser $f_t(\cdot) = f_{\mathrm{opt}}(\cdot; \bar{v}_{\mathrm{A}\to\mathrm{B},t})$ computed via the SE equations [46]

$$\bar{v}_{\mathrm{A}\to\mathrm{B},t} = \bar{\gamma}_t - \bar{v}_{\mathrm{B}\to\mathrm{A},t}, \quad \bar{v}_{\mathrm{B}\to\mathrm{A},0} = 1, \tag{88}$$

$$\frac{1}{\bar{v}_{\mathrm{B}\to\mathrm{A},t+1}} = \frac{1}{\mathrm{MMSE}(\bar{v}_{\mathrm{A}\to\mathrm{B},t})} - \frac{1}{\bar{v}_{\mathrm{A}\to\mathrm{B},t}}, \tag{89}$$

with

$$\bar{\gamma}_t^{-1} = \lim_{M=\delta N\to\infty} \frac{1}{N}\sum_{m=0}^{M-1} \frac{\sigma_m^2}{\sigma^2 + \bar{v}_{\mathrm{B}\to\mathrm{A},t}\sigma_m^2}. \tag{90}$$

To improve the convergence property, we applied the damping technique: The messages $\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t+1}$ and $v_{\mathrm{B}\to\mathrm{A},t+1}$ in (18) were replaced by the damped messages $\zeta\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t+1} + (1-\zeta)\boldsymbol{x}_{\mathrm{B}\to\mathrm{A},t}$ and $\zeta\bar{v}_{\mathrm{B}\to\mathrm{A},t+1} + (1-\zeta)\bar{v}_{\mathrm{B}\to\mathrm{A},t}$, respectively. Note that damped EP cannot be described via SE.

B. Ill-Conditioned Sensing Matrices

We first consider the parameter θ in the CAMP defined in Section III-E. From Theorem 5, we know that the CAMP is Bayes-optimal for any θ if it converges. Thus, the parameter θ only affects the convergence property of the CAMP.

Figure 1 shows the MSEs of the CAMP for a sensing matrix with condition number $\kappa = 5$ defined in Corollary 3. As a baseline, we plotted the asymptotic MSE of the Bayes-optimal signal recovery [18]–[20]. The CAMP with $\theta = 2$ and $\zeta = 0.9$ converges to the Bayes-optimal performance more slowly than that with $\theta = 0$ and $\zeta = 0.85$. This observation does not necessarily imply that $\theta = 0$ is the best option. When the damping factor $\zeta = 0.9$ is used, the CAMP converges for $\theta = 2$ in the finite-sized system while it diverges for $\theta = 0$. Thus, we conclude that using non-zero $\theta \neq 0$ improves the stability of the CAMP in finite-sized systems.

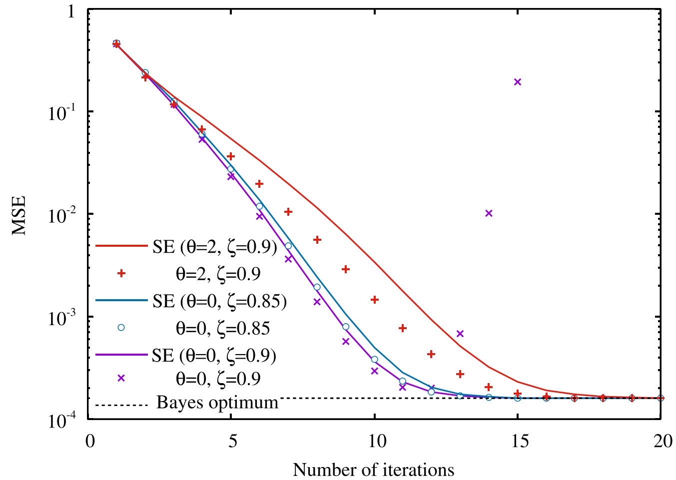

Fig. 1. MSE versus the number of iterations t for the CAMP. $M = 2^{12}$, $N = 2^{13}$, $\rho = 0.1$, $\kappa = 5$, and $1/\sigma^2 = 30$ dB.

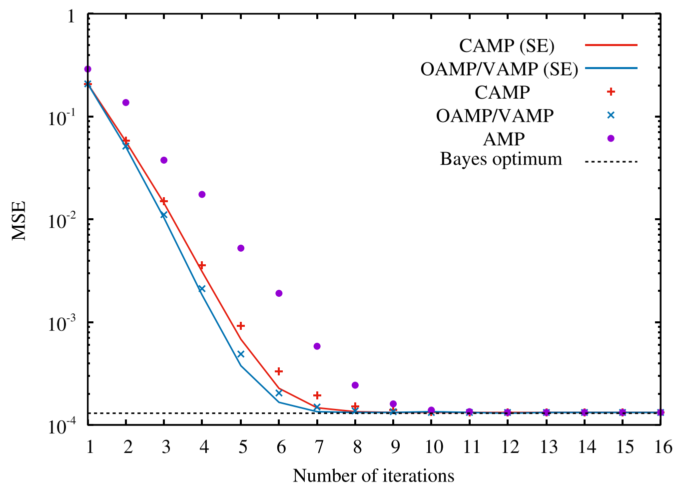

Fig. 2. MSE versus the number of iterations t for the CAMP with $\theta = 0$. $M = 2^{11}$, $N = 2^{12}$, $\rho = 0.1$, $\kappa = 1$, $1/\sigma^2 = 30$ dB, and $\zeta = 1$.

The CAMP is compared to the AMP and OAMP/VAMP for sensing matrices with unit condition number, i.e. orthogonal rows. As noted in Remark 1, without loss of generality, we can use $\theta = 0$ for this case. In this case, the OAMP/VAMP has comparable complexity to the AMP since the SVD of the sensing matrix is not required. Figure 2 shows that the OAMP/VAMP is the best in terms of the convergence speed among the three MP algorithms.

We next consider a sensing matrix with condition number $\kappa = 10$. As shown in Fig. 3, the AMP cannot approach the Bayes-optimal performance. The CAMP converges to the Bayes-optimal performance more slowly than the OAMP/VAMP while the CAMP does not require high-complexity SVD of the sensing matrix. Especially in

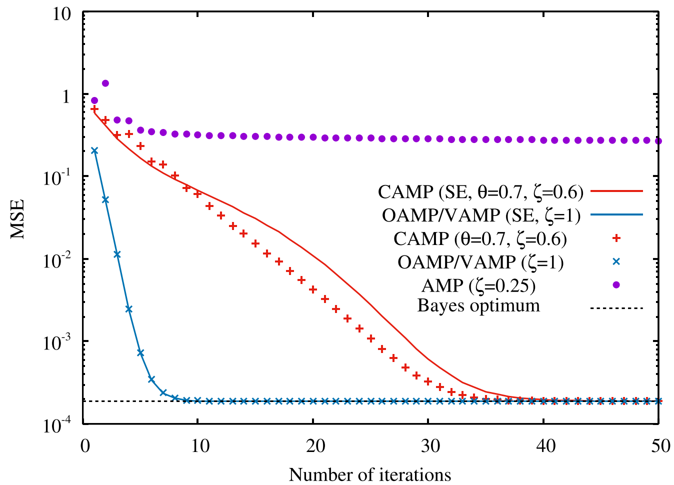

Fig. 3. MSE versus the number of iterations t for the CAMP. $M = 2^{13}$, $N = 2^{14}$, $\rho = 0.1$, $\kappa = 10$, and $1/\sigma^2 = 30$ dB.

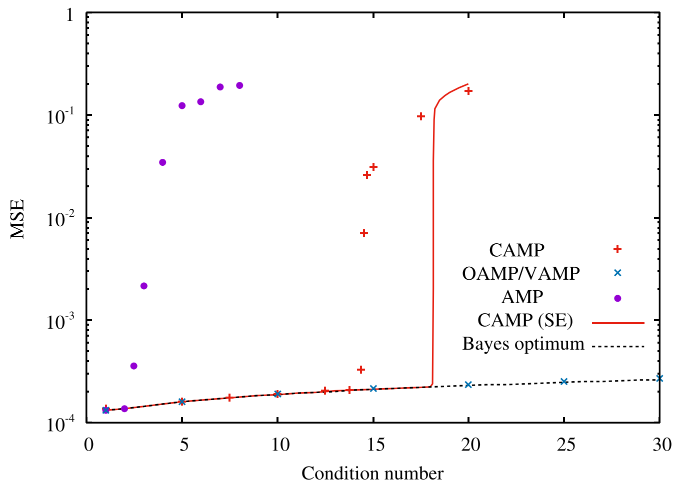

Fig. 4. MSE versus the condition number κ for the CAMP. $M = 512$, $N = 1024$, $\rho = 0.1$, $1/\sigma^2 = 30$ dB, and 150 iterations.

TABLE II

PARAMETERS USED IN FIG. 4

CAMP	OAMP/VAMP	AMP
(κ, θ, ζ)	(κ, ζ)	(κ, ζ)
(1, 0, 0.8)	(1, 0.9)	(1, 1)
(5, 1.65, 0.75)	(5, 0.75)	(2, 0.8)
(7.5, 1.1, 0.6)	(10, 0.7)	(2.5, 0.6)
(10, 0.75, 0.5)	(15, 0.7)	(3, 0.55)
(12.5, 0.75, 0.45)	(20, 0.7)	(4, 0.45)
(13.75, 0.35, 0.25)	(25, 0.7)	(5, 0.35)
(14.375–14.6875, 0.35, 0.2)	(30, 0.7)	(6, 0.35)
(15, 0.3, 0.2)		(7, 0.3)
(17.5, 0.2, 0.1)		(8, 0.3)
(20, 0.1, 0.05)		

large systems, thus, the CAMP should need lower complexity to achieve the Bayes-optimal performance than the OAMP/VAMP.

We investigate the influence of the condition number κ shown in Fig. 4. In evaluating the SE of the CAMP as a baseline, the parameter θ was optimized for each condition number while no damping was employed. In particular, the parameter θ was set to -0.7 for $\kappa \geq 17$. Otherwise, $\theta = 0$ was used. See Table II for the parameters used in the three algorithms,

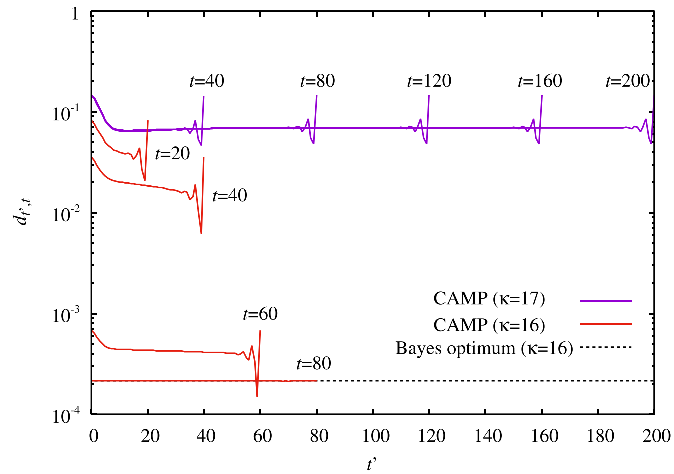

Fig. 5. Correlation $d_{t',t}$ versus $t' = 0, \dots, t$ for the CAMP. $\delta = 0.5$, $\rho = 0.1$, $1/\sigma^2 = 30$ dB, $\theta = 0$, and $\zeta = 1$.

which were numerically optimized for each condition number. More precisely, the parameters were selected so as to achieve the fastest convergence among all possible parameters that approach the best MSE in the last iteration.

The AMP has poor performance with the exception of small condition numbers. The CAMP achieves the Bayes-optimal performance for low-to-moderate condition numbers. However, it is inferior to the high-complexity OAMP/VAMP for large condition numbers. These observations are consistent with the SE results of the CAMP. The SE prediction of the MSE changes rapidly from the Bayes-optimal performance to a large value around a condition number $\kappa \approx 18$ while the OAMP/VAMP still achieves the Bayes-optimal performance for $\kappa > 18$. This is because the CAMP fails to converge for $\kappa > 18$. As a result, we cannot use Theorem 5 to claim the Bayes-optimality of the CAMP. Thus, we conclude that the CAMP is Bayes-optimal in a strictly smaller class of sensing matrices than the OAMP/VAMP.

Finally, we investigate the convergence properties of the CAMP for high condition numbers. Figure 5 shows the correlation $d_{t',t}$ in the CAMP for $t' = 0, \dots, t$. For the condition number $\kappa = 16$, the correlation $d_{t',t}$ converges toward the Bayes-optimal MSE for all t' as t increases. This provides numerical evidence for the assumption in Theorem 5: the convergence of the CAMP toward a fixed-point.

The results for $\kappa = 17$ imply that the CAMP fails to converge. A soliton-like quasi-steady wave propagates as t grows, while the CAMP does not diverge. As implied from Fig. 4, using non-zero $\theta \neq 0$ allows us to avoid the occurrence of such a wave for $\kappa = 17$. However, such waves occur for any θ when the condition number is larger than $\kappa \approx 18$.

Intuitively, the occurrence of soliton-like waves can be understood as follows: The SE equation (77) in time domain becomes unstable for high condition numbers, so that $a_{t',t}$ increases as t grows. However, larger $a_{t',t}$ results in a geometrically smaller forgetting factor $\bar{\xi}_{t-\tau}^{(t-1)}$ in (77), which suppresses the divergence of $a_{t',t}$. As a result, a soliton-like quasi-steady wave occurs for high condition numbers.

V. Conclusion

The Bayes-optimal CAMP solves the disadvantages of AMP and OAMP/VAMP, and realizes their advantages for orthogonally invariant sensing matrices with low-to-moderate condition numbers: The Bayes-optimal CAMP is an efficient MP algorithm that has comparable complexity to AMP. Furthermore, the CAMP has been proved to be Bayes-optimal for all orthogonally invariant sensing matrices if it converges. High-complexity OAMP/VAMP is Bayes-optimal for this class of sensing matrices while AMP is not. The CAMP converges for sensing matrices with low-to-moderate condition numbers while it fails to converge for high condition numbers.

A disadvantage of CAMP is that it needs all moments of the asymptotic singular-value distribution of the sensing matrix. In general, computation of the moments requires high complexity unless their closed-form is available. To circumvent this issue, deep unfolding [60], [61] might be utilized to learn the tap coefficients in the Onsager correction without using the asymptotic singular-value distribution.

The CAMP has a room for improvement especially in finite-sized and ill-conditioned sensing matrices. One option is a replacement of scalar parameters determined via the SE equation with empirical estimators that depend on the measurements, as considered in AMP and OAMP/VAMP.

Another option is a damping technique that keeps the asymptotic Gaussianity of estimation errors. This paper used a heuristic damping technique to improve the convergence property of the CAMP. However, the heuristic damping breaks the asymptotic Gaussianity. Damped CAMP should be designed via Theorem 1 to guarantee the asymptotic Gaussianity. A recent paper [62] proposed long-memory damping in the MF-based interference cancellation to improve the convergence property of long-memory MP. A possible direction for future work is to design CAMP with long-memory damping.

Appendix A
Proof of Theorem 1

A. Formulation

We use Bolthausen's conditioning technique [32] to prove Theorem 1. In the technique, the random variables are classified into three groups: $\boldsymbol{V}$, $\mathfrak{F} = \{\boldsymbol{\lambda}, \tilde{\boldsymbol{w}}, \boldsymbol{x}\}$, and $\mathfrak{E}_{t,t'} = \{\boldsymbol{B}_{t'}, \tilde{\boldsymbol{M}}_{t'}, \boldsymbol{H}_t, \tilde{\boldsymbol{Q}}_{t+1}\}$ with $\tilde{\boldsymbol{Q}}_{t+1} = (\tilde{\boldsymbol{q}}_0, \ldots, \tilde{\boldsymbol{q}}_t)$ and $\tilde{\boldsymbol{M}}_t = (\tilde{\boldsymbol{m}}_0, \ldots, \tilde{\boldsymbol{m}}_{t-1})$. The random variables in $\mathfrak{F}$ are fixed throughout the proof of Theorem 1 while $\boldsymbol{V}$ is averaged out.

The set $\mathfrak{E}_{t,t}$ contains all messages just before updating $\boldsymbol{b}_t = \boldsymbol{V}^{\mathrm{T}}\tilde{\boldsymbol{q}}_t$ while $\mathfrak{E}_{t,t+1}$ includes all messages just before updating $\boldsymbol{h}_t = \boldsymbol{V}\tilde{\boldsymbol{m}}_t$. The main part in the conditioning technique is evaluation of the conditional distribution of $\boldsymbol{b}_t$ given $\mathfrak{E}_{t,t}$ and $\mathfrak{F}$ via that of $\boldsymbol{V}$.

Theorem 1 is proved by induction. More precisely, we prove a theorem obtained by adding several technical results to Theorem 1. Before presenting the theorem, we first define several notations.

The notation $\boldsymbol{o}(1)$ denotes a finite-dimensional vector with vanishing norm. For a tall matrix $\boldsymbol{M} \in \mathbb{R}^{N\times t}$ with rank $r \leq t$, the SVD of $\boldsymbol{M}$ is denoted by $\boldsymbol{M} = \boldsymbol{\Phi}_{\boldsymbol{M}}\boldsymbol{\Sigma}_{\boldsymbol{M}}\boldsymbol{\Psi}_{\boldsymbol{M}}^{\mathrm{T}}$, with $\boldsymbol{\Phi}_{\boldsymbol{M}} = (\boldsymbol{\Phi}_{\boldsymbol{M}}^{\parallel}, \boldsymbol{\Phi}_{\boldsymbol{M}}^{\perp})$. The matrix $\boldsymbol{\Phi}_{\boldsymbol{M}}^{\parallel} \in \mathcal{O}_{N\times r}$ consists of all left-singular vectors corresponding to r non-zero singular values while $\boldsymbol{\Phi}_{\boldsymbol{M}}^{\perp} \in \mathcal{O}_{N\times(N-r)}$ is composed of left-singular vectors corresponding to $N-r$ zero singular values. The matrix $\boldsymbol{P}_{\boldsymbol{M}}^{\parallel} = \boldsymbol{M}(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M})^{-1}\boldsymbol{M}^{\mathrm{T}}$ is the projection to the space spanned by the columns of $\boldsymbol{M}$ while $\boldsymbol{P}_{\boldsymbol{M}}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{M}}^{\parallel}$ is the projection to the orthogonal complement. Note that $\boldsymbol{P}_{\boldsymbol{M}}^{\parallel} = \boldsymbol{\Phi}_{\boldsymbol{M}}^{\parallel}(\boldsymbol{\Phi}_{\boldsymbol{M}}^{\parallel})^{\mathrm{T}}$ and $\boldsymbol{P}_{\boldsymbol{M}}^{\perp} = \boldsymbol{\Phi}_{\boldsymbol{M}}^{\perp}(\boldsymbol{\Phi}_{\boldsymbol{M}}^{\perp})^{\mathrm{T}}$ hold.

In the following theorem, we call the system with respect to $\{\boldsymbol{B}_t, \tilde{\boldsymbol{M}}_t\}$ module A while we refer to that for $\{\boldsymbol{H}_t, \tilde{\boldsymbol{Q}}_{t+1}\}$ as module B.

Theorem 6: Suppose that Assumptions 1–4 hold. Then, the following properties in module A hold for all $\tau = 0, 1, \ldots$ in the large system limit.

(A1) Let $\boldsymbol{\beta}_\tau = (\tilde{\boldsymbol{Q}}_\tau^{\mathrm{T}}\tilde{\boldsymbol{Q}}_\tau)^{-1}\tilde{\boldsymbol{Q}}_\tau^{\mathrm{T}}\tilde{\boldsymbol{q}}_\tau$, $\tilde{\boldsymbol{q}}_\tau^{\perp} = \boldsymbol{P}_{\tilde{\boldsymbol{Q}}_\tau}^{\perp}\tilde{\boldsymbol{q}}_\tau$, and

$$\tilde{\boldsymbol{\omega}}_\tau = \tilde{\boldsymbol{V}}^{\mathrm{T}}(\boldsymbol{\Phi}_{(\tilde{\boldsymbol{Q}}_\tau, \boldsymbol{H}_\tau)}^{\perp})^{\mathrm{T}}\tilde{\boldsymbol{q}}_\tau, \tag{91}$$

where $\tilde{\boldsymbol{V}} \in \mathcal{O}_{N-2\tau}$ is a Haar orthogonal matrix and independent of $\mathfrak{F}$ and $\mathfrak{E}_{\tau,\tau}$. Then, for $\tau > 0$

$$\boldsymbol{b}_\tau \sim \boldsymbol{B}_\tau\boldsymbol{\beta}_\tau + \tilde{\boldsymbol{M}}_\tau\boldsymbol{o}(1) + \boldsymbol{B}_\tau\boldsymbol{o}(1) + \boldsymbol{\Phi}_{(\boldsymbol{B}_\tau, \tilde{\boldsymbol{M}}_\tau)}^{\perp}\tilde{\boldsymbol{\omega}}_\tau \tag{92}$$

conditioned on $\mathfrak{F}$ and $\mathfrak{E}_{\tau,\tau}$ in the large system limit, with

$$\lim_{M=\delta N\to\infty}\frac{1}{N}\left\{\|\tilde{\boldsymbol{\omega}}_\tau\|^2 - \|\tilde{\boldsymbol{q}}_\tau^{\perp}\|^2\right\} \overset{\mathrm{a.s.}}{=} 0. \tag{93}$$

(A2) Suppose that $\tilde{\boldsymbol{\phi}}_\tau(\boldsymbol{B}_{\tau+1}, \tilde{\boldsymbol{w}}, \boldsymbol{\lambda}) : \mathbb{R}^{N\times(\tau+3)} \to \mathbb{R}^N$ is separable, pseudo-Lipschitz of order k with respect to the first $\tau+2$ variables, and proper. If $N^{-1}\tilde{\boldsymbol{q}}_t^{\mathrm{T}}\tilde{\boldsymbol{q}}_{t'}$ converges almost surely to some constant $\kappa_{t,t'} \in \mathbb{R}$ in the large system limit for all $t, t' = 0, \ldots, \tau$, then

$$\langle\tilde{\boldsymbol{\phi}}_\tau(\boldsymbol{B}_{\tau+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle - \mathbb{E}\left[\langle\tilde{\boldsymbol{\phi}}_\tau(\tilde{\boldsymbol{Z}}_{\tau+1}, \tilde{\boldsymbol{w}}, \boldsymbol{\lambda})\rangle\right] \overset{\mathrm{a.s.}}{\to} 0. \tag{94}$$

In (94), $\tilde{\boldsymbol{Z}}_{\tau+1} = (\tilde{\boldsymbol{z}}_0, \ldots, \tilde{\boldsymbol{z}}_\tau) \in \mathbb{R}^{N\times(\tau+1)}$ denotes a zero-mean Gaussian random matrix with covariance $\mathbb{E}[\tilde{\boldsymbol{z}}_t\tilde{\boldsymbol{z}}_{t'}^{\mathrm{T}}] = \kappa_{t,t'}\boldsymbol{I}_N$ for all $t, t' = 0, \ldots, \tau$. In evaluating the expectation in (94), $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{w}$ in (13) follows the zero-mean Gaussian distribution with covariance $\sigma^2\boldsymbol{I}_M$. In particular, for $k = 1$ we have

$$\langle\partial_{\tau'}\tilde{\boldsymbol{\phi}}_\tau(\boldsymbol{B}_{\tau+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle - \mathbb{E}\left[\langle\partial_{\tau'}\tilde{\boldsymbol{\phi}}_\tau(\tilde{\boldsymbol{Z}}_{\tau+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle\right] \overset{\mathrm{a.s.}}{\to} 0 \tag{95}$$

for all $\tau' = 0, \ldots, \tau$.

(A3) Suppose that $\tilde{\boldsymbol{\phi}}_\tau(\boldsymbol{B}_{\tau+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda}) : \mathbb{R}^{N\times(\tau+3)} \to \mathbb{R}^N$ is separable, Lipschitz-continuous with respect to the first $\tau+2$ variables, and proper. Then,

$$\frac{1}{N}\boldsymbol{b}_{\tau'}^{\mathrm{T}}\left(\tilde{\boldsymbol{\phi}}_\tau - \sum_{t'=0}^{\tau}\left\langle\partial_{t'}\tilde{\boldsymbol{\phi}}_\tau\right\rangle\boldsymbol{b}_{t'}\right) \overset{\mathrm{a.s.}}{\to} 0 \tag{96}$$

for all $\tau' = 0, \ldots, \tau$.

(A4) The inner product $N^{-1}\tilde{\boldsymbol{m}}_{\tau'}^{\mathrm{T}}\tilde{\boldsymbol{m}}_\tau$ converges almost surely to some constant $\pi_{\tau',\tau} \in \mathbb{R}$ for all $\tau' = 0, \ldots, \tau$.

(A5) For some $\epsilon > 0$ and $C > 0$,

$$\lim_{M=\delta N\to\infty}\mathbb{E}\left[|\tilde{m}_{\tau,n}|^{2k-2+\epsilon}\right] < \infty, \tag{97}$$

$$\liminf_{M=\delta N\to\infty}\lambda_{\min}\left(\frac{1}{N}\tilde{\boldsymbol{M}}_{\tau+1}^{\mathrm{T}}\tilde{\boldsymbol{M}}_{\tau+1}\right) \overset{\mathrm{a.s.}}{>} C. \tag{98}$$

The following properties in module B hold for all $\tau = 0, 1, \ldots$ in the large system limit.

(B1) Let $\boldsymbol{\alpha}_\tau = (\tilde{\boldsymbol{M}}_\tau^{\mathrm{T}} \tilde{\boldsymbol{M}}_\tau)^{-1} \tilde{\boldsymbol{M}}_\tau^{\mathrm{T}} \tilde{\boldsymbol{m}}_\tau$, $\tilde{\boldsymbol{m}}_0^{\perp} = \tilde{\boldsymbol{m}}_0$, $\tilde{\boldsymbol{m}}_\tau^{\perp} = \boldsymbol{P}_{\tilde{\boldsymbol{M}}_\tau}^{\perp} \tilde{\boldsymbol{m}}_\tau$, and

$$\boldsymbol{\omega}_\tau = \begin{cases} \tilde{\boldsymbol{V}}(\boldsymbol{\Phi}_{\boldsymbol{b}_0}^{\perp})^{\mathrm{T}} \tilde{\boldsymbol{m}}_0 & \text{for } \tau = 0, \\ \tilde{\boldsymbol{V}}(\boldsymbol{\Phi}_{(\tilde{\boldsymbol{M}}_\tau, \boldsymbol{B}_{\tau+1})}^{\perp})^{\mathrm{T}} \tilde{\boldsymbol{m}}_\tau & \text{for } \tau > 0, \end{cases} \tag{99}$$

where $\tilde{\boldsymbol{V}} \in \mathcal{O}_{N-(2\tau+1)}$ is a Haar orthogonal matrix and independent of $\mathfrak{F}$ and $\mathfrak{E}_{\tau,\tau+1}$. Then, we have

$$\boldsymbol{h}_0 \sim o(1)\tilde{\boldsymbol{q}}_0 + \boldsymbol{\Phi}_{\tilde{\boldsymbol{q}}_0}^{\perp} \boldsymbol{\omega}_\tau, \tag{100}$$

conditioned on $\mathfrak{F}$ and $\mathfrak{E}_{0,1} = \{\boldsymbol{b}_0, \tilde{\boldsymbol{m}}_0, \tilde{\boldsymbol{q}}_0\}$ in the large system limit. For $\tau > 0$

$$\boldsymbol{h}_\tau \sim \boldsymbol{H}_\tau \boldsymbol{\alpha}_\tau + \tilde{\boldsymbol{Q}}_{\tau+1} \boldsymbol{o}(1) + \boldsymbol{H}_\tau \boldsymbol{o}(1) + \boldsymbol{\Phi}_{(\boldsymbol{H}_\tau, \tilde{\boldsymbol{Q}}_{\tau+1})}^{\perp} \boldsymbol{\omega}_\tau, \tag{101}$$

conditioned on $\mathfrak{F}$ and $\mathfrak{E}_{\tau,\tau+1}$ in the large system limit, with

$$\lim_{M=\delta N \to \infty} \frac{1}{N} \left\{ \|\boldsymbol{\omega}_\tau\|^2 - \|\tilde{\boldsymbol{m}}_\tau^{\perp}\|^2 \right\} \overset{\text{a.s.}}{=} 0. \tag{102}$$

(B2) Suppose that $\tilde{\boldsymbol{\psi}}_\tau(\boldsymbol{H}_{\tau+1}, \boldsymbol{x}) : \mathbb{R}^{N \times (\tau+2)} \to \mathbb{R}^N$ is a separable and proper pseudo-Lipschitz function of order k. If $N^{-1} \tilde{\boldsymbol{m}}_t^{\mathrm{T}} \tilde{\boldsymbol{m}}_{t'}$ converges almost surely to some constant $\pi_{t,t'} \in \mathbb{R}$ in the large system limit for all $t, t' = 0, \ldots, \tau$, then

$$\langle \tilde{\boldsymbol{\psi}}_\tau(\boldsymbol{H}_{\tau+1}, \boldsymbol{x}) \rangle - \mathbb{E}\left[\langle \tilde{\boldsymbol{\psi}}_\tau(\boldsymbol{Z}_{\tau+1}, \boldsymbol{x}) \rangle \right] \overset{\text{a.s.}}{\to} 0, \tag{103}$$

where $\boldsymbol{Z}_{\tau+1} = (\boldsymbol{z}_0, \ldots, \boldsymbol{z}_\tau) \in \mathbb{R}^{N \times (\tau+1)}$ denotes a zero-mean Gaussian random matrix with covariance $\mathbb{E}[\boldsymbol{z}_t \boldsymbol{z}_{t'}^{\mathrm{T}}] = \pi_{t,t'} \boldsymbol{I}_N$ for all $t, t' = 0, \ldots, \tau$. In particular, for $k = 1$ we have

$$\langle \partial_{\tau'} \tilde{\boldsymbol{\psi}}_\tau(\boldsymbol{H}_{\tau+1}, \boldsymbol{x}) \rangle - \mathbb{E}\left[\langle \partial_{\tau'} \tilde{\boldsymbol{\psi}}_\tau(\boldsymbol{Z}_{\tau+1}, \boldsymbol{x}) \rangle \right] \overset{\text{a.s.}}{\to} 0 \tag{104}$$

for all $\tau' = 0, \ldots, \tau$.

(B3) Suppose that $\tilde{\boldsymbol{\psi}}_\tau(\boldsymbol{H}_{\tau+1}, \boldsymbol{x}) : \mathbb{R}^{N \times (\tau+2)} \to \mathbb{R}^N$ is a separable and proper Lipschitz-continuous function. Then,

$$\frac{1}{N} \boldsymbol{h}_{\tau'}^{\mathrm{T}} \left(\tilde{\boldsymbol{\psi}}_\tau - \sum_{t'=0}^{\tau} \left\langle \partial_{t'} \tilde{\boldsymbol{\psi}}_\tau \right\rangle \boldsymbol{h}_{t'} \right) \overset{\text{a.s.}}{\to} 0 \tag{105}$$

for all $\tau' = 0, \ldots, \tau$.

(B4) The inner product $N^{-1} \tilde{\boldsymbol{q}}_{\tau'}^{\mathrm{T}} \tilde{\boldsymbol{q}}_{\tau+1}$ converges almost surely to some constant $\pi_{\tau',\tau+1} \in \mathbb{R}$ for all $\tau' = 0, \ldots, \tau+1$.

(B5) For some $\epsilon > 0$ and $C > 0$,

$$\lim_{M=\delta N \to \infty} \mathbb{E}\left[|\tilde{q}_{\tau+1,n}|^{2+\epsilon} \right] < \infty, \tag{106}$$

$$\liminf_{M=\delta N \to \infty} \lambda_{\min} \left(\frac{1}{N} \tilde{\boldsymbol{Q}}_{\tau+2}^{\mathrm{T}} \tilde{\boldsymbol{Q}}_{\tau+2} \right) \overset{\text{a.s.}}{>} C. \tag{107}$$

We summarize useful lemmas used in the proof of Theorem 6 by induction.

Lemma 1 ([42], [46]): Suppose that $\boldsymbol{X} \in \mathbb{R}^{N \times t}$ has full rank for $0 < t < N$, and consider the noiseless and compressed observation $\boldsymbol{Y} \in \mathbb{R}^{N \times t}$ of $\boldsymbol{V}$ given by

$$\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{X}. \tag{108}$$

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

$$\boldsymbol{V}|_{\boldsymbol{X},\boldsymbol{Y}} \sim \boldsymbol{Y}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y})^{-1}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{\Phi}_{\boldsymbol{Y}}^{\perp} \tilde{\boldsymbol{V}} (\boldsymbol{\Phi}_{\boldsymbol{X}}^{\perp})^{\mathrm{T}}, \tag{109}$$

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

The following lemma is a generalization of Stein's lemma. The lemma is proved under a different assumption from in [59].

Lemma 2: Let $\boldsymbol{z} = (z_1, \ldots, z_t)^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ for any positive definite covariance matrix $\boldsymbol{\Sigma}$. If $f : \mathbb{R}^t \to \mathbb{R}$ is Lipschitz-continuous, then we have

$$\mathbb{E}[z_1 f(\boldsymbol{z})] = \sum_{t'=1}^{t} \mathbb{E}[z_1 z_{t'}] \mathbb{E}[\partial_{t'} f(\boldsymbol{z})]. \tag{110}$$

Proof: We first confirm that both sides of (110) are bounded. For the left-hand side (LHS), we find $f(\boldsymbol{z}) = \mathcal{O}(\|\boldsymbol{z}\|)$ as $\|\boldsymbol{z}\| \to \infty$ since f is Lipschitz-continuous. Thus, $\mathbb{E}[z_1 f(\boldsymbol{z})]$ is bounded for $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$.

For the RHS, we use the Lipschitz-continuity of f to find that there is some Lipschitz-constant $L > 0$ such that

$$\left| \frac{f(\boldsymbol{z} + \Delta \boldsymbol{e}_{t'}) - f(\boldsymbol{z})}{\Delta} \right| \leq L \tag{111}$$

holds for any $\Delta \neq 0$, where $\boldsymbol{e}_{t'} \in \mathbb{R}^t$ is the t'th column of $\boldsymbol{I}_t$. This implies that each partial derivative $\partial_{t'} f$ is bounded almost everywhere since the partial derivatives of any Lipschitz-continuous function exist almost everywhere. Thus, $\mathbb{E}[\partial_{t'} f(\boldsymbol{z})]$ is bounded. These observations indicate the boundedness of both sides in (110).

For the eigen-decomposition $\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathrm{T}}$, we use the change of variables $\tilde{\boldsymbol{z}} = \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{z}$ to obtain

$$\mathbb{E}[z_1 f(\boldsymbol{z})] = \sum_{\tau=1}^{t} [\boldsymbol{\Phi}]_{1,\tau} \mathbb{E}[\tilde{z}_\tau f(\boldsymbol{\Phi}\tilde{\boldsymbol{z}})] = \sum_{\tau=1}^{t} [\boldsymbol{\Phi}]_{1,\tau} \mathbb{E}[\tilde{z}_\tau g(\tilde{z}_\tau)], \tag{112}$$

with $g(\tilde{z}_\tau) = \mathbb{E}[f(\boldsymbol{\Phi}\tilde{\boldsymbol{z}}) | \tilde{z}_\tau]$.

We prove that g is Lipschitz-continuous. Let $\tilde{\boldsymbol{z}}_x$ denote the vector obtained by replacing $\tilde{z}_\tau$ in $\tilde{\boldsymbol{z}}$ with x. Since $\tilde{\boldsymbol{z}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda})$ has independent elements, we have

$$\begin{aligned} |g(x) - g(y)| &\leq \mathbb{E}\left[|f(\boldsymbol{\Phi}\tilde{\boldsymbol{z}}_x) - f(\boldsymbol{\Phi}\tilde{\boldsymbol{z}}_y)| \right] \\ &\leq L\mathbb{E}\left[\|\boldsymbol{\Phi}(\tilde{\boldsymbol{z}}_x - \tilde{\boldsymbol{z}}_y)\| \right] \\ &= L\mathbb{E}\left[\|\tilde{\boldsymbol{z}}_x - \tilde{\boldsymbol{z}}_y\| \right] = L|x - y|, \end{aligned} \tag{113}$$

where the second inequality follows from the Lipschitz-continuity of f with a Lipschitz-constant $L > 0$. Thus, $g(\tilde{z}_\tau)$ is Lipschitz-continuous, so that $g(\tilde{z}_\tau)$ is differentiable almost everywhere.

Since $\tilde{\boldsymbol{z}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda})$ holds, Stein's lemma [63] yields

$$\begin{aligned} \mathbb{E}[z_1 f(\boldsymbol{z})] &= \sum_{\tau=1}^{t} [\boldsymbol{\Phi}]_{1,\tau} \mathbb{E}[\tilde{z}_\tau^2] \mathbb{E}\left[g'(\tilde{z}_\tau) \right] \\ &= \sum_{\tau=1}^{t} [\boldsymbol{\Phi}]_{1,\tau} [\boldsymbol{\Lambda}]_{\tau,\tau} \mathbb{E}\left[\sum_{t'=1}^{t} [\boldsymbol{\Phi}]_{t',\tau} \partial_{t'} f(\boldsymbol{z}) \right]. \end{aligned} \tag{114}$$

Using the identity

$$\sum_{\tau=1}^{t}[\boldsymbol{\Phi}]_{1,\tau}[\boldsymbol{\Lambda}]_{\tau,\tau}[\boldsymbol{\Phi}]_{t',\tau} = [\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathrm{T}}]_{1,t'} = \mathbb{E}[z_1 z_{t'}], \tag{115}$$

we arrive at Lemma 2. ■

Lemma 3 ([46]): For $t \in \mathbb{N}$, suppose that $\boldsymbol{f} : \mathbb{R}^{N\times(t+1)} \to \mathbb{R}^N$ is separable and pseudo-Lipschitz of order k. Let $L_n > 0$ denote a Lipschitz constant of the nth element $[\boldsymbol{f}]_n$. The sequence of Lipschitz constants is assumed to satisfy

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}L_n^2 < \infty. \tag{116}$$

Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)^{\mathrm{T}} \in \mathbb{R}^N$ denote a vector that satisfies

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}L_n\epsilon_n^2 \overset{\text{a.s.}}{=} 0, \tag{117}$$

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}L_n\epsilon_n^{2k-2} \overset{\text{a.s.}}{<} \infty. \tag{118}$$

Suppose that $\boldsymbol{A}_{t+1} = (\boldsymbol{a}_0, \ldots, \boldsymbol{a}_t) \in \mathbb{R}^{N\times(t+1)}$ satisfies

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}L_n^i a_{t',n}^{2k-2} \overset{\text{a.s.}}{<} \infty \quad \text{for } i = 1, 2. \tag{119}$$

For $t' > 0$, let $\boldsymbol{E} = (\boldsymbol{e}_1^{\mathrm{T}}, \ldots, \boldsymbol{e}_N^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{N\times t'}$ denote a matrix that satisfies

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}L_n\|\boldsymbol{e}_n\|^{\max\{2,2k-2\}} \overset{\text{a.s.}}{<} \infty, \tag{120}$$

$$\liminf_{N\to\infty}\lambda_{\min}\left(\frac{1}{N}\boldsymbol{E}^{\mathrm{H}}\boldsymbol{E}\right) \overset{\text{a.s.}}{>} C \tag{121}$$

for some constant $C > 0$. Suppose that $\boldsymbol{\omega} \in \mathbb{R}^{N-t'}$ is an orthogonally invariant random vector conditioned on $\boldsymbol{\epsilon}$, $\boldsymbol{A}_{t+1}$, and $\boldsymbol{E}$. For some $v > 0$, postulate the following:

$$\lim_{N\to\infty}\frac{1}{N}\|\boldsymbol{\omega}\|^2 \overset{\text{a.s.}}{=} v > 0. \tag{122}$$

Let $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, v\boldsymbol{I}_N)$ denote a standard Gaussian random vector independent of the other random variables. Then,

$$\lim_{N\to\infty}\left\langle \boldsymbol{f}(\boldsymbol{A}_t, \boldsymbol{a}_t+\boldsymbol{\epsilon}+\boldsymbol{\Phi}_{\boldsymbol{E}}^{\perp}\boldsymbol{\omega}) - \mathbb{E}_{\boldsymbol{z}}[\boldsymbol{f}(\boldsymbol{A}_t, \boldsymbol{a}_t+\boldsymbol{z})]\right\rangle \overset{\text{a.s.}}{=} 0. \tag{123}$$

B. Module a for $\tau = 0$

Proof of Property (A2) for $\tau = 0$: The latter property (95) follows from the former property (94) and a technical result proved in [30, Lemma 5]. Thus, we only prove the former property for $\tau = 0$.

Property (94) follows from Lemma 3 for $\boldsymbol{f}(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{b}}_0) = \tilde{\boldsymbol{\phi}}_0(\tilde{\boldsymbol{b}}_0, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})$ with $\boldsymbol{a}_0 = \tilde{\boldsymbol{w}}$, $\boldsymbol{a}_1 + \boldsymbol{\epsilon} = \boldsymbol{0}$, $\boldsymbol{\Phi}_{\boldsymbol{E}}^{\perp} = \boldsymbol{I}_N$, and $\boldsymbol{\omega} = \tilde{\boldsymbol{b}}_0$. We confirm all conditions in Lemma 3. Applying Hölder's inequality for any $\epsilon > 0$, we have

$$\frac{1}{N}\sum_{n=1}^{N}L_n^i\tilde{w}_n^{2k-2} \leq \left(\frac{1}{N}\sum_{n=1}^{N}L_n^{ip}\right)^{1/p}\left(\frac{1}{N}\sum_{n=1}^{N}\tilde{w}_n^{2k-2+\epsilon}\right)^{1/q} \tag{124}$$

for $i = 1, 2$, with $q = 1 + \epsilon/(2k-2)$ and $p^{-1} = 1 - q^{-1}$, which is bounded because of Assumption 3. Furthermore, the definition $\boldsymbol{b}_0 = -\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}$ implies the orthogonal invariance and $N^{-1}\|\boldsymbol{b}_0\|^2 \overset{\text{a.s.}}{\to} 1$. Thus, all conditions in Lemma 3 hold. Using Lemma 3, we obtain

$$\langle\tilde{\boldsymbol{\phi}}_0(\tilde{\boldsymbol{b}}_0, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle - \mathbb{E}_{\tilde{\boldsymbol{z}}_0}\left[\langle\tilde{\boldsymbol{\phi}}_0(\tilde{\boldsymbol{z}}_0, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle\right] \overset{\text{a.s.}}{\to} 0, \tag{125}$$

with $\tilde{\boldsymbol{z}}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$.

We repeat the use of Lemma 3 for $\boldsymbol{f}(\tilde{\boldsymbol{z}}_0, \tilde{\boldsymbol{w}}) = \tilde{\boldsymbol{\phi}}_0(\tilde{\boldsymbol{z}}_0, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})$ with $\boldsymbol{a}_0 = \tilde{\boldsymbol{z}}_0$ and $\boldsymbol{\omega} = \tilde{\boldsymbol{w}}$. Using Lemma 3 from Assumption 3 and applying Assumption 2, we obtain

$$\langle\tilde{\boldsymbol{\phi}}_0(\tilde{\boldsymbol{z}}_0, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle - \mathbb{E}\left[\langle\tilde{\boldsymbol{\phi}}_0(\tilde{\boldsymbol{z}}_0, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle\right] \overset{\text{a.s.}}{\to} 0. \tag{126}$$

In evaluating the expectation over $\tilde{\boldsymbol{w}}$, the first M elements $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{w}$ in (13) follow $\mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_M)$. Combining these results, we arrive at (94) for $\tau = 0$. ■

Proof of (A3) for $\tau = 0$: The LHS of (96) is a separable and proper pseudo-Lipschitz function of order 2. We can use (94) for $\tau = 0$ to find that the LHS of (96) converges almost surely to its expectation in which $\boldsymbol{b}_0$ and $\langle\partial_0\tilde{\boldsymbol{\phi}}_0\rangle$ are replaced by $\tilde{\boldsymbol{z}}_0 \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ and the expected one, respectively. Thus, it is sufficient to evaluate the expectation.

The function $\boldsymbol{f}(\tilde{\boldsymbol{z}}_0; \tilde{\boldsymbol{w}}, \boldsymbol{\lambda}) = \tilde{\boldsymbol{\phi}}_0(\tilde{\boldsymbol{z}}_0, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda}) - \mathbb{E}[\langle\partial_0\tilde{\boldsymbol{\phi}}_0\rangle]\tilde{\boldsymbol{z}}_0$ is a separable Lipschitz-continuous function of $\tilde{\boldsymbol{z}}_0$. Thus, we can use Lemma 2 to obtain

$$\begin{aligned}&\frac{1}{N}\mathbb{E}\left[\tilde{\boldsymbol{z}}_0^{\mathrm{T}}\left(\tilde{\boldsymbol{\phi}}_0 - \mathbb{E}\left[\left\langle\partial_0\tilde{\boldsymbol{\phi}}_0\right\rangle\right]\tilde{\boldsymbol{z}}_0\right)\right] \\ =&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[\tilde{z}_{0,n}^2\right]\mathbb{E}\left[\partial_0\tilde{\phi}_{0,n}\right] - \mathbb{E}\left[\left\langle\partial_0\tilde{\boldsymbol{\phi}}_0\right\rangle\right] = 0.\end{aligned} \tag{127}$$

Thus, (96) holds for $\tau = 0$. ■

Proof of (A4) for $\tau = 0$: From the definition (11) of $\tilde{\boldsymbol{m}}_0$ and (96), we find the orthogonality $N^{-1}\boldsymbol{b}_0^{\mathrm{T}}\tilde{\boldsymbol{m}}_0 \overset{\text{a.s.}}{\to} 0$. Using this orthogonality and (95) for $\tau = 0$ yields

$$\begin{aligned}\frac{1}{N}\|\tilde{\boldsymbol{m}}_0\|^2 &\overset{\text{a.s.}}{=} \frac{1}{N}\boldsymbol{m}_0^{\mathrm{T}}\tilde{\boldsymbol{m}}_0 + o(1) \\ &= \frac{1}{N}\boldsymbol{m}_0^{\mathrm{T}}\boldsymbol{m}_0 - \mathbb{E}\left[\langle\partial_0\boldsymbol{\phi}_0\rangle\right]\frac{\boldsymbol{m}_0^{\mathrm{T}}\boldsymbol{b}_0}{N} + o(1).\end{aligned} \tag{128}$$

The first and second terms are separable and proper pseudo-Lipschitz functions of order 2. From (94) for $\tau = 0$, they converge almost surely to their expected terms. Thus, $N^{-1}\|\tilde{\boldsymbol{m}}_0\|^2$ converges almost surely to a constant. ■

Proof of Property (A5) for $\tau = 0$: The latter property (98) for $\tau = 0$ follows from the nonlinearity of $\boldsymbol{\phi}_0$ in Assumption 4. Thus, we only prove the former property (97) for $\tau = 0$.

The proper Lipschitz-continuity in Assumption 4 implies the upper bound $|\tilde{m}_{0,n}| \leq C_n(1 + |b_{0,n}| + |\tilde{w}_{0,n}|)$ for some λ_n-dependent constant C_n. From Assumptions 1 and 3, we find that $\boldsymbol{b}_0$ and $\tilde{\boldsymbol{w}}$ have bounded $(2k-2+\epsilon)$th moments for some $\epsilon > 0$. Thus, we obtain the former property (97) for $\tau = 0$. ■

C. Module B for $\tau = 0$

Proof of Property (B1) for $\tau = 0$: Lemma 1 for the constraint $\boldsymbol{V}\boldsymbol{b}_0 = \tilde{\boldsymbol{q}}_0$ implies

$$\boldsymbol{V} \sim \frac{\tilde{\boldsymbol{q}}_0\boldsymbol{b}_0^{\mathrm{T}}}{\|\tilde{\boldsymbol{q}}_0\|^2} + \boldsymbol{\Phi}_{\tilde{\boldsymbol{q}}_0}^{\perp}\tilde{\boldsymbol{V}}(\boldsymbol{\Phi}_{\boldsymbol{b}_0}^{\perp})^{\mathrm{T}} \tag{129}$$

conditioned on $\mathfrak{F}$ and $\mathfrak{E}_{0,0}$, where $\tilde{\boldsymbol{V}} \in \mathcal{O}_{N-1}$ is Haar orthogonal and independent of $\boldsymbol{b}_0$ and $\tilde{\boldsymbol{q}}_0$. Using the definition (11) of $\boldsymbol{h}_0$ and the orthogonality $N^{-1}\boldsymbol{b}_0^{\mathrm{T}}\tilde{\boldsymbol{m}}_0 \overset{\text{a.s.}}{\to} 0$ obtained from Property (A3) for $\tau = 0$, we obtain (100).

To complete the proof of Property (B1) for $\tau = 0$, we prove (102) for $\tau = 0$. By definition,

$$\frac{1}{N}\|\tilde{\boldsymbol{\omega}}_0\|^2 = \frac{1}{N}\tilde{\boldsymbol{m}}_0^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{b}_0}^{\perp}\tilde{\boldsymbol{m}}_0 \overset{\text{a.s.}}{=} \frac{1}{N}\|\tilde{\boldsymbol{m}}_0\|^2, \tag{130}$$

where the last equality follows from the orthogonality $N^{-1}\boldsymbol{b}_0^{\mathrm{T}}\tilde{\boldsymbol{m}}_0 \overset{\text{a.s.}}{\to} 0$. Thus, (102) holds for $\tau = 0$, because of the notational convention $\tilde{\boldsymbol{m}}_0^{\perp} = \tilde{\boldsymbol{m}}_0$. ∎

Proof of Property (B2) for $\tau = 0$: Since the latter property (104) follows from the former property (103), we only prove the former property for $\tau = 0$. Using Property (B1) for $\tau = 0$ and Lemma 3 for $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{h}_0) = \tilde{\boldsymbol{\psi}}_0(\boldsymbol{h}_0, \boldsymbol{x})$ with $\boldsymbol{a}_0 = \boldsymbol{x}$, $\boldsymbol{a}_1 = \boldsymbol{0}$, $\boldsymbol{\epsilon} = o(1)\tilde{\boldsymbol{q}}_0$, $\boldsymbol{E} = \tilde{\boldsymbol{q}}_0$, and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, we obtain

$$\langle \tilde{\boldsymbol{\psi}}_0(\boldsymbol{h}_0, \boldsymbol{x})\rangle - \mathbb{E}_{\boldsymbol{z}_0}\left[\langle \tilde{\boldsymbol{\psi}}_0(\boldsymbol{z}_0, \boldsymbol{x})\rangle\right] \overset{\text{a.s.}}{\to} 0, \tag{131}$$

with $\boldsymbol{z}_0 \sim \mathcal{N}(\boldsymbol{0}, \pi_{0,0}\boldsymbol{I}_N)$. Applying Assumption 1 to the second term, we arrive at (103) for $\tau = 0$. ∎

Proof of Properties (B3) and (B4) for $\tau = 0$: Repeat the proofs of Properties (A3) and (A4) for $\tau = 0$. ∎

Proof of Property (B5) for $\tau = 0$: The former property (106) for $\tau = 0$ is obtained by repeating the proof of (97) for $\tau = 0$. See [46, p. 377] for the proof of the latter property (107) for $\tau = 0$. ∎

D. Proof by Induction

Suppose that Theorem 6 is correct for all $\tau < t$. In a proof by induction we need to prove all properties in modules A and B for $\tau = t$. Since the properties for module B can be proved by repeating the proofs for module A, we only prove the properties for module A.

Proof of Property (A1) for $\tau = t$: The matrix $(\boldsymbol{B}_t, \tilde{\boldsymbol{M}}_t)$ has full rank from the induction hypotheses (98) and (107) for $\tau = t-1$, as well as the orthogonality $N^{-1}\boldsymbol{b}_\tau^{\mathrm{T}}\tilde{\boldsymbol{m}}_{\tau'} \overset{\text{a.s.}}{\to} 0$ for all $\tau, \tau' < t$. Using Lemma 1 for the constraint $(\tilde{\boldsymbol{Q}}_t, \boldsymbol{H}_t) = \boldsymbol{V}(\boldsymbol{B}_t, \tilde{\boldsymbol{M}}_t)$, we obtain

$$\begin{aligned}\boldsymbol{V} = \; & (\tilde{\boldsymbol{Q}}_t, \boldsymbol{H}_t)\begin{bmatrix}\tilde{\boldsymbol{Q}}_t^{\mathrm{T}}\tilde{\boldsymbol{Q}}_t & \tilde{\boldsymbol{Q}}_t^{\mathrm{T}}\boldsymbol{H}_t \\ \boldsymbol{H}_t^{\mathrm{T}}\tilde{\boldsymbol{Q}}_t & \boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{H}_t\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{B}_t^{\mathrm{T}} \\ \tilde{\boldsymbol{M}}_t^{\mathrm{T}}\end{bmatrix} \\ & + \boldsymbol{\Phi}_{(\tilde{\boldsymbol{Q}}_t, \boldsymbol{H}_t)}^{\perp}\tilde{\boldsymbol{V}}(\boldsymbol{\Phi}_{(\boldsymbol{B}_t, \tilde{\boldsymbol{M}}_t)}^{\perp})^{\mathrm{T}}\end{aligned} \tag{132}$$

conditioned on $\mathfrak{F}$ and $\mathfrak{E}_{t,t}$. Applying the orthogonality $N^{-1}\boldsymbol{b}_\tau^{\mathrm{T}}\tilde{\boldsymbol{m}}_{\tau'} \overset{\text{a.s.}}{\to} 0$ and $N^{-1}\boldsymbol{h}_\tau^{\mathrm{T}}\tilde{\boldsymbol{q}}_{\tau'} \overset{\text{a.s.}}{\to} 0$ obtained from the induction hypotheses (A3) and (B3) for $\tau < t$, as well as the definition (9) of $\boldsymbol{b}_t$, we have

$$\begin{aligned}\boldsymbol{b}_t \sim \; & \boldsymbol{B}_t(\tilde{\boldsymbol{Q}}_t^{\mathrm{T}}\tilde{\boldsymbol{Q}}_t)^{-1}\tilde{\boldsymbol{Q}}_t^{\mathrm{T}}\tilde{\boldsymbol{q}}_t + \boldsymbol{B}_t\boldsymbol{o}(1) + \tilde{\boldsymbol{M}}_t\boldsymbol{o}(1) \\ & + \boldsymbol{\Phi}_{(\boldsymbol{B}_t, \tilde{\boldsymbol{M}}_t)}^{\perp}\tilde{\boldsymbol{V}}^{\mathrm{T}}(\boldsymbol{\Phi}_{(\tilde{\boldsymbol{Q}}_t, \boldsymbol{H}_t)}^{\perp})^{\mathrm{T}}\tilde{\boldsymbol{q}}_t\end{aligned} \tag{133}$$

conditioned on $\mathfrak{F}$ and $\mathfrak{E}_{t,t}$, which is equivalent to (92) for $\tau = t$.

To complete the proof of Property (A1) for $\tau = t$, we shall prove (93). By definition,

$$\frac{\|\tilde{\boldsymbol{\omega}}_t\|^2}{N} = \frac{\tilde{\boldsymbol{q}}_t^{\mathrm{T}}\boldsymbol{P}_{(\tilde{\boldsymbol{Q}}_t, \boldsymbol{H}_t)}^{\perp}\tilde{\boldsymbol{q}}_t}{N} \overset{\text{a.s.}}{=} \frac{\tilde{\boldsymbol{q}}_t^{\mathrm{T}}\boldsymbol{P}_{\tilde{\boldsymbol{Q}}_t}^{\perp}\tilde{\boldsymbol{q}}_t}{N} + o(1), \tag{134}$$

where the last equality follows from the orthogonality $N^{-1}\boldsymbol{h}_\tau^{\mathrm{T}}\tilde{\boldsymbol{q}}_{\tau'} \overset{\text{a.s.}}{\to} 0$. Thus, (93) holds for $\tau = t$. ∎

Proof of Property (A2) for $\tau = t$: Since the latter property (95) follows from the former property (94), we only prove the former property for $\tau = t$.

We use Property (A1) for $\tau = t$ and Lemma 3 for the function $\boldsymbol{f}(\tilde{\boldsymbol{w}}, \boldsymbol{B}_t, \boldsymbol{b}_t) = \tilde{\boldsymbol{\phi}}_t(\boldsymbol{B}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})$ with $\boldsymbol{A}_{t+1} = (\tilde{\boldsymbol{w}}, \boldsymbol{B}_t)$, $\boldsymbol{a}_{t+1} = \boldsymbol{B}_t\boldsymbol{\beta}_t$, $\boldsymbol{\epsilon} = \tilde{\boldsymbol{M}}_t\boldsymbol{o}(1) + \boldsymbol{B}_t\boldsymbol{o}(1)$, $\boldsymbol{E} = (\boldsymbol{B}_t, \tilde{\boldsymbol{M}}_t)$, and $\boldsymbol{\omega} = \tilde{\boldsymbol{\omega}}$. Then,

$$\langle \tilde{\boldsymbol{\phi}}_t(\boldsymbol{B}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle - \mathbb{E}_{\tilde{\boldsymbol{z}}_t}\left[\langle \tilde{\boldsymbol{\phi}}_t(\boldsymbol{B}_t, \boldsymbol{B}_t\boldsymbol{\beta}_t + \tilde{\boldsymbol{z}}_t, \tilde{\boldsymbol{w}}, \boldsymbol{\lambda})\rangle\right] \overset{\text{a.s.}}{\to} 0, \tag{135}$$

where $\tilde{\boldsymbol{z}}_t$ has independent zero-mean Gaussian elements with variance $\mu_t \overset{\text{a.s.}}{=} N^{-1}\|\tilde{\boldsymbol{q}}_t^{\perp}\|^2$. Repeating this argument yields

$$\langle \tilde{\boldsymbol{\phi}}_t(\boldsymbol{B}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda})\rangle - \mathbb{E}\left[\langle \tilde{\boldsymbol{\phi}}_t(\tilde{\boldsymbol{Z}}_{t+1}, \tilde{\boldsymbol{w}}, \boldsymbol{\lambda})\rangle\right] \overset{\text{a.s.}}{\to} 0, \tag{136}$$

where $\tilde{\boldsymbol{Z}}_{t+1}$ is a zero-mean Gaussian random matrix having independent elements. In evaluating the expectation over $\tilde{\boldsymbol{w}}$, $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{w}$ in (13) follows the zero-mean Gaussian distribution with covariance $\sigma^2\boldsymbol{I}_M$.

To complete the proof of (94) for $\tau = t$, we evaluate the covariance of $\boldsymbol{Z}_{t+1}$. By construction, we have $N^{-1}\mathbb{E}[\boldsymbol{z}_\tau^{\mathrm{T}}\boldsymbol{z}_{\tau'}] = N^{-1}\boldsymbol{b}_\tau^{\mathrm{T}}\boldsymbol{b}_{\tau'} \overset{\text{a.s.}}{=} \kappa_{\tau,\tau'} + o(1)$. Thus, the former property (94) is correct for $\tau = t$. ∎

Proof of Property (A3) for $\tau = t$: The LHS of (96) is a separable and proper pseudo-Lipschitz function of order 2. We can use (94) for $\tau = t$ to find that the LHS of (96) converges almost surely to its expectation in which $\boldsymbol{B}_{t+1}$ and $\langle \partial_{t'}\tilde{\boldsymbol{\phi}}_t\rangle$ are replaced by $\tilde{\boldsymbol{Z}}_{t+1}$ and the expected one, respectively. Thus, it is sufficient to evaluate the expectation.

Since the function $\boldsymbol{f}(\tilde{\boldsymbol{Z}}_{t+1}; \tilde{\boldsymbol{w}}, \boldsymbol{\lambda}) = \tilde{\boldsymbol{\phi}}_t(\tilde{\boldsymbol{Z}}_{t+1}, \tilde{\boldsymbol{w}}; \boldsymbol{\lambda}) - \sum_{t'=0}^{t}\mathbb{E}[\langle \partial_{t'}\tilde{\boldsymbol{\phi}}_t\rangle]\tilde{\boldsymbol{z}}_{t'}$ is separable and Lipschitz-continuous with respect to $\tilde{\boldsymbol{Z}}_{t+1}$, we can use Lemma 2 to obtain

$$\begin{aligned}&\frac{1}{N}\mathbb{E}\left[\tilde{\boldsymbol{z}}_{\tau'}^{\mathrm{T}}\left(\tilde{\boldsymbol{\phi}}_t - \sum_{t'=0}^{t}\mathbb{E}\left[\left\langle \partial_{t'}\tilde{\boldsymbol{\phi}}_t\right\rangle\right]\tilde{\boldsymbol{z}}_{t'}\right)\right] \\ =&\frac{1}{N}\sum_{n=1}^{N}\sum_{t'=0}^{t}\mathbb{E}[\tilde{z}_{\tau',n}\tilde{z}_{t,n}]\mathbb{E}\left[\partial_{t'}\tilde{\phi}_{t,n}\right] \\ &- \sum_{t'=0}^{t}\mathbb{E}\left[\left\langle \partial_{t'}\tilde{\boldsymbol{\phi}}_t\right\rangle\right]\frac{\mathbb{E}[\tilde{\boldsymbol{z}}_{\tau'}^{\mathrm{T}}\tilde{\boldsymbol{z}}_{t'}]}{N} = 0.\end{aligned} \tag{137}$$

Thus, (96) holds for $\tau = t$. ∎

Proof of Properties (A4) and (A5) for $\tau = t$: Repeat the proofs of Properties (A4) and (A5) for $\tau = 0$. In particular, see [46, p. 378] for the proof of (98) for $\tau = t$. ∎

APPENDIX B
PROOF OF THEOREM 2

In evaluating the derivative in $g_{t',t}^{(j)}$, the parameter ξ_t requires a careful treatment since it depends on $\boldsymbol{B}_{t+1}$ via $\boldsymbol{h}_t$. If the general error model contained the error model of the CAMP, we could use (28) in Theorem 1 to prove that ξ_t converges almost surely to a $\boldsymbol{B}_{t+1}$-independent constant $\bar{\xi}_t$ in the large system limit. To use Theorem 1, however, we have to prove the inclusion of the CAMP error model into the general error model. To circumvent this dilemma, we prove $g_{t-\tau,t}^{(j)} \overset{\text{a.s.}}{=} \xi_{t-\tau}^{(t-1)} g_{\tau}^{(j)} + o(1)$ for all t and $\tau = 0, \ldots, t$ by induction.

We consider the case $\tau = 0$, in which the expression (41) requires no special treatments in computing the derivative. Differentiating (41) with respect to the tth variable yields

$$g_{t,t}^{(j)} = \mu_{j+1} - \mu_j, \tag{138}$$

where μ_j denotes the jth moment (42) of the asymptotic eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. Comparing (43) and (138), we have $g_{t,t}^{(j)} = g_0^{(j)}$ for all t.

Suppose that there is some $t > 0$ such that $g_{t'-\tau,t'}^{(j)} \overset{\text{a.s.}}{=} \xi_{t'-\tau}^{(t'-1)} g_{\tau}^{(j)} + o(1)$ is correct for all $t' < t$ and $\tau = 0, \ldots, t'$. Then, (28) in Theorem 1 implies that $\xi_{t'}$ converges almost surely to a constant $\bar{\xi}_{t'}$ for any $t' < t$. We need to prove $g_{t-\tau,t}^{(j)} \overset{\text{a.s.}}{=} \xi_{t-\tau}^{(t-1)} g_{\tau}^{(j)} + o(1)$ for all $\tau = 0, \ldots, t$.

We first consider the case $\tau = 1$ since we have already proved the case $\tau = 0$. Differentiating (41) with respect to the $(t-1)$th variable yields

$$\begin{aligned} g_{t-1,t}^{(j)} = {} & \bar{\xi}_{t-1}(g_{t-1,t-1}^{(j)} - g_{t-1,t-1}^{(j+1)}) - \bar{\xi}_{t-1} g_1 (g_{t-1,t-1}^{(j)} + \mu_j) \\ & + \bar{\xi}_{t-1}\theta_1 (g_{t-1,t-1}^{(j+1)} + \mu_{j+1}). \end{aligned} \tag{139}$$

Using $g_{t,t}^{(j)} = g_0^{(j)}$ and (44), we arrive at $g_{t-1,t}^{(j)} \overset{\text{a.s.}}{=} \xi_{t-1} g_1^{(j)} + o(1)$.

We next consider the case $\tau > 1$. Differentiating (41) with respect to the $(t-\tau)$th variable, we have

$$\begin{aligned} g_{t-\tau,t}^{(j)} &= \bar{\xi}_{t-1}(g_{t-\tau,t-1}^{(j)} - g_{t-\tau,t-1}^{(j+1)}) \\ &+ \sum_{\tau'=t-\tau}^{t-1} \bar{\xi}_{\tau'}^{(t-1)} (\theta_{t-\tau'} g_{t-\tau,\tau'}^{(j+1)} - g_{t-\tau'} g_{t-\tau,\tau'}^{(j)}) \\ &- \sum_{\tau'=t-\tau+1}^{t-1} \bar{\xi}_{\tau'-1}^{(t-1)} (\theta_{t-\tau'} g_{t-\tau,\tau'-1}^{(j+1)} - g_{t-\tau'} g_{t-\tau,\tau'-1}^{(j)}) \\ &+ \bar{\xi}_{t-\tau}^{(t-1)} (\theta_\tau \mu_{j+1} - g_\tau \mu_j). \end{aligned} \tag{140}$$

Using (45) and the induction hypothesis $g_{t'-\tau,t'}^{(j)} \overset{\text{a.s.}}{=} \xi_{t'-\tau}^{(t'-1)} g_{\tau}^{(j)} + o(1)$ for all $t' < t$ and $\tau = 0, \ldots, t'$, we find $g_{t-\tau,t}^{(j)} \overset{\text{a.s.}}{=} \xi_{t-\tau}^{(t-1)} g_{\tau}^{(j)} + o(1)$.

Appendix C
Proof of Theorem 3

Let $G(x,z)$ denote the generating function of $\{g_\tau^{(j)}\}$ given by

$$G(x,z) = \sum_{j=0}^{\infty} G_j(z) x^j, \tag{141}$$

with

$$G_j(z) = \sum_{\tau=0}^{\infty} g_\tau^{(j)} z^{-\tau}. \tag{142}$$

It is possible to prove that $G(x,z)$ is given by

$$G(x,z) = \frac{\{\Theta(z) - xG(z)\}\eta(-x) - \Theta(z)}{x\tilde{G}(z) + 1 - \tilde{\Theta}(z)}, \tag{143}$$

with $\tilde{G}(z) = (1 - z^{-1})G(z)$ and $\tilde{\Theta}(z) = (1 - z^{-1})\Theta(z)$. Let $-x^*$ denote a pole of the generating function, i.e. $x^* = [1 - \tilde{\Theta}(z)]/\tilde{G}(z)$. Since the generating function is analytical, the numerator of (143) at $x = -x^*$ must be zero.

$$\{\Theta(z) + x^* G(z)\}\eta(x^*) - \Theta(z) = 0, \tag{144}$$

which is equivalent to (52).

To complete the proof of Theorem 3, we prove (143). The proof is a simple exercise of the Z-transform. We first compute $G_j(z)$ given by

$$G_j(z) = g_0^{(j)} + g_1^{(j)} z^{-1} + \sum_{\tau=2}^{\infty} g_\tau^{(j)} z^{-\tau}. \tag{145}$$

To evaluate the last term with (45), we note

$$\sum_{\tau=2}^{\infty} g_{\tau-1}^{(j)} z^{-\tau} = z^{-1} \sum_{\tau=1}^{\infty} g_\tau^{(j)} z^{-\tau} = z^{-1}\left\{ G_j(z) - g_0^{(j)} \right\}, \tag{146}$$

$$\begin{aligned} &\sum_{\tau=2}^{\infty}\sum_{\tau'=0}^{\tau-1} g_{\tau-\tau'} g_{\tau'}^{(j)} z^{-\tau} \\ &= g_0^{(j)} \sum_{\tau=2}^{\infty} g_\tau z^{-\tau} + \sum_{\tau'=1}^{\infty}\sum_{\tau=\tau'+1}^{\infty} g_{\tau-\tau'} g_{\tau'}^{(j)} z^{-\tau} \\ &= [G(z) - 1]\, G_j(z) - g_1 g_0^{(j)} z^{-1}, \end{aligned} \tag{147}$$

$$\begin{aligned} &\sum_{\tau=2}^{\infty}\sum_{\tau'=1}^{\tau-1} g_{\tau-\tau'} g_{\tau'-1}^{(j)} z^{-\tau} \\ &= \sum_{\tau'=1}^{\infty}\sum_{\tau=\tau'+1}^{\infty} g_{\tau-\tau'} g_{\tau'-1}^{(j)} z^{-\tau} \\ &= [G(z) - 1]\, z^{-1} G_j(z). \end{aligned} \tag{148}$$

Combining (43), (44), (45), and these results, we arrive at

$$\begin{aligned} G_j(z) = {} & [1 - \tilde{G}(z)]G_j(z) - [1 - \tilde{\Theta}(z)]G_{j+1}(z) \\ & - \mu_j G(z) + \mu_{j+1}\Theta(z). \end{aligned} \tag{149}$$

We next evaluate $G(x,z)$. Substituting (149) into the definition of $G(x,z)$ yields

$$\begin{aligned} G(x,z) = {} & [1 - \tilde{G}(z)]G(x,z) - [1 - \tilde{\Theta}(z)]\tfrac{G(x,z)}{x} \\ & - \eta(-x)G(z) + \tfrac{\eta(-x)-1}{x}\Theta(z), \end{aligned} \tag{150}$$

where we have used the definition (50) and the identity $G_0(z) = 0$ obtained from Theorem 2. Solving this equation with respect to $G(x,z)$, we obtain (143).

Appendix D
Proof of Theorem 4

A. SE Equations

The proof of Theorem 4 consists of four steps: A first step is a derivation of the SE equations, which is a dynamical system that describes the dynamics of five variables with three indices. A second step is evaluation of the generating functions for the five variables. The step is a simple exercise of the Z-transform. In a third step, we evaluate the obtained generating functions at poles to prove the SE equation (75) in terms of the generating functions. The last step is a derivation of the SE equation (77) in time domain via the inverse Z-transform.

Let $a_{t',t}^{(j)} = N^{-1}\boldsymbol{m}_{t'}^{\mathrm{T}}\boldsymbol{\Lambda}^j\boldsymbol{m}_t$, $b_{t',t}^{(j)} = N^{-1}\boldsymbol{b}_{t'}^{\mathrm{T}}\boldsymbol{\Lambda}^j\boldsymbol{m}_t$, $c_{t',t} = N^{-1}\tilde{\boldsymbol{q}}_{t'}^{\mathrm{T}}\tilde{\boldsymbol{q}}_t$, $d_{t',t} = N^{-1}\boldsymbol{q}_{t'}^{\mathrm{T}}\boldsymbol{q}_t$, and $e_t^{(j)} = N^{-1}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^j\boldsymbol{m}_t$.

Theorem 2 implies the asymptotic orthogonality between $\boldsymbol{b}_{t'}$ and $\boldsymbol{m}_t$. We use the definition (41) to obtain

$$\begin{aligned} a_{t',t}^{(j)} \overset{\text{a.s.}}{=} & b_{t,t'}^{(j)} - b_{t,t'}^{(j+1)} + \bar{\xi}_{t-1}(a_{t',t-1}^{(j)} - a_{t',t-1}^{(j+1)}) + e_{t'}^{(j)} \\ & + \sum_{\tau=0}^{t-1} \bar{\xi}_{\tau}^{(t-1)} \theta_{t-\tau}(a_{t',\tau}^{(j+1)} - b_{\tau,t'}^{(j+1)} - \bar{\xi}_{\tau-1} a_{t',\tau-1}^{(j+1)}) \\ & - \sum_{\tau=0}^{t-1} \bar{\xi}_{\tau}^{(t-1)} g_{t-\tau}(a_{t',\tau}^{(j)} - b_{\tau,t'}^{(j)} - \bar{\xi}_{\tau-1} a_{t',\tau-1}^{(j)}) + o(1), \end{aligned} \tag{151}$$

where we have replaced ξ_t with the asymptotic value $\bar{\xi}_t$. Applying (31) in Theorem 1 and (9) yields

$$\begin{aligned} b_{t',t}^{(j)} \overset{\text{a.s.}}{=} & (\mu_j - \mu_{j+1}) c_{t',t} + \bar{\xi}_{t-1}(b_{t',t-1}^{(j)} - b_{t',t-1}^{(j+1)}) + o(1) \\ & + \sum_{\tau=0}^{t-1} \bar{\xi}_{\tau}^{(t-1)} \theta_{t-\tau}(b_{t',\tau}^{(j+1)} - \mu_{j+1} c_{t',\tau} - \bar{\xi}_{\tau-1} b_{t',\tau-1}^{(j+1)}) \\ & - \sum_{\tau=0}^{t-1} \bar{\xi}_{\tau}^{(t-1)} g_{t-\tau}(b_{t',\tau}^{(j)} - \mu_j c_{t',\tau} - \bar{\xi}_{\tau-1} b_{t',\tau-1}^{(j)}). \end{aligned} \tag{152}$$

Using (30) in Theorem 1, (36), and (11), we have

$$c_{t'+1,t+1} \overset{\text{a.s.}}{=} \frac{\boldsymbol{q}_{t'+1}^{\mathrm{T}} \tilde{\boldsymbol{q}}_{t+1}}{N} + o(1) \overset{\text{a.s.}}{=} d_{t'+1,t+1} - \bar{\xi}_t \bar{\xi}_{t'} a_{t',t}^{(0)} + o(1). \tag{153}$$

Applying (26) in Theorem 1 yields

$$d_{t'+1,t+1} \overset{\text{a.s.}}{\to} \mathbb{E}\left[\{f_{t'}(x_1 + z_{t'}) - x_1\}\{f_t(x_1 + z_t) - x_1\}\right], \tag{154}$$

where $\{z_t\}$ are zero-mean Gaussian random variables with covariance $\mathbb{E}[z_{t'} z_t] = a_{t',t}^{(0)}$. Finally, we use (31) in Theorem 1 to obtain

$$\begin{aligned} e_t^{(j)} \overset{\text{a.s.}}{=} & \bar{\xi}_{t-1}(e_{t-1}^{(j)} - e_{t-1}^{(j+1)}) + \sigma^2 \mu_{j+1} + o(1) \\ & + \sum_{\tau=0}^{t-1} \bar{\xi}_{\tau}^{(t-1)} \theta_{t-\tau}(e_{\tau}^{(j+1)} - \bar{\xi}_{\tau-1} e_{\tau-1}^{(j+1)}) \\ & - \sum_{\tau=0}^{t-1} \bar{\xi}_{\tau}^{(t-1)} g_{t-\tau}(e_{\tau}^{(j)} - \bar{\xi}_{\tau-1} e_{\tau-1}^{(j)}). \end{aligned} \tag{155}$$

To transform the summations in these equations to convolution, we use the change of variables $a_{t',t}^{(j)} = \bar{\xi}_0^{(t'-1)} \bar{\xi}_0^{(t-1)} \tilde{a}_{t',t}^{(j)}$. Similarly, we define $\tilde{b}_{t',t}^{(j)}$, $\tilde{c}_{t',t}$, and $\tilde{d}_{t',t}$ while we use $e_{t'}^{(j)} = \bar{\xi}_0^{(t'-1)} \bar{\xi}_0^{(t-1)} \tilde{e}_{t',t}^{(j)}$. Then, the SE equations (151)–(155) reduce to

$$\begin{aligned} \tilde{a}_{t',t}^{(j)} \overset{\text{a.s.}}{=} & \tilde{b}_{t,t'}^{(j)} - \tilde{b}_{t,t'}^{(j+1)} + \tilde{a}_{t',t-1}^{(j)} - \tilde{a}_{t',t-1}^{(j+1)} + \tilde{e}_{t',t}^{(j)} \\ & + \sum_{\tau=0}^{t-1} \theta_{t-\tau}(\tilde{a}_{t',\tau}^{(j+1)} - \tilde{b}_{\tau,t'}^{(j+1)} - \tilde{a}_{t',\tau-1}^{(j+1)}) \\ & - \sum_{\tau=0}^{t-1} g_{t-\tau}(\tilde{a}_{t',\tau}^{(j)} - \tilde{b}_{\tau,t'}^{(j)} - \tilde{a}_{t',\tau-1}^{(j)}) + o(1), \end{aligned} \tag{156}$$

$$\begin{aligned} \tilde{b}_{t',t}^{(j)} \overset{\text{a.s.}}{=} & (\mu_j - \mu_{j+1}) \tilde{c}_{t',t} + \tilde{b}_{t',t-1}^{(j)} - \tilde{b}_{t',t-1}^{(j+1)} + o(1) \\ & + \sum_{\tau=0}^{t-1} \theta_{t-\tau}(\tilde{b}_{t',\tau}^{(j+1)} - \mu_{j+1} \tilde{c}_{t',\tau} - \tilde{b}_{t',\tau-1}^{(j+1)}) \\ & - \sum_{\tau=0}^{t-1} g_{t-\tau}(\tilde{b}_{t',\tau}^{(j)} - \mu_j \tilde{c}_{t',\tau} - \tilde{b}_{t',\tau-1}^{(j)}), \end{aligned} \tag{157}$$

$$\tilde{c}_{t'+1,t+1} \overset{\text{a.s.}}{=} \tilde{d}_{t'+1,t+1} - \tilde{a}_{t',t}^{(0)} + o(1), \tag{158}$$

$$\begin{aligned} \tilde{e}_{t',t}^{(j)} \overset{\text{a.s.}}{=} & \tilde{e}_{t'-1,t}^{(j)} - \tilde{e}_{t'-1,t}^{(j+1)} + \mu_{j+1} \sigma_{t',t}^2 + o(1) \\ & + \sum_{\tau=0}^{t'-1} \theta_{t'-\tau}(\tilde{e}_{\tau,t}^{(j+1)} - \tilde{e}_{\tau-1,t}^{(j+1)}) \\ & - \sum_{\tau=0}^{t'-1} g_{t'-\tau}(\tilde{e}_{\tau,t}^{(j)} - \tilde{e}_{\tau-1,t}^{(j)}), \end{aligned} \tag{159}$$

with

$$\sigma_{t',t}^2 = \frac{\sigma^2}{\bar{\xi}_0^{(t'-1)} \bar{\xi}_0^{(t-1)}}. \tag{160}$$

In principle, it is possible to solve the coupled dynamical system (154), (156)–(159) numerically. However, numerical evaluation is a challenging task due to instability against numerical errors.

B. Generating Functions

We solve the coupled dynamical system via the Z-transform. Define the generating function of $\tilde{a}_{t',t}^{(j)}$ as

$$A(x, y, z) = \sum_{j=0}^{\infty} x^j A_j(y, z), \tag{161}$$

with

$$A_j(y, z) = \sum_{t',t=0}^{\infty} \tilde{a}_{t',t}^{(j)} y^{-t'} z^{-t}. \tag{162}$$

Similarly, we write the generating functions of $\{\tilde{b}_{t',t}^{(j)}\}$, $\{\tilde{c}_{t',t}\}$, $\{\tilde{d}_{t',t}\}$, $\{\tilde{e}_{t',t}^{(j)}\}$, and $\{\sigma_{t',t}^2\}$ as $B(x, y, z)$, $C(y, z)$, $D(y, z)$, $E(x, y, z)$, and $\Sigma(y, z)$, respectively.

To evaluate the generating function $A_j(y, z)$, we utilize

$$\begin{aligned} & \sum_{t'=0}^{\infty} y^{-t'} \sum_{t=1}^{\infty} z^{-t} \sum_{\tau=0}^{t-1} g_{t-\tau} \tilde{a}_{t',\tau-k}^{(j)} \\ = & \sum_{t'=0}^{\infty} y^{-t'} \sum_{\tau=0}^{\infty} \sum_{t=\tau+1}^{\infty} z^{-t} g_{t-\tau} \tilde{a}_{t',\tau-k}^{(j)} \\ = & z^{-k} \left[G(z) - 1\right] A_j(y, z) \end{aligned} \tag{163}$$

for any integer k, where we have used the definition (51) of $G(z)$. From (156), we have

$$\begin{aligned} A_j(y, z) \overset{\text{a.s.}}{=} & B_j(z, y) - B_{j+1}(z, y) + \frac{A_j(y, z)}{z} - \frac{A_{j+1}(y, z)}{z} \\ & + [\Theta(z) - 1] \left\{ A_{j+1}(y, z) - B_{j+1}(z, y) - \frac{A_{j+1}(y, z)}{z} \right\} \\ & - [G(z) - 1] \left\{ A_j(y, z) - B_j(z, y) - \frac{A_j(y, z)}{z} \right\} \\ & + E_j(y, z). \end{aligned} \tag{164}$$

Similarly, we can derive

$$\begin{aligned} B_j(y, z) \overset{\text{a.s.}}{=} & (\mu_j - \mu_{j+1}) C(y, z) + \frac{B_j(y, z)}{z} - \frac{B_{j+1}(y, z)}{z} \\ & + [\Theta(z) - 1] \left\{ B_{j+1}(y, z) - \mu_{j+1} C(y, z) - \frac{B_{j+1}(y, z)}{z} \right\} \\ & - [G(z) - 1] \left\{ B_j(y, z) - \mu_j C(y, z) - \frac{B_j(y, z)}{z} \right\} + o(1), \end{aligned} \tag{165}$$

$$C(y,z) \overset{\text{a.s.}}{=} D(y,z) - (yz)^{-1}A_0(y,z) + o(1), \tag{166}$$

$$E_j(y,z) \overset{\text{a.s.}}{=} \frac{E_j(y,z)}{y} - \frac{E_{j+1}(y,z)}{y} + \mu_{j+1}\Sigma(y,z) + o(1) + (1-y^{-1})[\Theta(y)-1]E_{j+1}(y,z) - (1-y^{-1})[G(y)-1]E_j(y,z). \tag{167}$$

We next substitute (164) into (161) to obtain

$$\left\{x\tilde{G}(z) + 1 - \tilde{\Theta}(z)\right\} A(x,y,z) \overset{\text{a.s.}}{=} [1-\tilde{\Theta}(z)]A_0(y,z) + \left\{x\tilde{G}(z) - \tilde{\Theta}(z)\right\} \frac{B(x,z,y)}{1-z^{-1}} + xE(x,y,z) + o(1), \tag{168}$$

with $\tilde{G}(z) = (1-z^{-1})G(z)$ and $\tilde{\Theta}(z) = (1-z^{-1})\Theta(z)$, where we have used the identity $B_0(y,z) \overset{\text{a.s.}}{=} o(1)$ obtained from the asymptotic orthogonality between $\boldsymbol{b}_{t'}$ and $\boldsymbol{m}_t$. Similarly, we use (50) and (165) to obtain

$$B(x,y,z) \overset{\text{a.s.}}{=} \frac{[x\tilde{G}(z)-\tilde{\Theta}(z)]\eta(-x)+\tilde{\Theta}(z)}{x\tilde{G}(z)+1-\tilde{\Theta}(z)} \frac{C(y,z)}{1-z^{-1}} + o(1). \tag{169}$$

Furthermore, we have

$$E(x,y,z) \overset{\text{a.s.}}{=} \frac{1-\tilde{\Theta}(y)}{x\tilde{G}(y)+1-\tilde{\Theta}(y)}E_0(y,z) + \frac{\eta(-x)-1}{x\tilde{G}(y)+1-\tilde{\Theta}(y)}\Sigma(y,z) + o(1). \tag{170}$$

C. Evaluation at Poles

The equations (166), (168), (169), and (170) provide all information about the generating functions. However, we are interested only in those at $x=0$. To extract this information, we focus on the poles of $A(x,y,z)$ and $E(x,y,z)$. Let $-x^*$ denote the pole of $A(x,y,z)$ given by

$$x^* = \frac{1-\tilde{\Theta}(z)}{\tilde{G}(z)}. \tag{171}$$

Since $A(x,y,z)$ is analytical, the RHS of (168) has to be zero at $x=-x^*$.

$$\frac{B(-x^*,z,y)}{1-z^{-1}} \overset{\text{a.s.}}{=} [1-\tilde{\Theta}(z)]A_0(y,z) - x^*E(-x^*,y,z) + o(1). \tag{172}$$

Similarly, we use (170) and Theorem 3 to obtain

$$E_0(y,z) \overset{\text{a.s.}}{=} \Sigma(y,z) + o(1). \tag{173}$$

Thus, (170) reduces to

$$E(-x^*,y,z) \overset{\text{a.s.}}{=} \frac{[\tilde{\Theta}(z)-\tilde{\Theta}(y)]\tilde{G}(z)\Sigma(y,z)}{\tilde{G}(y)\tilde{\Theta}(z) - \tilde{\Theta}(y)\tilde{G}(z) + \tilde{G}(z) - \tilde{G}(y)} + o(1). \tag{174}$$

Evaluating $B(x,z,y)$ given via (169) at $x=-x^*$ yields

$$\frac{B(-x^*,z,y)}{1-z^{-1}} \overset{\text{a.s.}}{=} \frac{\Theta(y)G(z)-G(y)\Theta(z)}{\tilde{G}(y)\tilde{\Theta}(z)-\tilde{\Theta}(y)\tilde{G}(z)+\tilde{G}(z)-\tilde{G}(y)} \cdot [1-\tilde{\Theta}(z)]C(y,z) + o(1), \tag{175}$$

where we have used $\tilde{\Theta}(z) = (1-z^{-1})\Theta(z)$, $\tilde{G}(z) = (1-z^{-1})G(z)$, and the symmetry $C(z,y) = C(y,z)$.

Substituting (166), (174), and (175) into (172), we obtain

$$F_{G,\Theta}(y,z)A_0(y,z) \overset{\text{a.s.}}{=} \frac{\Theta(y)G(z)-G(y)\Theta(z)}{y^{-1}-z^{-1}}D(y,z) + \frac{(1-z^{-1})\Theta(z)-(1-y^{-1})\Theta(y)}{y^{-1}-z^{-1}}\Sigma(y,z) + o(1), \tag{176}$$

with

$$F_{G,\Theta}(y,z) = \frac{(y^{-1}+z^{-1}-1)[\Theta(y)G(z)-G(y)\Theta(z)]}{y^{-1}-z^{-1}} + \frac{(1-z^{-1})G(z)-(1-y^{-1})G(y)}{y^{-1}-z^{-1}}. \tag{177}$$

We transform the SE equation (176) into another generating-function representation that is suited for deriving time-domain representation. Let S denote the generating function of some sequence $\{s_t\}$. We use the notations $S_1(z) = z^{-1}S(z)$, Δ_S, and Δ_{S_1}, given by

$$\Delta_S = \frac{S(y)-S(z)}{y^{-1}-z^{-1}}, \tag{178}$$

which is a function of y and z. The inverse Z-transform of these generating functions can be evaluated straightforwardly, as shown shortly. We use these notations to re-write the SE equation (176) as

$$F_{G,\Theta}(y,z)A_0(y,z) \overset{\text{a.s.}}{=} \{G(z)\Delta_\Theta - \Theta(z)\Delta_G\}D(y,z) + (\Delta_{\Theta_1} - \Delta_\Theta)\Sigma(y,z) + o(1), \tag{179}$$

with

$$F_{G,\Theta}(y,z) = (y^{-1}+z^{-1}-1)[G(z)\Delta_\Theta - \Theta(z)\Delta_G] + \Delta_{G_1} - \Delta_G, \tag{180}$$

where $G_1(z) = z^{-1}G(z)$ and $\Theta_1(z) = z^{-1}\Theta(z)$ are defined in the same manner as in $S_1(z)$. The SE equation (179) is equivalent to the former statement in Theorem 4.

D. Time-Domain Representation

We transform the SE equation (179) into a time-domain representation that is suitable for numerical evaluation. Suppose that $G(z)$ is represented as $G(z) = P(z)/Q(z)$. Let $R(z)$ denote the generating function of $\{r_t\}$, i.e. $R(z) = Q(z)\Theta(z)$. We multiply both sides of the SE equation (179) by $Q(y)Q(z)$ to obtain

$$F_{P,Q,\Theta}(y,z)A_0(y,z) \overset{\text{a.s.}}{=} \{P(z)\Delta_R - R(z)\Delta_P\}D(y,z) + Q(y)Q(z)(\Delta_{\Theta_1} - \Delta_\Theta)\Sigma(y,z) + o(1), \tag{181}$$

with

$$F_{P,Q,\Theta}(y,z) = [\Delta_{P_1} - \Delta_P]Q(z) + (1-z^{-1})P(z)\Delta_Q + (z^{-1}-1)[P(z)\Delta_R - R(z)\Delta_P] + y^{-1}[P(z)\Delta_R - R(z)\Delta_P]. \tag{182}$$

It is possible to evaluate the inverse Z-transform of $S_1(z)$, Δ_S, Δ_{S_1}, and $z^{-1}\Delta_S$ for any generating function $S(z)$. By definition, we have

$$S_1(z) = \sum_{t=0}^{\infty} s_t z^{-(t+1)} = \sum_{t=0}^{\infty} s_{t-1}z^{-t}, \tag{183}$$

TABLE III
Z-TRANSFORM OF 2-DIMENSIONAL ARRAYS

Array $s_{t',t}$	Z-transform
$\delta_{t',0}s_{t-1}$	$S_1(z)$
$s_{t'+t+1}$	Δ_S
$s_{t'+t}$	Δ_{S_1}
$s_{t'+t} - \delta_{t',0}s_t$	$y^{-1}\Delta_S$

where the convention $s_{-1} = 0$ has been used. Thus, $S_1(z)$ is the generating function of the sequence $\{s_{t-1}\}$.

For Δ_S, we obtain

$$\Delta_S = \sum_{\tau=1}^{\infty} s_\tau \frac{y^{-\tau} - z^{-\tau}}{y^{-1} - z^{-1}} = \sum_{\tau=1}^{\infty}\sum_{\tau'=0}^{\tau-1} s_\tau y^{-\tau'} z^{-(\tau-\tau'-1)}$$
$$= \sum_{\tau'=0}^{\infty}\sum_{\tau=\tau'+1}^{\infty} s_\tau y^{-\tau'} z^{-(\tau-\tau'-1)} = \sum_{\tau'=0}^{\infty}\sum_{\tau=0}^{\infty} s_{\tau'+\tau+1} y^{-\tau'} z^{-\tau}, \tag{184}$$

which implies that Δ_S is the generating function of the two-dimensional array $s_{t',t} = s_{t'+t+1}$.

We combine these results to evaluate the inverse Z-transform of the remaining generating functions. For $S_1(z)$, Δ_{S_1} is the generating function of $\{s_{t'+t}\}$. Since y^{-1} is the generating function of $\delta_{t',1}\delta_{t,0}$ and since Δ_S is the generating function of $s_{t',t} = s_{t'+t+1}$, $y^{-1}\Delta_S$ is the generating function of the two-dimensional convolution:

$$(\delta_{t',1}\delta_{t,0}) * s_{t',t} = s_{t'-1,t} = s_{t'+t} - \delta_{t',0}s_t, \tag{185}$$

where the last expression is due to the convention $s_{-1,t} = 0$. See Table III for a summary of these results.

We evaluate the inverse Z-transform of (181). It is a simple exercise to confirm that (181) is equal to the Z-transform of the following difference equation:

$$\mathfrak{D}_{t',t} * \tilde{a}^{(0)}_{t',t} \overset{\text{a.s.}}{=} (p_t * r_{t'+t+1} - r_t * p_{t'+t+1}) * \tilde{d}_{t',t}$$
$$+ (q_{t'}q_t) * (\theta_{t'+t} - \theta_{t'+t+1}) * \sigma^2_{t',t} + o(1), \tag{186}$$

with

$$\begin{aligned}\mathfrak{D}_{t',t} = \ & (p_{t'+t} - p_{t'+t+1}) * q_t + (p_t - p_{t-1}) * q_{t'+t+1} \\ + \ & (p_{t-1} - p_t) * r_{t'+t+1} + (r_t - r_{t-1}) * p_{t'+t+1} \\ + \ & p_t * (r_{t'+t} - \delta_{t',0}r_t) - r_t * (p_{t'+t} - \delta_{t',0}p_t),\end{aligned} \tag{187}$$

where all variables with negative indices are set to zero. Multiplying (186) by $\bar{\xi}_0^{(t'-1)}\bar{\xi}_0^{(t-1)}$ and using the definitions $\tilde{a}^{(0)}_{\tau',\tau} = a^{(0)}_{\tau',\tau}/(\bar{\xi}_0^{(\tau'-1)}\bar{\xi}_0^{(\tau-1)})$, $\tilde{d}^{(0)}_{\tau',\tau} = d_{\tau',\tau}/(\bar{\xi}_0^{(\tau'-1)}\bar{\xi}_0^{(\tau-1)})$, and $\sigma^2_{\tau',\tau} = \sigma^2/(\bar{\xi}_0^{(\tau'-1)}\bar{\xi}_0^{(\tau-1)})$, we arrive at the SE equation (77) in time domain, with the superscript in $a^{(0)}_{\tau',\tau}$ omitted.

Finally, we use the notational convention $f_{-1}(\cdot) = 0$ to obtain initial and boundary conditions. From the definition (70) of $d_{t'+1,t+1}$, we have the initial condition $d_{0,0} = \mathbb{E}[x_1^2] = 1$. Similarly, we use (70) to obtain the boundary condition $d_{0,\tau+1} = -\mathbb{E}[x_1\{f_\tau(x_1 + z_\tau) - x_1\}]$. The boundary condition $d_{\tau+1,0} = d_{0,\tau+1}$ follows from the symmetry.

APPENDIX E
PROOF OF THEOREM 5

Without the loss of generality, we assume $p_t = g_t$ and $q_t = \delta_{t,0}$. Then, the SE equation (77) in time domain reduces to

$$\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} \bar{\xi}^{(t'-1)}_{t'-\tau'}\bar{\xi}^{(t-1)}_{t-\tau}\Big\{\mathfrak{D}_{\tau',\tau}a_{t'-\tau',t-\tau}$$
$$-(g_\tau * \theta_{\tau'+\tau+1} - \theta_\tau * g_{\tau'+\tau+1})d_{t'-\tau',t-\tau}$$
$$-\sigma^2(\theta_{\tau'+\tau} - \theta_{\tau'+\tau+1})\Big\} = 0, \tag{188}$$

with

$$\begin{aligned}\mathfrak{D}_{\tau',\tau} =& g_{\tau'+\tau} - g_{\tau'+\tau+1} + (g_{\tau-1} - g_\tau) * \theta_{\tau'+\tau+1} \\ &+ (\theta_\tau - \theta_{\tau-1}) * g_{\tau'+\tau+1} + g_\tau * (\theta_{\tau'+\tau} - \delta_{\tau',0}\theta_\tau) \\ &- \theta_\tau * (g_{\tau'+\tau} - \delta_{\tau',0}g_\tau).\end{aligned} \tag{189}$$

We evaluate a fixed-point of the reduced SE equation (188) for the Bayes-optimal denoiser f_{opt}. Suppose that $\lim_{t',t\to\infty} a_{t',t} = a_{\text{s}}$, $\lim_{t',t\to\infty} d_{t',t} = d_{\text{s}}$, and $\lim_{t\to\infty}\bar{\xi}_t = \xi_{\text{s}}$ hold. The main feature of the Bayes-optimal denoiser is the identity $\xi_{\text{s}} = d_{\text{s}}/a_{\text{s}}$ [46, Lemma 2]. We use this identity and the assumptions in Theorem 5 to prove the fixed-point (80).

Taking the limits $t', t \to \infty$ in (188) yields

$$a_{\text{s}}\sum_{\tau',\tau=0}^{\infty}\mathfrak{D}_{\tau',\tau}(\xi_{\text{s}}^{-1})^{-\tau'-\tau}$$
$$= d_{\text{s}}\sum_{\tau',\tau=0}^{\infty}(g_\tau * \theta_{\tau'+\tau+1} - \theta_\tau * g_{\tau'+\tau+1})(\xi_{\text{s}}^{-1})^{-\tau'-\tau}$$
$$+ \sigma^2\sum_{\tau',\tau=0}^{\infty}(\theta_{\tau'+\tau} - \theta_{\tau'+\tau+1})(\xi_{\text{s}}^{-1})^{-\tau'-\tau}. \tag{190}$$

We use the properties of the Z-transform in Table III and the identity $\xi_{\text{s}} = d_{\text{s}}/a_{\text{s}}$ to find

$$F_{G,\Theta}(y,z)\frac{d_{\text{s}}}{\xi_{\text{s}}} = \{G(z)\Delta_\Theta - \Theta(z)\Delta_G\}d_{\text{s}} + (\Delta_{\Theta_1} - \Delta_\Theta)\sigma^2 \tag{191}$$

in the limit $y, z \to \xi_{\text{s}}^{-1}$, where $F_{G,\Theta}$ is given by (76).

Series-expanding Δ_S with respect to z^{-1} at $z = y$ up to the first order yields

$$\lim_{y,z\to\xi_{\text{s}}^{-1}}\Delta_S = \frac{dS}{dz^{-1}}(\xi_{\text{s}}^{-1}). \tag{192}$$

Similarly, we have

$$\lim_{y,z\to\xi_{\text{s}}^{-1}}\Delta_{S_1} = S(\xi_{\text{s}}^{-1}) + \xi_{\text{s}}\frac{dS}{dz^{-1}}(\xi_{\text{s}}^{-1}), \tag{193}$$

Applying these results to (191) with (76) yields

$$\left\{1 + (\xi_{\text{s}} - 1)\frac{d\Theta}{dz^{-1}}(\xi_{\text{s}}^{-1})\right\}\left\{\frac{G(\xi_{\text{s}}^{-1})d_{\text{s}}}{\xi_{\text{s}}} - \sigma^2\right\} = 0, \tag{194}$$

where we have used the assumption $\Theta(\xi_{\text{s}}^{-1}) = 1$. Since $1 + (\xi_{\text{s}} - 1)d\Theta(\xi_{\text{s}}^{-1})/(dz^{-1}) \neq 0$ has been assumed, we arrive at

$$\frac{G(\xi_{\text{s}}^{-1})}{\xi_{\text{s}}} = \frac{\sigma^2}{d_{\text{s}}}. \tag{195}$$

To prove the fixed-point (80), we use the relationship (55) between the η-transform and the R-transform. Evaluating (55) at $x = x^*$ given in (171) and using Theorem 3, we obtain

$$G(z) = \Theta(z) R\left(-\frac{1-(1-z^{-1})\Theta(z)}{G(z)}\Theta(z)\right). \quad (196)$$

Letting $z = \xi_{\mathrm{s}}^{-1}$ and applying the assumption $\Theta(\xi_{\mathrm{s}}^{-1}) = 1$ yields

$$G(\xi_{\mathrm{s}}^{-1}) = R\left(-\frac{\xi_{\mathrm{s}}}{G(\xi_{\mathrm{s}}^{-1})}\right). \quad (197)$$

Substituting (195) into this identity and using $\xi_{\mathrm{s}} = d_{\mathrm{s}}/a_{\mathrm{s}}$, we arrive at

$$a_{\mathrm{s}} = \frac{\sigma^2}{R(-d_{\mathrm{s}}/\sigma^2)}. \quad (198)$$

Appendix F
Evaluation of (70) for Bernoulli-Gaussian Signals

A. Summary

We evaluate the correlation (70) for the Bernoulli-Gaussian signals. This appendix is organized as an independent section of the other parts. Thus, we use different notations from the other parts.

Let $A \in \{0, 1\}$ denote a Bernoulli random variable taking 1 with probability $\rho \in [0, 1]$. Suppose that $Z \sim \mathcal{N}(0, \rho^{-1})$ is independent of A and a zero-mean Gaussian random variable with variance ρ^{-1}. We consider estimation of a Bernoulli-Gaussian signal $X = AZ$ on the basis of two dependent noisy observations,

$$Y_{t'} = X + W_{t'}, \quad Y_t = X + W_t, \quad (199)$$

with

$$\begin{pmatrix} W_{t'} \\ W_t \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} a_{t',t'} & a_{t',t} \\ a_{t',t} & a_{t,t} \end{pmatrix}, \quad (200)$$

where $\mathbf{\Sigma}$ is positive definite. The goal of this appendix is to evaluate the correlation $d_{t'+1,t+1}$ of the estimation errors for the Bayes-optimal denoiser $f_{\mathrm{opt}}(Y_t; a_{t,t}) = \mathbb{E}[X|Y_t]$,

$$d_{t'+1,t+1} = \mathbb{E}[\{f_{\mathrm{opt}}(Y_{t'}; a_{t',t'}) - X\}\{f_{\mathrm{opt}}(Y_t; a_{t,t}) - X\}]. \quad (201)$$

Before presenting the derived expression of the correlation (201), we first introduce several definitions. We write the pdf of a zero-mean Gaussian random variable Y with variance σ^2 as $p_{\mathrm{G}}(y; \sigma^2)$, with

$$p_{\mathrm{G}}(y; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right). \quad (202)$$

The pdf of a Gaussian mixture is defined as

$$p_{\mathrm{GM}}(y; a_{t,t}) = \rho p_{\mathrm{G}}(y; \rho^{-1} + a_{t,t}) + (1-\rho) p_{\mathrm{G}}(y; a_{t,t}), \quad (203)$$

which is used to represent the marginal pdf of Y_t. As proved in Appendix F-B, the probability of $A = 1$ given Y_t is given by $\Pr(A = 1|Y_t = y) = \pi(y; a_{t,t})$, with

$$\pi(y, a_{t,t}) = \frac{\rho p_{\mathrm{G}}(y; \rho^{-1} + a_{t,t})}{p_{\mathrm{GM}}(y; a_{t,t})}. \quad (204)$$

The Bayes-optimal denoiser $f_{\mathrm{opt}}(Y_t; a_{t,t})$ is derived in the same appendix:

$$f_{\mathrm{opt}}(y; a_{t,t}) = \frac{y}{1 + \rho a_{t,t}} \pi(y, a_{t,t}), \quad (205)$$

where the conditional probability $\pi(y, a_{t,t})$ is given by (204).

We write the MSE function $\mathrm{MSE}(a_{t,t})$ as

$$\begin{aligned} &\mathrm{MSE}(a_{t,t}) \\ &= \frac{a_{t,t}}{1 + \rho a_{t,t}} + \mathbb{E}[\{1 - \pi(Y_t, a_{t,t})\}\{f_{\mathrm{opt}}(Y_t; a_{t,t})\}^2] \\ &+ \mathbb{E}\left[\pi(Y_t, a_{t,t})\left\{\frac{Y_t}{1 + \rho a_{t,t}} - f_{\mathrm{opt}}(Y_t; a_{t,t})\right\}^2\right], \end{aligned} \quad (206)$$

where the Bayes-optimal denoiser f_{opt} is given in (205). In (206), the expectation is over $Y_t \sim p_{\mathrm{GM}}(y; a_{t,t})$ given in (203).

The joint pdf of $\{Y_{t'}, Y_t\}$ is represented as

$$p(Y_{t'}, Y_t) = \rho p(Y_{t'}, Y_t|A = 1) + (1-\rho) p(Y_{t'}, Y_t|A = 0). \quad (207)$$

As proved in Appendix F-F, the conditional pdf $p(Y_{t'}, Y_t|A)$ is given by

$$\begin{aligned} p(Y_{t'}, Y_t|A = a) &= p_{\mathrm{G}}(Y_{t'}; \rho^{-1}a + a_{t',t'}) \\ \cdot p_{\mathrm{G}}&\left(Y_t - \tfrac{a + \rho a_{t',t}}{a + \rho a_{t',t'}} Y_{t'}; \tfrac{a + \rho a_{t,t}}{\rho} - \tfrac{(a + \rho a_{t',t})^2}{\rho(a + \rho a_{t',t'})}\right) \end{aligned} \quad (208)$$

for $a = 0, 1$.

Proposition 1:
- Let $\mathrm{MSE}(a_{t,t})$ denote the MSE function (206). Then,

$$d_{t+1,t+1} = \mathrm{MSE}(a_{t,t}). \quad (209)$$

- For $t' \neq t$, let

$$v_{t',t} = \frac{a_{t',t'} a_{t,t} - a_{t',t}^2}{a_{t',t'} + a_{t,t} - 2a_{t',t}}. \quad (210)$$

Then, the correlation $d_{t'+1,t+1}$ for $t' \neq t$ is given by

$$\begin{aligned} d_{t'+1,t+1} &= \mathbb{E}[f_{\mathrm{opt}}(Y_{t'}; a_{t',t'}) f_{\mathrm{opt}}(Y_t; a_{t,t})] \\ &+ \mathbb{E}\left[\pi(Y_{t',t}; v_{t',t})\left\{\left(\frac{Y_{t',t}}{1 + \rho v_{t',t}}\right)^2 + \frac{\rho^{-1} v_{t',t}}{\rho^{-1} + v_{t',t}}\right.\right. \\ &\left.\left. - \frac{Y_{t',t}[f_{\mathrm{opt}}(Y_{t'}; a_{t',t'}) + f_{\mathrm{opt}}(Y_t; a_{t,t})]}{1 + \rho v_{t',t}}\right\}\right], \end{aligned} \quad (211)$$

with

$$Y_{t',t} = \frac{(a_{t,t} - a_{t',t}) Y_{t'} + (a_{t',t'} - a_{t',t}) Y_t}{a_{t',t'} + a_{t,t} - 2a_{t',t}}, \quad (212)$$

where the expectation in (211) over $\{Y_{t'}, Y_t\}$ is evaluated via the joint pdf (207).

Proof: See from Appendix F-B to Appendix F-F. ■

Proposition 1 implies that $d_{t'+1,t+1}$ for $t' \neq t$ requires numerical computation of the double integrals.

B. Bayes-Optimal Denoiser

We compute the Bayes-optimal denoiser $f_{\mathrm{opt}}(Y_t; a_{t,t}) = \mathbb{E}[X|Y_t]$, given by

$$\begin{aligned} f_{\mathrm{opt}}(Y_t; a_{t,t}) &= \mathbb{E}\left[\mathbb{E}[AZ|Y_t, A]\middle| Y_t\right] \\ &= \mathbb{E}[Z|Y_t, A = 1]\Pr(A = 1|Y_t). \end{aligned} \quad (213)$$

Note that f_{opt} is different from the true posterior mean estimator (PME) $\mathbb{E}[X|Y_{t'}, Y_t]$.

We first evaluate the former factor $\mathbb{E}[Z|Y_t, A = 1]$. Since $Y_t = Z + W_t$ given $A = 1$ is the AWGN observation of

$Z \sim \mathcal{N}(0, \rho^{-1})$, we obtain the well-known LMMSE estimator

$$\mathbb{E}[Z|Y_t, A=1] = \frac{\rho^{-1}Y_t}{\rho^{-1}+a_{t,t}}, \tag{214}$$

which implies the Bayes-optimal denoiser (205).

We next prove that the latter factor $\Pr(A=1|Y_t)$ is equal to $\pi(Y_t; a_{t,t})$ given in (204). By definition,

$$\Pr(A=1|Y_t) = \frac{\rho p(Y_t|A=1)}{p(Y_t)}. \tag{215}$$

For the numerator, we have

$$\begin{aligned} p(Y_t|A=1) &= \mathbb{E}_Z[p(Y_t|A=1,Z)] \\ =&\mathbb{E}_Z[p_{\mathrm{G}}(Y_t - Z; a_{t,t})] = p_{\mathrm{G}}(Y_t; \rho^{-1}+a_{t,t}), \end{aligned} \tag{216}$$

where the last equality follows from the fact that $Z + W_t$ is a zero-mean Gaussian random variable with variance $\rho^{-1} + a_{t,t}$.

The denominator $p(Y_t)$ is computed in the same manner,

$$\begin{aligned} p(Y_t) &= \rho p(Y_t|A=1) + (1-\rho)p(Y_t|A=0) \\ =&\rho p_{\mathrm{G}}(Y_t; \rho^{-1}+a_{t,t}) + (1-\rho)p_{\mathrm{G}}(Y_t; a_{t,t}), \end{aligned} \tag{217}$$

which is equal to $p_{\mathrm{GM}}(Y_t; a_{t,t})$ given in (203). Combining these results, we arrive at $\Pr(A=1|Y_t) = \pi(Y_t; a_{t,t})$ given in (204).

C. MSE

To evaluate the MSE $d_{t+1,t+1} = \mathbb{E}[\{X - f_{\mathrm{opt}}(Y_t; a_{t,t})\}^2]$, we focus on the posterior variance $\mathbb{E}[\{X - f_{\mathrm{opt}}(Y_t; a_{t,t})\}^2|Y_t]$. By definition,

$$\begin{aligned} &\mathbb{E}[\{X - f_{\mathrm{opt}}(Y_t; a_{t,t})\}^2|Y_t] \\ =&\Pr(A=1|Y_t)\mathbb{E}[\{Z - f_{\mathrm{opt}}(Y_t; a_{t,t})\}^2|Y_t, A=1] \\ &+ \{1 - \Pr(A=1|Y_t)\}\{f_{\mathrm{opt}}(Y_t; a_{t,t})\}^2, \end{aligned} \tag{218}$$

with $\Pr(A=1|Y_t) = \pi(Y_t, a_{t,t})$ given in (204).

Let $\mathbb{E}[Z|Y_t, A=1]$ denote the PME of Z conditioned on Y_t and $A=1$, given in (214). The conditional expectation in the first term can be evaluated as follows:

$$\begin{aligned} &\mathbb{E}[\{Z - f_{\mathrm{opt}}(Y_t; a_{t,t})\}^2|Y_t, A=1] = \mathbb{E}[\{Z - \mathbb{E}[Z|Y_t, A=1] \\ &+\mathbb{E}[Z|Y_t, A=1] - f_{\mathrm{opt}}(Y_t; a_{t,t})\}^2 \Big| Y_t, A=1\Big] \\ &= \frac{\rho^{-1}a_{t,t}}{\rho^{-1}+a_{t,t}} + \left\{\frac{Y_t}{1+\rho a_{t,t}} - f_{\mathrm{opt}}(Y_t; a_{t,t})\right\}^2. \end{aligned} \tag{219}$$

Combining these results and taking the expectation over $Y_t \sim p(Y_t) = p_{\mathrm{GM}}(Y_t; a_{t,t})$ given in (203), we arrive at the MSE (209).

D. Sufficient Statistic

As a preliminary step for computing the correlation (201) for $t' \neq t$, we derive a sufficient statistic of X based on the two correlated observations $\{Y_{t'}, Y_t\}$.

Let $\boldsymbol{\Sigma}^{-1/2}$ denote a square root of $\boldsymbol{\Sigma}^{-1}$, i.e. $(\boldsymbol{\Sigma}^{-1/2})^2 = \boldsymbol{\Sigma}^{-1}$. Applying the noise whitening filter $\boldsymbol{\Sigma}^{-1/2}$ to the observation vector $(Y_{t'}, Y_t)^{\mathrm{T}}$ yields

$$\boldsymbol{\Sigma}^{-1/2}\begin{pmatrix} Y_{t'} \\ Y_t \end{pmatrix} = \boldsymbol{\Sigma}^{-1/2}\mathbf{1}_2 X + \boldsymbol{\Sigma}^{-1/2}\begin{pmatrix} W_{t'} \\ W_t \end{pmatrix}, \tag{220}$$

with $\mathbf{1}_2 = (1,1)^{\mathrm{T}}$. Note that the effective noise vector—the second term on the RHS—follows the standard Gaussian distribution. It is well-known that the MF output is a sufficient statistic of X when the effective noise vector has zero-mean i.i.d. Gaussian elements. Applying the MF $(\boldsymbol{\Sigma}^{-1/2}\mathbf{1}_2)^{\mathrm{T}}/\mathbf{1}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{1}_2$ to (220), we arrive at a sufficient statistic $Y_{t',t}$, given by

$$Y_{t',t} = \frac{\mathbf{1}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}}{\mathbf{1}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{1}_2}\begin{pmatrix} Y_{t'} \\ Y_t \end{pmatrix} = X + W_{t',t}, \tag{221}$$

with

$$W_{t',t} = \frac{\mathbf{1}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}}{\mathbf{1}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{1}_2}\begin{pmatrix} W_{t'} \\ W_t \end{pmatrix}. \tag{222}$$

It is straightforward to confirm that the sufficient statistic (221) reduces to (212). Furthermore, we find $W_{t',t} \sim \mathcal{N}(0, v_{t',t})$, with $v_{t',t} = (\mathbf{1}_2^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{1}_2)^{-1}$, which reduces to (210).

E. Correlation

To evaluate the correlation (201) for $t' \neq t$, we first derive a few quantities associated with the sufficient statistic (221).

The probability of $A=1$ given $Y_{t'}$ and Y_t is equal to that of $A=1$ given the sufficient statistic (221). Thus, repeating the derivation of $\Pr(A=1|Y_t) = \pi(Y_t; a_{t,t})$ given in (204), we have

$$\Pr(A=1|Y_{t'}, Y_t) = \pi(Y_{t',t}; v_{t',t}), \tag{223}$$

where $Y_{t',t}$ and $v_{t',t}$ are given by (212) and (210). Similarly, repeating the derivation of (214) implies that the PME $\mathbb{E}[Z|Y_{t'}, Y_t, A=1]$ reduces to

$$\mathbb{E}[Z|Y_{t'}, Y_t, A=1] = \frac{Y_{t',t}}{1+\rho v_{t',t}}. \tag{224}$$

Furthermore, the true PME $\mathbb{E}[X|Y_{t'}, Y_y]$ is given by

$$\mathbb{E}[X|Y_{t'}, Y_t] = f_{\mathrm{opt}}(Y_{t',t}; v_{t',t}). \tag{225}$$

We next evaluate the posterior covariance

$$\begin{aligned} &\mathbb{E}[\{f_{\mathrm{opt}}(Y_{t'}; a_{t',t'}) - X\}\{f_{\mathrm{opt}}(Y_t; a_{t,t}) - X\}|Y_{t'}, Y_t] \\ =&\Pr(A=0|Y_{t'}, Y_t)f_{\mathrm{opt}}(Y_{t'}; a_{t',t'})f_{\mathrm{opt}}(Y_t; a_{t,t}) \\ &+ \Pr(A=1|Y_{t'}, Y_t)\mathbb{E}[\{f_{\mathrm{opt}}(Y_{t'}; a_{t',t'}) - Z\} \\ &\cdot \{f_{\mathrm{opt}}(Y_t; a_{t,t}) - Z\}|Y_{t'}, Y_t, A=1]. \end{aligned} \tag{226}$$

Substituting (223) into (226) and using $f_{\mathrm{opt}}(Y_\tau; a_{\tau,\tau}) - Z = \{f_{\mathrm{opt}}(Y_\tau; a_{\tau,\tau}) - \mathbb{E}[Z|Y_{t'}, Y_t, A=1]\} + \{\mathbb{E}[Z|Y_{t'}, Y_t, A=1] - Z\}$ with (224) for $\tau = t', t$, we have

$$\begin{aligned} &\mathbb{E}[\{f_{\mathrm{opt}}(Y_{t'}; a_{t',t'}) - X\}\{f_{\mathrm{opt}}(Y_t; a_{t,t}) - X\}|Y_{t'}, Y_t] \\ =&\{1 - \pi(Y_{t',t}; v_{t',t})\}f_{\mathrm{opt}}(Y_{t'}; a_{t',t'})f_{\mathrm{opt}}(Y_t; a_{t,t}) \\ &+ \pi(Y_{t',t}; v_{t',t})\left\{\left[f_{\mathrm{opt}}(Y_{t'}; a_{t',t'}) - \frac{Y_{t',t}}{1+\rho v_{t',t}}\right]\right. \\ &\left.\cdot\left[f_{\mathrm{opt}}(Y_t; a_{t,t}) - \frac{Y_{t',t}}{1+\rho v_{t',t}}\right] + \frac{\rho^{-1}v_{t',t}}{\rho^{-1}+v_{t',t}}\right\}, \end{aligned} \tag{227}$$

where $Y_{t',t}$ is computed with $\{Y_{t'}, Y_t\}$, as given in (212).

Finally, we derive the correlation (201). Taking the expectation of the posterior covariance (227) over $Y_{t'}$ and Y_t, we arrive at (211).

F. Joint Pdf

To compute the expectation in (211), we need the conditional pdf $p(Y_{t'}, Y_t|A)$ in the joint pdf (207) of $\{Y_{t'}, Y_t\}$.

We first evaluate the conditional distribution of W_t given $W_{t'}$. Let

$$W_t = \alpha W_{t'} + \sqrt{\beta}\tilde{W}, \tag{228}$$

with some constants $\alpha \in \mathbb{R}$ and $\beta > 0$, where $\tilde{W}$ is a standard Gaussian random variable independent of $W_{t'}$. Computing the correlation $\mathbb{E}[W_{t'}W_t]$ and variance $\mathbb{E}[W_t^2]$, we obtain

$$\mathbb{E}[W_{t'}W_t] = \alpha\mathbb{E}[W_{t'}^2], \tag{229}$$

$$\mathbb{E}[W_t^2] = \alpha^2\mathbb{E}[W_{t'}^2] + \beta. \tag{230}$$

We use the definitions $\mathbb{E}[W_\tau^2] = a_{\tau,\tau}$ for $\tau = t', t$ and $\mathbb{E}[W_{t'}W_t] = a_{t',t}$ to have $\alpha = a_{t',t}/a_{t',t'}$ and $\beta = a_{t,t} - a_{t',t}^2/a_{t',t'}$. Thus, (228) implies

$$W_t \text{ conditioned on } W_{t'} \sim \mathcal{N}\left(\frac{a_{t',t}W_{t'}}{a_{t',t'}}, a_{t,t} - \frac{a_{t',t}^2}{a_{t',t'}}\right). \tag{231}$$

We next evaluate the conditional pdf $p(Y_{t'}, Y_t|A)$ for $A = 0$. Since $Y_\tau = W_\tau$ holds for $A = 0$, we have

$$p(Y_{t'}, Y_t|A = 0) = p(W_{t'} = Y_{t'}, W_t = Y_t)$$
$$= p_{\mathrm{G}}\left(Y_t - \frac{a_{t',t}}{a_{t',t'}}Y_{t'}; a_{t,t} - \frac{a_{t',t}^2}{a_{t',t'}}\right) p_{\mathrm{G}}(Y_{t'}; a_{t',t'}). \tag{232}$$

For $A = 1$, we use $Y_\tau = Z + W_\tau$ to find that $\{Y_{t'}, Y_t\}$ given $A = 1$ are zero-mean Gaussian random variables with covariance,

$$\mathbb{E}[Y_\tau^2|A = 1] = \rho^{-1} + a_{\tau,\tau} \quad \text{for } \tau = t', t, \tag{233}$$

$$\mathbb{E}[Y_{t'}Y_t|A = 1] = \rho^{-1} + a_{t',t}. \tag{234}$$

Repeating the derivation of (231), we obtain

$$p(Y_{t'}, Y_t|A = 1) = p_{\mathrm{G}}(Y_{t'}; \rho^{-1} + a_{t',t'})$$
$$\cdot\, p_{\mathrm{G}}\left(Y_t - \frac{\rho^{-1} + a_{t',t}}{\rho^{-1} + a_{t',t'}}Y_{t'}; \rho^{-1} + a_{t,t} - \frac{(\rho^{-1} + a_{t',t})^2}{\rho^{-1} + a_{t',t'}}\right). \tag{235}$$

Combining these results, we arrive at the conditional pdf (208).

ACKNOWLEDGMENT

The author thanks the anonymous reviewers for their suggestions that have improved the quality of the manuscript greatly.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[2] E. J. Candes, 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] Y. Wu and S. Verdu, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.
[4] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Mathematica Academiae Scientiarum Hungaricae*, vol. 10, nos. 1–2, pp. 193–215, Mar. 1959.
[5] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[6] E. J. Candes 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.
[7] 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.
[8] D. Guo and S. Verdu, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.
[9] Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*. Singapore: World Scientific, 1987.
[10] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*. New York, NY, USA: Oxford Univ. Press, 2001.
[11] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 665–669.
[12] 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.
[13] J. Barbier, M. Dia, N. Macris, and F. Krzakala, "The mutual information in random linear estimation," in *Proc. 54th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Urbana-Champaign, IL, USA, Sep. 2016, pp. 625–632.
[14] 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.
[15] Y. Wu and S. Verdu, "MMSE dimension," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4857–4879, Aug. 2011.
[16] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. Providence, RI, USA: Amer. Math. Soc., 2000.
[17] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Hanover, MA, USA: Now, 2004.
[18] 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, Dec. 2006.
[19] A. M. Tulino, G. Caire, S. Verdu, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, Jul. 2013.
[20] 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.
[21] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 3, pp. 397–399, May 1974.
[22] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1261–1266, Jul. 1994.
[23] K. Kitagawa and T. Tanaka, "Optimization of sequences in CDMA systems: A statistical–mechanics approach," *Comput. Netw.*, vol. 54, no. 6, pp. 917–924, Apr. 2010.
[24] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[25] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.
[26] 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.
[27] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A, Math. Gen.*, vol. 36, no. 43, pp. 11111–11121, Oct. 2003.
[28] D. J. Thouless, P. W. Anderson, and R. G. Palmer, "Solution of 'solvable model of a spin glass,'" *Philos. Mag.*, vol. 35, no. 3, pp. 593–601, 1977.
[29] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," *Phys. Rev. Lett.*, vol. 35, no. 26, pp. 1792–1796, Dec. 1975.
[30] 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.
[31] 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.
[32] E. Bolthausen, "An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model," *Commun. Math. Phys.*, vol. 325, no. 1, pp. 333–366, Jan. 2014.
[33] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge Univ. Press, 2008.
[34] 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.

[35] 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.

[36] 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.

[37] 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.

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

[39] 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.

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

[41] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 1588–1592.

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

[43] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 7th Conf. Uncertainty Artif. Intell.*, Seattle, WA, USA, Aug. 2001, pp. 362–369.

[44] J. Cespedes, P. M. Olmos, M. Sanchez-Fernandez, 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.

[45] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 501–505.

[46] 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.

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

[48] M. Opper and O. Winther, "Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 5, Oct. 2001, Art. no. 056131.

[49] W. Tatsuno and K. Takeuchi, "Pilot decontamination in spatially correlated massive MIMO uplink via expectation propagation," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. 104, no. 4, pp. 723–733, Apr. 2021.

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

[51] 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.

[52] Z. Fan, "Approximate message passing algorithms for rotationally invariant matrices," 2020, *arXiv:2008.11892*. [Online]. Available: http://arxiv.org/abs/2008.11892

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

[54] K. Takeuchi and C.-K. Wen, "Rigorous dynamics of expectation-propagation signal detection via the conjugate gradient method," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sapporo, Japan, Jul. 2017, pp. 88–92.

[55] K. Takeuchi, "Bayes-optimal convolutional AMP," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021. [Online]. Available: https://arxiv.org/abs/2008.11892

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

[57] 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.

[58] 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, no. 12, Dec. 2019, Art. no. 124021.

[59] S. Campese, "Fourth moment theorems for complex Gaussian approximation," 2015, *arXiv:1511.00547*. [Online]. Available: http://arxiv.org/abs/1511.00547

[60] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 399–406.

[61] M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4293–4308, Aug. 2017.

[62] L. Liu, S. Huang, and B. M. Kurkoski, "Memory approximate message passing," 2020, *arXiv:2012.10861*. [Online]. Available: http://arxiv.org/abs/2012.10861

[63] 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 with the Department of Communication Engineering and Informatics, The University of Electro-Communications, Tokyo, Japan. He is currently an Associate Professor with the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Toyohashi, Japan. He held visiting appointments at the Norwegian University of Science and Technology (NTNU), Norway, and at National Sun Yat-sen University (NSYSU), Taiwan. His research interests include the field of wireless communications, statistical signal processing, and statistical-mechanical informatics.

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