IEICE TRANS. FUNDAMENTALS, VOL.E105–A, NO.2 FEBRUARY 2022

PAPER

On the Convergence of Convolutional Approximate Message-Passing for Gaussian Signaling

Keigo TAKEUCHI[†a)], ***Member***

SUMMARY Convolutional approximate message-passing (CAMP) is an efficient algorithm to solve linear inverse problems. CAMP aims to realize advantages of both approximate message-passing (AMP) and orthogonal/vector AMP. CAMP uses the same low-complexity matched-filter as AMP. To realize the asymptotic Gaussianity of estimation errors for all right-orthogonally invariant matrices, as guaranteed in orthogonal/vector AMP, the Onsager correction in AMP is replaced with a convolution of all preceding messages. CAMP was proved to be asymptotically Bayes-optimal if a state-evolution (SE) recursion converges to a fixed-point (FP) and if the FP is unique. However, no proofs for the convergence were provided. This paper presents a theoretical analysis for the convergence of the SE recursion. Gaussian signaling is assumed to linearize the SE recursion. A condition for the convergence is derived via a necessary and sufficient condition for which the linearized SE recursion has a unique stationary solution. The SE recursion is numerically verified to converge toward the Bayes-optimal solution if and only if the condition is satisfied. CAMP is compared to conjugate gradient (CG) for Gaussian signaling in terms of the convergence properties. CAMP is inferior to CG for matrices with a large condition number while they are comparable to each other for a small condition number. These results imply that CAMP has room for improvement in terms of the convergence properties.
key words: *linear inverse problems, convolutional approximate message-passing, state evolution, Gaussian signaling, bifurcation analysis*

1. Introduction

1.1 Background

In noisy linear inverse problems, an unknown N-dimensional signal vector $\boldsymbol{x} = (x_1, \ldots, x_N)^{\mathrm{T}} \in \mathbb{R}^N$ is estimated from M-dimensional linear and noisy measurements $\boldsymbol{y} \in \mathbb{R}^M$,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}, \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_M). \tag{1}$$

In (1), $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is a known sensing matrix. The additive white Gaussian noise (AWGN) vector $\boldsymbol{w}$ has independent zero-mean Gaussian elements with variance σ^2. The 3-tuple $\{\boldsymbol{A}, \boldsymbol{x}, \boldsymbol{w}\}$ is composed of independent random variables.

For simplicity, the signal vector $\boldsymbol{x}$ is assumed to have independent and identically distributed (i.i.d.) elements with zero mean and unit variance. Such noisy linear inverse problems appear in compressed sensing [1], precoding [2] and multiuser detection [3] in wireless communications, and linear regression in machine learning [4].

Approximate message-passing (AMP) [5] is a low-complexity and powerful message-passing (MP) algorithm for noisy linear inverse problems. Bayes-optimal AMP is regarded as a large-system approximation of belief propagation [6]. Bayati and Montanari [7] used state evolution (SE) to prove that Bayes-optimal AMP can achieve the minimum mean-square error (MMSE) performance for zero-mean i.i.d. Gaussian sensing matrices in the large system limit—M and N tend to infinity while the ratio $\delta = M/N$ is kept constant. However, AMP cannot converge for non-zero mean [8] or ill-conditioned [9] sensing matrices.

Orthogonal AMP (OAMP) [10] or equivalently vector AMP (VAMP) [11] is an MP algorithm to solve the convergence issue of AMP. A prototype of OAMP/VAMP was originally proposed by Opper and Winther [12]. Bayes-optimal OAMP/VAMP is regarded as a large-system approximation [13], [14] of expectation propagation (EP) [15]. Bayes-optimal OAMP/VAMP was proved to achieve the MMSE performance in the large system limit for right-orthogonally invariant sensing matrices [11], [14]. However, Bayes-optimal OAMP/VAMP requires a high-complexity linear MMSE (LMMSE) filter in interference suppression.

Convolutional AMP (CAMP) [16], [17] aims to solve the disadvantages of AMP and OAMP/VAMP. The main feature of CAMP is Onsager correction via a convolution of messages in all preceding iterations while CAMP uses the same low-complexity matched-filter (MF) as AMP. CAMP has been proved to achieve the MMSE in the large system limit for all right-orthogonally invariant sensing matrices if an two-dimensional (2D) discrete system—called SE recursion—converges to a fixed-point (FP) [17] and if the FP is unique. The SE recursion was numerically shown to converge for sensing matrices with low-to-moderate condition numbers [17]. However, it is still open to analyze the convergence properties of the SE recursion theoretically.

1.2 Contributions

This paper presents a bifurcation analysis for investigating the convergence properties of CAMP, instead of analyzing the convergence of the SE recursion toward a certain solution. In the bifurcation analysis, we investigate conditions for which the SE recursion has multiple solutions. The occurrence of multiple solutions does not immediately indicate that the SE recursion converges to a suboptimal solution. However, it is a negative sign for the convergence of CAMP.

To derive a simple condition for the occurrence of mul-

Manuscript received May 28, 2021.
Manuscript revised July 29, 2021.
Manuscript publicized August 11, 2021.
[†]The author is with the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Toyohashi-shi, 441-8580 Japan.
a) E-mail: takeuchi@ee.tut.ac.jp
DOI: 10.1587/transfun.2021EAP1056

tiple solutions, we linearize the SE recursion and use a basic technique in linear algebra. Gaussian signaling is a key assumption to linearize the SE recursion. In this case, we know that the Bayes-optimal estimator of $\boldsymbol{x}$ in (1) is the LMMSE estimator. Thus, CAMP for Gaussian signaling may be regarded as an efficient algorithm for linear precoding [2]. See [9], [18] for existing works that assumed Gaussian signaling in the convergence analysis of MP.

The contributions of this paper are twofold: A first contribution is that a condition for the convergence is derived under Gaussian signaling via a necessary and sufficient condition for which the SE recursion has a unique stationary solution. CAMP is numerically verified to converge toward the MMSE solution when the condition is satisfied. Otherwise, CAMP cannot converge to the MMSE solution.

The other contribution is a comparison between conjugate gradient (CG) [19] and CAMP for Gaussian signaling in the noiseless case. CG is an efficient iterative algorithm to solve linear systems. Numerical results show that CAMP is inferior to CG for sensing matrices with a large condition number while they are comparable to each other for a small condition number. This implies that CAMP has room for improvement in terms of the convergence properties.

1.3 Organization & Notation

The remainder of this paper is organized as follows: After summarizing the notation in this paper, CAMP is reviewed in Sect. 2. The convergence of the SE recursion is theoretically analyzed in Sect. 3. Section 4 presents numerical results: verification of the theoretical analysis and comparison between CG and CAMP. Section 5 concludes this paper. Technical results are summarized in the appendices.

Throughout this paper, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The Euclidean norm, imaginary unit, and identity matrix of dimension M are represented as $\|\cdot\|$, i, and $\boldsymbol{I}_M$, respectively. The notations z^*, $\Re[z]$, and $\Im[z]$ mean the complex conjugate, real part, and imaginary part of a complex number $z \in \mathbb{C}$.

For a vector $\boldsymbol{v}$, the notation $[\boldsymbol{v}]_n$ denotes the nth element of $\boldsymbol{v}$. For a scalar function $f : \mathbb{R} \to \mathbb{R}$, the notation $f(\boldsymbol{v})$ is used to mean the element-wise application $[f(\boldsymbol{v})]_n = f([\boldsymbol{v}]_n)$. Furthermore, we use the notational convention $\sum_{\tau=t}^{t'} \cdots = 0$ and $\prod_{\tau=t}^{t'} \cdots = 1$ for $t > t'$.

For two sequences $\{a_t\}_{t=0}^{\infty}$ and $\{b_t\}_{t=0}^{\infty}$, the convolution operator $*$ is defined as $a_{t+i} * b_{t+j} = \sum_{\tau=0}^{t} a_{\tau+i} b_{t-\tau+j}$, in which all variables with negative indices are set to zero.

The one-dimensional (1D) and 2D Z-transforms $A(z)$ and $A(z_1, z_2)$ of $\{a_t\}_{t=0}^{\infty}$ and $\{a_{t',t}\}_{t',t=0}^{\infty}$ are defined as

$$A(z) = \sum_{t=0}^{\infty} a_t z^t, \quad A(z_1, z_2) = \sum_{t',t=0}^{\infty} a_{t',t} z_1^{t'} z_2^t, \tag{2}$$

respectively. Note that the 1D Z-transform is usually defined as $A(z^{-1})$. Nonetheless, we use the definition $A(z)$ since the 2D Z-transform $A(z_1, z_2)$ was used in the literature [20].

2. Convolutional Approximate Message-Passing

2.1 Algorithm

The goal in noisy linear inverse problems is estimation of the unknown signal vector $\boldsymbol{x}$ from the measurement vector $\boldsymbol{y}$ in (1). Bayes-optimal CAMP [17] computes an estimator $\boldsymbol{x}_{t+1} \in \mathbb{R}^N$ of $\boldsymbol{x}$ in iteration $t = 0, 1, \ldots$ as follows:

$$\boldsymbol{x}_{t+1} = f_{\mathrm{opt}}(\boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t; v_{t,t}), \tag{3}$$

with $\boldsymbol{x}_0 = \boldsymbol{0}$ and

$$\boldsymbol{z}_t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_t + \sum_{\tau=0}^{t-1} \xi_{\tau}^{(t-1)} (\theta_{t-\tau}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} - g_{t-\tau}\boldsymbol{I}_M)\boldsymbol{z}_{\tau}. \tag{4}$$

In (3), the denoiser $f_{\mathrm{opt}}(u_n; v_{t,t})$ is defined as the posterior mean estimator of x_n given a virtual AWGN measurement $u_n = x_n + w_{t,n}$ with $w_{t,n} \sim \mathcal{N}(0, v_{t,t})$, in which the variance $v_{t,t} > 0$ is determined in Sect. 2.2. For Gaussian signaling $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$, we have the LMMSE estimator

$$f_{\mathrm{opt}}(u; v_{t,t}) = \frac{u}{1 + v_{t,t}}. \tag{5}$$

The forgetting factor $\xi_{\tau}^{(t-1)} = \prod_{t'=\tau}^{t-1} \xi_{t'}$ in (4) is defined with

$$\xi_t = \frac{1}{N}\sum_{n=1}^{N} f_{\mathrm{opt}}'\left([\boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t]_n; v_{t,t}\right), \tag{6}$$

where the derivative is with respect to the first variable. For Gaussian signaling, we have the representation

$$\xi_t = \frac{1}{1 + v_{t,t}}. \tag{7}$$

Thus, the forgetting factor $\xi_{\tau}^{(t-1)}$ is geometrically small for sequences $\{v_{t',t'} > 0 : t' = \tau, \ldots, t-1\}$.

For given $\{\theta_t\}$, the tap coefficients $\{g_t\}$ in (4) are determined so as to realize the asymptotic Gaussianity of the estimation error $\boldsymbol{h}_t = (\boldsymbol{x}_t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}_t) - \boldsymbol{x}$ before denoising for all right-orthogonally invariance sensing matrices. See [17] for the details. The other tap coefficients $\{\theta_t\}$ are designed to improve the convergence properties of CAMP.

2.2 State Evolution

SE [17] tracks the asymptotic dynamics of the mean-square errors (MSEs) $N^{-1}\|\boldsymbol{h}_t\|^2$ and $N^{-1}\|\boldsymbol{q}_{t+1}\|^2$ before and after denoising, respectively, with $\boldsymbol{q}_t = \boldsymbol{x}_t - \boldsymbol{x}$. The former MSE corresponds to the variance parameter $v_{t,t}$ in CAMP while the latter MSE is associated with another variance parameter $d_{t+1,t+1}$.

The SE recursion derived in [17] is a 2D discrete system with respect to the asymptotic error covariance $v_{t',t}$ and $d_{t',t}$ in iterations t' and t, given by

$$v_{t',t} = \lim_{M=\delta N\to\infty} \frac{\boldsymbol{h}_{t'}^{\mathrm{T}}\boldsymbol{h}_t}{N}, \quad d_{t',t} = \lim_{M=\delta N\to\infty} \frac{\boldsymbol{q}_{t'}^{\mathrm{T}}\boldsymbol{q}_t}{N}. \quad (8)$$

Theorem 1 ([17]): Suppose that the signal vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ is Gaussian and that 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. Then, the covariance parameters $\{v_{t',t}\}$ and $\{d_{t',t}\}$ satisfy the SE recursion

$$\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} \xi_{t'-\tau'}^{(t'-1)}\xi_{t-\tau}^{(t-1)}\alpha_{\tau',\tau}v_{t'-\tau',t-\tau}$$
$$= \sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} \xi_{t'-\tau'}^{(t'-1)}\xi_{t-\tau}^{(t-1)}(\beta_{\tau',\tau}d_{t'-\tau',t-\tau} + \gamma_{\tau',\tau}), \quad (9)$$

with

$$d_{t'+1,t+1} = \frac{v_{t',t'}v_{t,t} + v_{t',t}}{(1+v_{t',t'})(1+v_{t,t})} \quad \text{for } t', t \geq 0, \quad (10)$$

where all variables with negative indices are set to zero, and where we impose the initial condition $d_{0,0} = 1$ and boundary conditions $d_{0,t+1} = d_{t+1,0} = -\mathbb{E}[x_n\{f_{\mathrm{opt}}(u_n; v_{t,t}) - x_n\}]$. In (9), the coefficients $\alpha_{\tau',\tau}$, $\beta_{\tau',\tau}$, and $\gamma_{\tau',\tau}$ are given by

$$\begin{aligned}\alpha_{\tau',\tau} =& g_{\tau'+\tau} - g_{\tau'+\tau+1} + (g_{\tau-1} - g_\tau) * \theta_{\tau'+\tau+1} \\ &+ g_{\tau'+\tau+1} * (\theta_\tau - \theta_{\tau-1}) + g_\tau * (\theta_{\tau'+\tau} - \delta_{\tau',0}\theta_\tau) \\ &- \theta_\tau * (g_{\tau'+\tau} - \delta_{\tau',0}g_\tau), \end{aligned} \quad (11)$$

$$\beta_{\tau',\tau} = g_\tau * \theta_{\tau'+\tau+1} - g_{\tau'+\tau+1} * \theta_\tau, \quad (12)$$

$$\gamma_{\tau',\tau} = \sigma^2(\theta_{\tau'+\tau} - \theta_{\tau'+\tau+1}), \quad (13)$$

with $g_0 = \theta_0 = 1$, where $\delta_{t',t}$ denotes the Kronecker delta.

Proof: The SE recursion (9) was proved in [17, Theorem 4]. The representation (10) for Gaussian signaling is obtained by taking the limit $\rho \to 1$ in [17, Appendix F].

To use [17, Theorem 4], however, we need the right-orthogonal invariance of the sensing matrix. It is induced from the orthogonal invariance of Gaussian signals $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$, i.e. $\boldsymbol{x} \sim \boldsymbol{\Phi}\boldsymbol{x}$ for any Haar-distributed orthogonal matrix $\boldsymbol{\Phi}$ independent of $\boldsymbol{x}$ and $\boldsymbol{A}$. Re-defining $\boldsymbol{A}\boldsymbol{\Phi}$ in $\boldsymbol{A}\boldsymbol{x} \sim \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{x}$ as the sensing matrix, we obtain the right-orthogonal invariance of the sensing matrix. □

CAMP uses the variance parameter $v_{t,t}$ in the Bayes-optimal denoiser $f_{\mathrm{opt}}(\cdot; v_{t,t})$ while $d_{t+1,t+1}$ predicts the asymptotic MSE after denoising. As proved in [17, Theorem 5], the SE recursion (9) has the uniform solution $\lim_{t',t\to\infty} v_{t',t} = v_{\mathrm{u}}$, with $v_{\mathrm{u}} > 0$ denoting the solution of the FP equation,

$$v_{\mathrm{u}} = \frac{\sigma^2}{R(-d_{\mathrm{u}}/\sigma^2)}, \quad d_{\mathrm{u}} = \frac{a_{\mathrm{u}}}{1+a_{\mathrm{u}}}, \quad (14)$$

where $R(x)$ denotes the R-transform of the asymptotic eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. This uniform solution d_{u} is unique for Gaussian signaling and corresponds to the asymptotic MMSE based on the posterior mean estimator [17]. In proving the optimality of CAMP in [17], the SE Eq. (9) was assumed to converge toward this uniform solution.

3. Bifurcation Analysis

We are interested in whether the SE recursion (9) converges to the uniform solution as both t' and t tend to infinity. To derive a bifurcation point at which the solution of the SE recursion (9) changes qualitatively, we investigate a necessary and sufficient condition for which the uniform solution is a unique stationary solution to the SE recursion (9).

To enable theoretical analysis of the SE recursion (9), we postulate the following assumption:

Assumption 1: The variance parameter $v_{t,t}$ converges toward a constant $v > 0$ as t tends to infinity.

Note that the convergence of $v_{t',t}$ is not assumed for $t' \neq t$. Assumption 1 is postulated to find soliton-like solutions shown in Figs. 1 and 4 while it excludes non-stationary solutions in which $v_{t,t}$ does not converge. It is difficult to verify Assumption 1 numerically, because of slow dynamics. Even if Assumption 1 did not hold, we could still use theoretical results in Sect. 3. When we focus on a time window $\mathcal{W}_t = \{t - W + 1, \ldots, t\}$ of size W, we can regard $v_{t',t'}$ for $t' \in \mathcal{W}_t$ as a constant approximately. Theoretical results in Sect. 3 can be applied to this finite window. Such a finite-window analysis is well-known in time series analysis [21].

Assumption 1 implies the boundedness of $v_{t',t}$ in (9). From the definition (8), we use the Cauchy-Schwarz inequality to obtain $|v_{t',t}|^2 \leq v_{t',t'}v_{t,t} < \infty$ due to the convergence $v_{t,t} \to v < \infty$. Also, we have the lower bound $v \geq v_{\mathrm{u}}$ with v_{u} denoting the solution of the FP Eq. (14) since $d_{\mathrm{u}} = v_{\mathrm{u}}/(1+v_{\mathrm{u}})$ corresponds to the asymptotic MMSE. This lower bound implies the upper bound $\xi \equiv (1+v)^{-1} \leq \xi_{\mathrm{u}}$ with $\xi_{\mathrm{u}} = (1+v_{\mathrm{u}})^{-1}$.

Assumption 2: The Z-transforms of the tap coefficients $\{g_t\}$ and $\{\theta_t\}$ exist in the region $\{z \in \mathbb{C} : |z| \leq \xi_{\mathrm{u}}\}$.

Assumption 2 implies that the 2D Z-transforms of $\alpha_{\tau',\tau}$, $\beta_{\tau',\tau}$, and $\gamma_{\tau',\tau}$ given in (11), (12), and (13) are convergent in the region $\{(z_1, z_2) \in \mathbb{C}^2 : |z_1| \leq \xi_{\mathrm{u}}, |z_2| \leq \xi_{\mathrm{u}}\}$.

Lemma 1: Postulate Assumptions 1 and 2, suppose that $\{v_{t',t}\}$ is the solution to the SE recursion (9). Then, $\{v_{t',t}\}$ satisfies the following truncated 2D discrete system in the limit $\lim_{\tau_{\mathrm{c}}\to\infty}\lim_{t',t\to\infty}$:

$$\sum_{\tau',\tau=0}^{\tau_{\mathrm{c}}} \xi^{\tau'+\tau}\alpha_{\tau',\tau}v_{t'-\tau',t-\tau} = \xi^2 v^2 B(\xi,\xi) + C(\xi,\xi)$$
$$+ \xi^2 \sum_{\tau',\tau=0}^{\tau_{\mathrm{c}}} \xi^{\tau'+\tau}\beta_{\tau',\tau}v_{t'-\tau'-1,t-\tau-1} + o(1) \quad (15)$$

for $\xi = (1+v)^{-1}$ with v in Assumption 1, in which $B(z_1, z_2)$ and $C(z_1, z_2)$ denote the 2D Z-transforms of $\{\beta_{\tau',\tau}\}$ and $\{\gamma_{\tau',\tau}\}$, respectively.

Proof: See Appendix A. □

Lemma 1 does not claim that the two recursions (9) and (15) converge toward the same solution as each other. Nonetheless, we analyze solutions of the 2D discrete system (15) instead of the original SE recursion (9). The 2D discrete system (15) is different from the original system (9) in a neighborhood of the boundaries. However, the boundary influences are forgotten due to the geometrically small factor $\xi_{t'-\tau'}^{(t-1)}\xi_{t-\tau}^{(t-1)}$ in the SE recursion (9). As a result, we expect that the two systems converge toward an identical solution.

The 2D discrete system (15) is linear when v or equivalently ξ is fixed. We focus on stationary solutions to the 2D discrete linear system (15) for fixed $\xi = \xi_{\mathrm{u}}$. The following theorem provides a necessary and sufficient condition for guaranteeing that the uniform solution is a unique stationary solution to the 2D discrete linear system (15).

Theorem 2: Define a function $Q(z_1, z_2)$ for the 2D discrete linear system (15) with $\xi = \xi_{\mathrm{u}}$ as

$$Q(z_1, z_2) = A(\xi_{\mathrm{u}}z_1, \xi_{\mathrm{u}}z_2) - \xi_{\mathrm{u}}^2 z_1 z_2 B(\xi_{\mathrm{u}}z_1, \xi_{\mathrm{u}}z_2), \quad (16)$$

where $A(z_1, z_2)$ and $B(z_1, z_2)$ denote the 2D Z-transforms of $\{\alpha_{t',t}\}$ and $\{\beta_{t',t}\}$. Assume the existence of a stationary solution $\lim_{t'\to\infty} v_{t',t'+t} = v_t$ in (15) with $\xi = \xi_{\mathrm{u}}$ for all t. Then, $v_t = v_{\mathrm{u}}$ holds for all t if and only if the following holds:

$$Q(z, z^*) \neq 0 \quad \text{in the region } \{z \in \mathbb{C} : |z| = 1\}. \quad (17)$$

Proof: See Appendix B. □

The function $Q(z_1, z_2)$ determines the stability of the 2D discrete linear system (15) with $\xi = \xi_{\mathrm{u}}$. In fact, we evaluate the 2D Z-transform of both sides in (15) with $\xi = \xi_{\mathrm{u}}$ in the limit $\lim_{\tau_{\mathrm{c}}\to\infty}\lim_{t',t\to\infty}$ to represent the 2D Z-transform $V(z_1, z_2)$ of $\{v_{t',t}\}$ as

$$V(z_1, z_2) = \frac{\xi_{\mathrm{u}}^2 v_{\mathrm{u}}^2 B(\xi_{\mathrm{u}}, \xi_{\mathrm{u}}) + C(\xi_{\mathrm{u}}, \xi_{\mathrm{u}})}{(1-z_1)(1-z_2)Q(z_1, z_2)} + o(1). \quad (18)$$

Thus, the 2D discrete linear system (15) with $\xi = \xi_{\mathrm{u}}$ is stable if and only if $Q(z_1, z_2)$ has no zeros in the region $\{(z_1, z_2) \in \mathbb{C}^2 : |z_1| \leq 1, |z_2| \leq 1\}$ [20, Theorem 5]. This implies that the 2D discrete linear system becomes unstable when the condition (17) in Theorem 2 is not satisfied.

Theorem 2 does not claim no existence of non-stationary solutions that depend on both t' and t in the limit $t', t \to \infty$ or the convergence of $v_{t',t}$ in (15) toward to the uniform solution v_{u}. It only claims that the uniform solution is a unique stationary solution to the 2D discrete linear system (15) with $\xi = \xi_{\mathrm{u}}$. As shown in Sect. 4, nonetheless, the condition (17) in Theorem 2 is useful for specifying the region in which the SE recursion (9) converges toward the uniform solution $v_{t',t} \to v_{\mathrm{u}}$ as t' and t tend to infinity.

Property 1: $\Im[Q(z, z^*)] = 0$ and $Q(z_1, z_2) = Q(z_2, z_1)$ hold.

Proof: We first prove the former property by showing that both $A(z, z^*)$ and $B(z, z^*)$ are real. We follow [17, Eq. (177)] to obtain

$$A(z_1, z_2) = \frac{(z_1 + z_2 - 1)[\Theta(z_1)G(z_2) - G(z_1)\Theta(z_2)]}{z_1 - z_2} + \frac{(1-z_2)G(z_2) - (1-z_1)G(z_1)}{z_1 - z_2}, \quad (19)$$

where $G(z)$ and $\Theta(z)$ denote the Z-transforms of $\{g_t\}$ and $\{\theta_t\}$, respectively. Since $[G(z)]^* = G(z^*)$ and $[\Theta(z)]^* = \Theta(z^*)$ hold, we find $\Im[A(z, z^*)] = 0$.

Similarly, we follow [17, Eq. (176)] to have

$$B(z_1, z_2) = \frac{\Theta(z_1)G(z_2) - G(z_1)\Theta(z_2)}{z_1 - z_2}, \quad (20)$$

which implies $\Im[B(z, z^*)] = 0$. Thus, $\Im[Q(z, z^*)] = 0$ holds.

We next prove the latter property. It is a simple exercise to confirm $A(z_1, z_2) = A(z_2, z_1)$ and $B(z_1, z_2) = B(z_2, z_1)$ from (19) and (20). Thus, $Q(z_1, z_2) = Q(z_2, z_1)$ holds. □

Property 1 implies $Q(e^{-\mathrm{i}\omega}, e^{\mathrm{i}\omega}) = Q(e^{\mathrm{i}\omega}, e^{-\mathrm{i}\omega})$ for $\omega \in \mathbb{R}$. Since $Q(e^{\mathrm{i}\omega}, e^{-\mathrm{i}\omega})$ is a real and continuous function of ω, thus, the condition (17) in Theorem 2 is satisfied if and only if the following holds:

$$\Re[Q(1,1)Q(e^{\mathrm{i}\omega}, e^{-\mathrm{i}\omega})] > 0 \quad \text{for all } \omega \in [0, \pi]. \quad (21)$$

The condition (21) can be used to determine the bifurcation points between two regions: In one region, the uniform solution $v_{t',t} = v_{\mathrm{u}}$ is a unique stationary solution to the 2D discrete linear system (15) with $\xi = \xi_{\mathrm{u}}$ as t' and t tend to infinity. In the other region, the 2D discrete linear system has multiple solutions, so that $v_{t',t}$ may converge toward a non-uniform solution. The bifurcation points are defined via the following condition

$$\min_{\omega\in[0,\pi]} \Re[Q(1,1)Q(e^{\mathrm{i}\omega}, e^{-\mathrm{i}\omega})] = 0. \quad (22)$$

4. Numerical Results

4.1 Bifurcation Analysis

To verify the correctness of the bifurcation condition (22), we consider Gaussian signaling $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ and an artificial model of the sensing matrix $\boldsymbol{A}$ [17, Corollary 3]. In the artificial model, the singular values of $\boldsymbol{A}$ are located at equal intervals. Thus, the eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is uniquely determined via the unit-first moment assumption in Theorem 1 and the condition number $\kappa = \sigma_{\max}/\sigma_{\min}$, in which $\sigma_{\max}$ and $\sigma_{\min}$ denote the maximum and minimum non-zero singular values of $\boldsymbol{A}$. See [17, Corollary 3] for how to compute the tap coefficients $\{g_t\}$ for given $\{\theta_t\}$.

We assume $\theta_0 = 1$ and $\theta_t = 0$ for all $t > 2$. When $\Theta(\xi_{\mathrm{u}}) = 1$ holds for $\xi_{\mathrm{u}} = (1 + v_{\mathrm{u}})^{-1}$, the solution v_{u} of the FP equation (14) corresponds to the MMSE performance [17, Theorem 5]. To impose the condition $\Theta(\xi_{\mathrm{u}}) = 1 + \theta_1\xi_{\mathrm{u}} + \theta_2\xi_{\mathrm{u}}^2 = 1$, we let $\theta_2 = \theta$ and $\theta_1 = -\xi_{\mathrm{u}}\theta$.

We first verify that the SE recursion (9) and 2D discrete linear system (15) converge to the same solution as each other. As shown in Fig. 1, the SE recursion (9) has a soliton-like quasi-steady solution. Since its dynamics is slow, the variance $v_{t,t}$ does not yet converge even for $t = 300$. When ξ in (15) is set appropriately, the 2D discrete linear system (15)

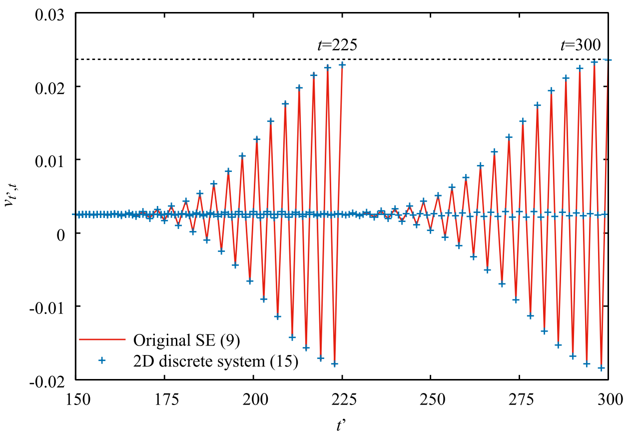

Fig. 1 Comparison between (9) and (15) for $\delta = 1$, $\kappa = 5$, $\theta = 0$, and SNR $1/\sigma^2 = 30$ dB. The parameter $v = v_{300,300}$ in (9) was used to simulate (15) with $\xi = (1+v)^{-1}$.

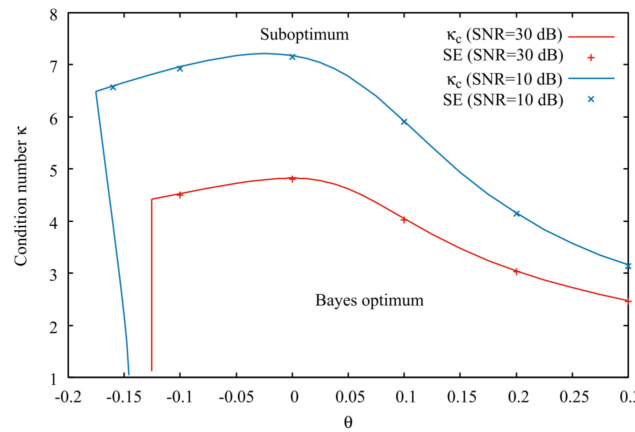

Fig. 3 Bifurcation diagram for $\delta = 1$.

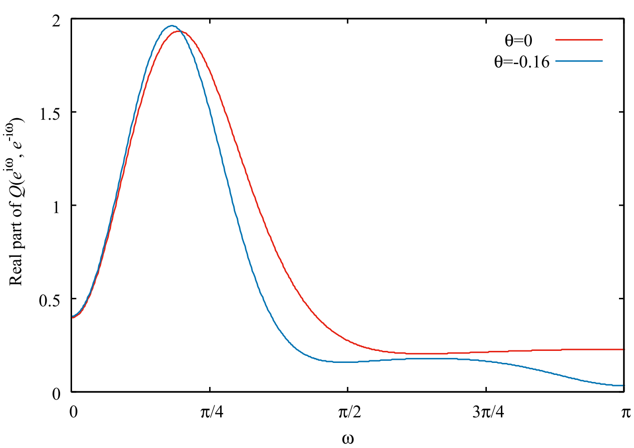

Fig. 2 $\Re[Q(e^{\mathrm{i}\omega}, e^{-\mathrm{i}\omega})]$ versus ω for $\delta = 1$, $\kappa = 5$, and SNR $1/\sigma^2 =$ 10 dB.

provides approximately the same solution as to (9).

We next plot the function $\Re[Q(e^{\mathrm{i}\omega}, e^{-\mathrm{i}\omega})]$ of $\omega \in [0, \pi]$ in Fig. 2 to evaluate the condition (22). For $\theta = 0$, the function has two local minima at $\omega = 0$ and $\omega = \omega_0$ for some $\omega_0 \in (0, \pi)$. As the condition number κ increases, $\Re[Q(e^{\mathrm{i}\omega_0}, e^{-\mathrm{i}\omega_0})]$ decreases and passes through zero while $\Re[Q(1,1)]$ remains separate from zero. As a result, the bifurcation point is given as the condition number κ such that $\Re[Q(e^{\mathrm{i}\omega_0}, e^{-\mathrm{i}\omega_0})] = 0$ holds.

For $\theta = -0.16$, the function $\Re[Q(e^{\mathrm{i}\omega}, e^{-\mathrm{i}\omega})]$ is qualitatively different from for $\theta = 0$. It has an additional local minimum at $\omega = \pi$. The local minimum $\Re[Q(e^{\mathrm{i}\omega_0}, e^{-\mathrm{i}\omega_0})]$ decreases and passes through zero as the condition number κ increases. On the other hand, $\Re[Q(e^{\mathrm{i}\pi}, e^{-\mathrm{i}\pi})]$ decreases and passes through zero as κ decreases. Thus, we have two bifurcation points defined via the condition (22). One bifurcation point is the condition number κ such that $\Re[Q(e^{\mathrm{i}\omega_0}, e^{-\mathrm{i}\omega_0})] = 0$ holds. The other bifurcation point is κ such that $\Re[Q(e^{\mathrm{i}\pi}, e^{-\mathrm{i}\pi})] = 0$ holds.

Figure 3 presents a bifurcation diagram obtained by solving the bifurcation points for different θ. The markers show numerical prediction of the true bifurcation points obtained via 10^3 iterations of the SE recursion (9). For each θ, the SE recursion (9) converges to the uniform solution $v_{t',t} = v_{\mathrm{u}}$ when the condition number κ is smaller than the corresponding marker. This uniform solution corresponds to the Bayes-optimal performance [17, Theorem 5]. When κ is larger than the marker, on the other hand, a soliton-like non-uniform solution appears, as shown in Fig. 1.

Theorem 2 claims that the uniform solution is a unique stationary solution of the 2D discrete linear system (15) when κ is between the upper bifurcation curve and the sheer bifurcation line for $\theta < 0$. The original SE recursion (9) converges to the uniform solution when κ is smaller than the upper bifurcation curve. Otherwise, it cannot converge to the uniform solution. In this sense, the upper bifurcation curve corresponds to a correct bifurcation curve.

The sheer line does not indicate any bifurcation in terms of the dynamics of the SE recursion (9). Theorem 2 claims that the 2D discrete linear system (15) has multiple solutions when κ is smaller than the sheer line. Fig. 3 implies that the initial conditions for the SE recursion (9) are included into the basin of attraction for the uniform solution when κ is smaller than the sheer line.

Finally, we show the covariance $v_{t',t}$ versus t' for condition numbers above and below the bifurcation point in Fig. 4. Since the condition number $\kappa = 4$ is below the bifurcation point, the covariance $v_{t',t}$ converges toward the uniform solution v_{u} as t' and t increase. Thus, the Bayes-optimality of CAMP is guaranteed from [17, Theorem 5]. For $\kappa = 5$, on the other hand, soliton-like solutions appear.

The occurrence of solution-like solutions is due to the instability of the 2D discrete linear system (15) and to the forgetting factor in the original system (9). As noted in Sect. 3, the 2D discrete linear system (15) is unstable when the condition number κ is above the bifurcation point. Thus, $v_{t,t}$ in (15) diverges as t increases. However, increasing $v_{t,t}$ decreases the forgetting factor $\xi_{t-\tau}^{(t-1)}$ in the original system (9) geometrically. This negative feedback produces soliton-like quasi-steady solutions.

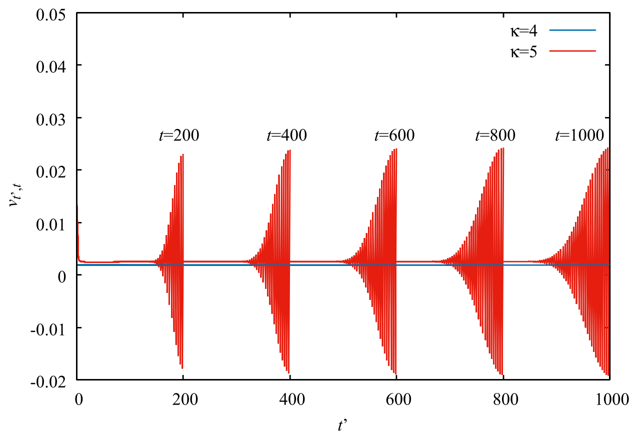

Fig. 4 Covariance $v_{t',t}$ versus $t' = 0, \ldots, t$ for several t. $\delta = 1$, $\theta = 0$, and SNR $1/\sigma^2 = 30$ dB.

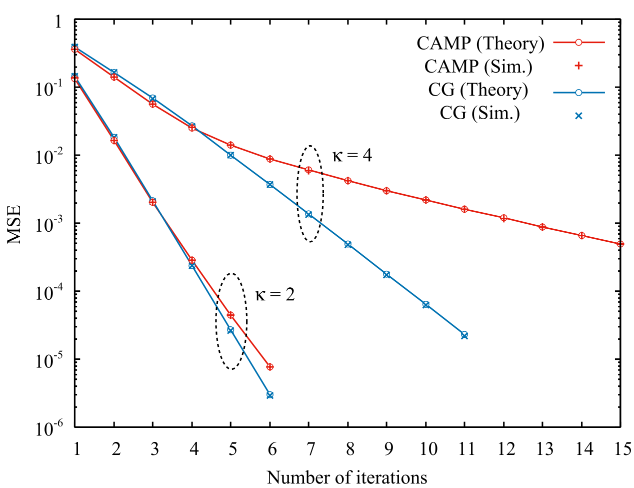

Fig. 5 MSE versus the number of iterations for $M = N = 256$, $\theta = 0$, and $\sigma^2 = 0$.

4.2 Convergence Properties

CAMP is compared to CG [19] for the noiseless case. As discussed in Appendix C, it is fair to compare the performance of the two algorithms for a fixed number of iterations. See Appendix C for the details of CG and its theoretical analysis.

In numerical simulations, the singular value decomposition (SVD) $\boldsymbol{A} = \boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ was postulated to reduce the complexity of the matrix-vector multiplication. In the SVD, $\boldsymbol{V}$ is a Hadamard matrix while $\boldsymbol{\Sigma}$ was determined according to [17, Corollary 3]. This SVD structure allows us to compute the matrix-vector multiplication in $O(N \log N)$ time.

Figure 5 shows the MSEs versus the number of iterations for CAMP and CG. The numerical simulations for $N = 256$ are in good agreement with the SE recursion (9) or a theoretical result in Appendix C, while the large system limit has been assumed in both theoretical analyses. CAMP converges more slowly than CG for the condition number $\kappa = 4$ while they are comparable to each other for $\kappa = 2$. This implies that CAMP has room for improvement in terms of the convergence property.

5. Conclusion

This paper has revealed negative aspects of CAMP theoretically: The SE recursion cannot converge to the Bayes-optimal solution for high condition numbers. CAMP converges more slowly than CG for high condition numbers.

The conclusion of this paper is that CAMP has room for improvement in terms of the convergence properties. A possible direction for improvement is a further addition of degrees of freedom for CAMP to handle high condition numbers. Long-memory damping [22] is a candidate to add degrees of freedom to CAMP. This paper has developed a methodology for linearizing the SE recursion under the assumption of Gaussian signaling. The methodology will be useful in designing CAMP with long memory damping.

Acknowledgments

K. Takeuchi was in part supported by the Grant-in-Aid for Scientific Research (B) (JSPS KAKENHI Grant Number 21H01326), Japan.

References

[1] D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol.52, no.4, pp.1289–1306, April 2006.
[2] A. Wiesel, Y.C. Eldar, and S. Shamai (Shitz), "Zero-forcing precoding and generalized inverses," IEEE Trans. Signal Process., vol.56, no.9, pp.4409–4418, Sept. 2008.
[3] S. Verdú, Multiuser Detection, Cambridge University Press, New York, USA, 1998.
[4] C.M. Bishop, Pattern Recognition and Machine Learning, 2nd ed., Springer, New York, USA, 2007.
[5] D.L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," Proc. Nat. Acad. Sci., vol.106, no.45, pp.18914–18919, Nov. 2009.
[6] 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.
[7] 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.
[8] F. Caltagirone, L. Zdeborová, and F. Krzakala, "On convergence of approximate message passing," Proc. 2014 IEEE Int. Symp. Inf. Theory, Honolulu, HI, USA, pp.1812–1816, July 2014.
[9] S. Rangan, P. Schniter, A. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," IEEE Trans. Inf. Theory, vol.65, no.9, pp.5339–5351, Sept. 2019.
[10] J. Ma and L. Ping, "Orthogonal AMP," IEEE Access, vol.5, pp.2020–2033, Jan. 2017.
[11] 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.
[12] M. Opper and O. Winther, "Expectation consistent approximate inference," J. Mach. Learn. Res., vol.6, pp.2177–2204, Dec. 2005.
[13] J. Céspedes, P.M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," IEEE Trans. Commun., vol.62, no.8, pp.2840–2849, Aug. 2014.
[14] 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.
[15] T.P. Minka, "Expectation propagation for approximate Bayesian inference," Proc. 17th Conf. Uncertainty Artif. Intell., Seattle, WA, USA, pp.362–369, Aug. 2001.
[16] K. Takeuchi, "Convolutional approximate message-passing," IEEE Signal Process. Lett., vol.27, pp.416–420, 2020.
[17] K. Takeuchi, "Bayes-optimal convolutional AMP," IEEE Trans. Inf. Theory, vol.67, no.7, pp.4405–4428, July 2021.
[18] L. Liu, C. Yuen, Y.L. Guan, Y. Li, and Y. Su, "Convergence analysis and assurance for Gaussian message passing iterative detector in massive MU-MIMO systems," IEEE Trans. Wireless Commun., vol.15, no.9, pp.6487–6501, Sept. 2016.
[19] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed., Society for Industrial and Applied Mathematics, 2003.
[20] B. O'connor and T. Huang, "Stability of general two-dimensional recursive digital filters," IEEE Trans. Acoust., Speech, Signal Process., vol.26, no.6, pp.550–560, Dec. 1978.
[21] J.D. Hamilton, Time Series Analysis, Princeton Univ. Press, Princeton, NJ, USA, 1994.
[22] L. Liu, S. Huang, and B.M. Kurkoski, "Memory approximate message passing," Proc. 2021 IEEE Int. Symp. Inf. Theory, pp.1379–1384, July 2021.
[23] I.I. Hirschman, "On a theorem of Szegö, Kac, and Baxter," J. Anal. Math., vol.14, pp.225–234, Dec. 1965.
[24] H. Widom, "Asymptotic behavior of block Toeplitz matrices and determinants. II," Adv. Math., vol.21, no.1, pp.1–29, July 1976.
[25] K. Takeuchi and C.K. Wen, "Rigorous dynamics of expectation-propagation signal detection via the conjugate gradient method," Proc. 18th IEEE Int. Workshop Sig. Process. Advances Wirel. Commun., Sapporo, Japan, pp.88–92, July 2017.

Appendix A: Proof of Lemma 1

We first prove the following lemma:

Lemma 2: Suppose that an array $\{a_{\tau',\tau}\}_{\tau',\tau=0}^{\infty}$ is summable and that $\{b_{\tau',\tau}^{(t',t)}\}_{\tau',\tau=0}^{\infty}$ is bounded. If $\lim_{t',t\to\infty} b_{\tau',\tau}^{(t',t)} = 0$ holds for all τ' and τ, we have

$$\lim_{t',t\to\infty}\left|\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} a_{\tau',\tau} b_{\tau',\tau}^{(t',t)}\right| = 0. \tag{A·1}$$

Proof: Let $\mathcal{L}_{\tau_c} = \{0,\dots,\tau_c\}\times\{0,\dots,\tau_c\}$ denote a 2D lattice for any non-negative integer τ_c. We use the triangle inequality to obtain

$$\lim_{t',t\to\infty}\left|\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} a_{\tau',\tau} b_{\tau',\tau}^{(t',t)}\right| < \lim_{t',t\to\infty}\left|\sum_{(\tau',\tau)\in\mathcal{L}_{\tau_c}} a_{\tau',\tau} b_{\tau',\tau}^{(t',t)}\right|$$
$$+ b_{\max}\lim_{t',t\to\infty}\left|\sum_{(\tau',\tau)\notin\mathcal{L}_{\tau_c}} a_{\tau',\tau}\right|, \tag{A·2}$$

with $b_{\max} = \sup|b_{\tau',\tau}^{(t',t)}|$. The assumption $\lim_{t',t\to\infty} b_{\tau',\tau}^{(t',t)} = 0$ implies that the first term on the upper bound converges to zero for any τ_c. Similarly, the boundedness of $b_{\tau',\tau}^{(t',t)}$ and the summability of $\{a_{\tau',\tau}\}$ imply that the second term tends to zero as $\tau_c \to \infty$. Thus, Lemma 2 holds. □

We first prove that $\{v_{t',t}\}$ satisfy the following 2D discrete system:

$$\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t} \xi^{\tau'+\tau}\alpha_{\tau',\tau} v_{t'-\tau',t-\tau} = \xi^2 v^2 B(\xi,\xi) + C(\xi,\xi)$$
$$+\xi^2\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\xi^{\tau'+\tau}\beta_{\tau',\tau} v_{t'-\tau'-1,t-\tau-1} + o(1). \tag{A·3}$$

We use the triangle inequality to evaluate the following difference:

$$\left|\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\beta_{\tau',\tau}\xi_{t'-\tau'}^{(t'-1)}\xi_{t-\tau}^{(t-1)} d_{t'-\tau',t-\tau} - \xi^2 v^2 B(\xi,\xi)\right.$$
$$\left.-\xi^2\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\beta_{\tau',\tau}\xi^{\tau'+\tau} v_{t'-\tau'-1,t-\tau-1}\right|$$
$$<\left|\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\beta_{\tau',\tau}(\xi_{t'-\tau'}^{(t'-1)}\xi_{t-\tau}^{(t-1)} - \xi^{\tau'+\tau}) d_{t'-\tau',t-\tau}\right|$$
$$+\left|\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\beta_{\tau',\tau}\xi^{\tau'+\tau} d_{t'-\tau',t-\tau} - \xi^2 v^2 B(\xi,\xi)\right.$$
$$\left.-\xi^2\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\beta_{\tau',\tau}\xi^{\tau'+\tau} v_{t'-\tau'-1,t-\tau-1}\right|. \tag{A·4}$$

To evaluate the first term on the upper bound, we let $a_{\tau',\tau} = \beta_{\tau',\tau}\xi_{\mathrm{u}}^{\tau'+\tau}$ and $b_{\tau',\tau}^{(t',t)} = d_{t'-\tau',t-\tau}(\xi_{t'-\tau}^{(t'-1)}\xi_{t-\tau}^{(t-1)} - \xi^{\tau'+\tau})/\xi_{\mathrm{u}}^{\tau'+\tau}$ in Lemma 2. Using Assumptions 1 and 2, we can confirm that $a_{\tau',\tau}$ and $b_{\tau',\tau}^{(t',t)}$ satisfy all assumptions in Lemma 2. Thus, the first term converges to zero as t' and t tend to infinity. Similarly, we can prove the convergence of the second term on the upper bound toward zero.

It is exercise to evaluate the differences between the other terms in the SE recursion (9) and its approximation (A·3), by repeating the same argument. Thus, we arrive at the 2D discrete system (A·3).

The truncated 2D discrete system (15) follows from the boundedness of the summations in (A·3). Since $v_{t',t}$ is bounded, we let $v_{\max} = \sup_{t',t}|v_{t',t}|$ to have

$$\lim_{t',t\to\infty}\left|\sum_{\tau'=0}^{t'}\sum_{\tau=0}^{t}\xi^{\tau'+\tau}\alpha_{\tau',\tau} v_{t'-\tau',t-\tau}\right| < v_{\max} A(\xi,\xi), \tag{A·5}$$

which is bounded from Assumption 2. Similarly, we have the boundedness of the other summation in (A·3). Thus, Lemma 1 holds.

Appendix B: Proof of Theorem 2

We investigate properties of the truncated system (15). Under the stationarity assumption $v_{t',t+t'} \to v_t$ as t' tends to infinity with t kept constant, (15) with $\xi = \xi_{\mathrm{u}}$ reduces to

$$\sum_{\tau',\tau=0}^{\tau_c}\xi_{\mathrm{u}}^{\tau'+\tau}\alpha_{\tau',\tau} v_{t-(\tau-\tau')} = \xi_{\mathrm{u}}^2 v^2 B(\xi_{\mathrm{u}},\xi_{\mathrm{u}}) + C(\xi_{\mathrm{u}},\xi_{\mathrm{u}})$$

$$+\xi_{\mathrm{u}}^2\sum_{\tau',\tau=0}^{\tau_{\mathrm{c}}}\xi_{\mathrm{u}}^{\tau'+\tau}\beta_{\tau',\tau}v_{t-(\tau-\tau')}+o(1). \tag{A· 6}$$

Applying the change of (τ',τ) into $(\tau',\tilde{\tau})$ with $\tilde{\tau}=\tau-\tau'$ in the limit $\tau_{\mathrm{c}}\to\infty$, we have

$$\sum_{\tilde{\tau}=-\infty}^{\infty}(\alpha_{\tilde{\tau}}-\xi_{\mathrm{u}}^2\beta_{\tilde{\tau}})v_{t-\tilde{\tau}}=\xi_{\mathrm{u}}^2v^2B(\xi_{\mathrm{u}},\xi_{\mathrm{u}})+C(\xi_{\mathrm{u}},\xi_{\mathrm{u}})+o(1), \tag{A· 7}$$

with

$$\alpha_\tau=\sum_{\tau'=\max\{0,-\tau\}}^{\infty}\xi_{\mathrm{u}}^{2\tau'+\tau}\alpha_{\tau',\tau'+\tau}, \tag{A· 8}$$

$$\beta_\tau=\sum_{\tau'=\max\{0,-\tau\}}^{\infty}\xi_{\mathrm{u}}^{2\tau'+\tau}\beta_{\tau',\tau'+\tau}. \tag{A· 9}$$

As proved in [17, Theorem 5], the Toeplitz system (A· 7) has the uniform solution $v_t=v_{\mathrm{u}}$ with v_{u} given in (14). To prove the uniqueness of this solution, it is necessary and sufficient to show the asymptotic non-singularity of the Toeplitz matrix $[\boldsymbol{T}]_{t',t}=\alpha_{t'-t}-\xi_{\mathrm{u}}^2\beta_{t'-t}$. For that purpose, we use the so-called Szegö theorem.

Lemma 3 ([23], [24]): Consider the $T\times T$ Toeplitz matrix $[\boldsymbol{T}]_{i,j}=c_{i-j}$ for Fourier coefficients $\{c_t\in\mathbb{C}:t\in\mathbb{Z}\}$ that satisfy

$$\sum_{t=-\infty}^{\infty}|c_t|+\sum_{t=-\infty}^{\infty}|t||c_t|^2<\infty. \tag{A· 10}$$

Define a function $c(\theta)$ on the interval $[0,1)$ as

$$c(\theta)=\sum_{t=-\infty}^{\infty}c_te^{2\pi\mathrm{i}t\theta}. \tag{A· 11}$$

Suppose that $c(\theta)$ has no zeros for $\theta\in[0,1)$ and that $\log c(\theta)$ is continuous. Then,

$$\lim_{T\to\infty}\frac{\det(\boldsymbol{T})}{e^{(T+1)d_0}}=\exp\left(\sum_{t=0}^{\infty}td_td_{-t}\right), \tag{A· 12}$$

where d_t is the Fourier coefficient of $\log c(\theta)$, given by

$$d_t=\int_0^1e^{-2\pi\mathrm{i}t\theta}\log c(\theta)d\theta. \tag{A· 13}$$

Lemma 3 provides the asymptotic determinant formula of a non-singular Toeplitz matrix $\boldsymbol{T}$. The function $c(\theta)$ corresponds to the eigenvalue spectrum of $\boldsymbol{T}$. Thus, we need to investigate zeros of $c(\theta)$ for $c_t=\alpha_t-\xi_{\mathrm{u}}^2\beta_t$ to prove the non-singularity of $[\boldsymbol{T}]_{t',t}=c_{t'-t}$.

By definition, we have

$$c(\theta)=\sum_{\tilde{\tau}=-\infty}^{\infty}(\alpha_{\tilde{\tau}}-\xi_{\mathrm{u}}^2\beta_{\tilde{\tau}})e^{2\pi\mathrm{i}\tilde{\tau}\theta}. \tag{A· 14}$$

Substituting the definitions (A· 8) and (A· 9), and changing the variables from $(\tau',\tilde{\tau})$ back to (τ',τ) with $\tau=\tau'+\tilde{\tau}$, we obtain

$$\begin{aligned}c(\theta)&=\sum_{\tau',\tau=0}^{\infty}\xi_{\mathrm{u}}^{\tau'+\tau}(\alpha_{\tau',\tau}-\xi_{\mathrm{u}}^2\beta_{\tau',\tau})e^{2\pi\mathrm{i}(\tau-\tau')\theta}\\&=A(\xi_{\mathrm{u}}e^{-2\pi\mathrm{i}\theta},\xi_{\mathrm{u}}e^{2\pi\mathrm{i}\theta})-\xi_{\mathrm{u}}^2B(\xi_{\mathrm{u}}e^{-2\pi\mathrm{i}\theta},\xi_{\mathrm{u}}e^{2\pi\mathrm{i}\theta}).\end{aligned} \tag{A· 15}$$

The condition (17) implies that $c(\theta)$ has no zeros on the interval $[0,1)$. Thus, Theorem 2 holds.

Appendix C: Conjugate Gradient

Consider the noiseless measurement (1) with $\boldsymbol{w}=0$. When $\boldsymbol{A}$ has full rank, the solution to this noiseless system is given by $\boldsymbol{x}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$ for $\delta\geq1$. CG [19] is an efficient method for computing this solution via the linear system

$$\boldsymbol{b}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{G}\boldsymbol{x}, \tag{A· 16}$$

where $\boldsymbol{G}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is symmetric and positive-definite.

Let $\boldsymbol{x}_t^{\mathrm{CG}}\in\mathbb{R}^N$, $\boldsymbol{r}_t=\boldsymbol{b}-\boldsymbol{G}\boldsymbol{x}_t^{\mathrm{CG}}$, and $\boldsymbol{p}_t\in\mathbb{R}^N$ denote a tentative solution of $\boldsymbol{x}$, the residual, and a conjugate and gradient vector in CG iteration t, respectively. The conjugate and gradient vectors $\{\boldsymbol{p}_\tau\}_{\tau=0}^{t-1}$ satisfy $\boldsymbol{p}_{t'}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{p}_t=0$ for all $t'\neq t$. Using the initial conditions $\boldsymbol{x}_0^{\mathrm{CG}}=\boldsymbol{0}$, $\boldsymbol{r}_0=\boldsymbol{b}$, and $\boldsymbol{p}_0=\boldsymbol{b}$, we compute these vectors recursively as

$$\boldsymbol{x}_{t+1}^{\mathrm{CG}}=\boldsymbol{x}_t^{\mathrm{CG}}+a_t\boldsymbol{p}_t,\quad a_t=\frac{\|\boldsymbol{r}_t\|^2}{\boldsymbol{p}_t^{\mathrm{T}}\boldsymbol{G}\boldsymbol{p}_t}, \tag{A· 17}$$

$$\boldsymbol{r}_{t+1}=\boldsymbol{r}_t-a_t\boldsymbol{G}\boldsymbol{p}_t, \tag{A· 18}$$

$$\boldsymbol{p}_{t+1}=\boldsymbol{r}_{t+1}+b_t\boldsymbol{p}_t,\quad b_t=\frac{\|\boldsymbol{r}_{t+1}\|^2}{\|\boldsymbol{r}_t\|^2}. \tag{A· 19}$$

The per-iteration complexity of CG is dominated by computation of $\boldsymbol{G}\boldsymbol{p}_t=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_t$, which requires two matrix-vector multiplications. This requirement is the same as in each iteration of CAMP. Thus, it is fair to compare the performance of CG and CAMP for a fixed number of iterations in terms of the computational complexity.

We follow [25] to evaluate the MSE $N^{-1}\|\boldsymbol{x}-\boldsymbol{x}_t^{\mathrm{CG}}\|^2$ for CG in the large system limit. Let $\mu_t>0$ denote the tth moment of the asymptotic eigenvalue distribution of $\boldsymbol{G}$, i.e.

$$\mu_t=\lim_{M=\delta N\to\infty}\frac{1}{N}\mathrm{Tr}(\boldsymbol{G}^t). \tag{A· 20}$$

Theorem 3: Suppose that $N^{-1}\|\boldsymbol{x}\|^2$ converges almost surely to 1 as $N\to\infty$, that $\boldsymbol{A}$ is right-orthogonally invariant, that the empirical eigenvalue distribution of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ converges almost surely to a compactly-supported deterministic distribution in the large system limit, and that $\{r_{\tau,t}\}$ and $\{p_{\tau,t}\}$ satisfy

$$r_{\tau,t+1}=r_{\tau,t}-\bar{a}_tp_{\tau-1,t},\quad r_{0,0}=1, \tag{A· 21}$$

$$p_{\tau,t+1}=r_{\tau,t+1}+\bar{b}_tp_{\tau,t},\quad p_{0,0}=1, \tag{A· 22}$$

where $r_{\tau,t}=p_{\tau,t}=0$ holds for $\tau\notin\{0,\dots,t\}$, with $\bar{a}_t=$

$\bar{r}_t/\bar{p}_t$, $\bar{b}_t = \bar{r}_{t+1}/\bar{r}_t$, and

$$\bar{r}_t = \sum_{\tau',\tau=0}^{t} r_{\tau',t} r_{\tau,t} \mu_{\tau'+\tau+2}, \tag{A· 23}$$

$$\bar{p}_t = \sum_{\tau',\tau=0}^{t} p_{\tau',t} p_{\tau,t} \mu_{\tau'+\tau+3}. \tag{A· 24}$$

Then, the MSE $N^{-1}\|\boldsymbol{x} - \boldsymbol{x}_t^{\mathrm{CG}}\|^2$ converges almost surely to

$$\frac{1}{N}\|\boldsymbol{x} - \boldsymbol{x}_t^{\mathrm{CG}}\|^2 \to \sum_{\tau,\tau'=0}^{t} r_{\tau',t} r_{\tau,t} \mu_{\tau+\tau'} \tag{A· 25}$$

in the large system limit.

Proof: For notational simplicity, we omit the superscript CG in $\boldsymbol{x}_t^{\mathrm{CG}}$. The asymptotic MSE (A· 25) is obtained via the following asymptotic representation:

$$\boldsymbol{x}_t = \boldsymbol{x} - \sum_{\tau=0}^{t} r_{\tau,t} \boldsymbol{G}^{\tau}\boldsymbol{x} + o(1), \tag{A· 26}$$

$$\boldsymbol{r}_t = \sum_{\tau=0}^{t} r_{\tau,t} \boldsymbol{G}^{\tau+1}\boldsymbol{x} + o(1), \tag{A· 27}$$

$$\boldsymbol{p}_t = \sum_{\tau=0}^{t} p_{\tau,t} \boldsymbol{G}^{\tau+1}\boldsymbol{x} + o(1). \tag{A· 28}$$

Since $\boldsymbol{G} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is orthogonally invariant and symmetric, we have the eigen-decomposition $\boldsymbol{G} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}$, in which $\boldsymbol{V}$ is a Haar-distributed orthogonal matrix and independent of $\boldsymbol{\Lambda}$. Using (A· 26) and [14, Corollary 1], we obtain

$$\frac{1}{N}\|\boldsymbol{x} - \boldsymbol{x}_t\|^2 = \frac{1}{N}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{V}\left(\sum_{\tau=0}^{t} r_{\tau,t}\boldsymbol{\Lambda}^{\tau}\right)^2 \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x} + o(1)$$
$$\to \lim_{M=\delta N\to\infty} \frac{1}{N}\mathrm{Tr}\left\{\left(\sum_{\tau=0}^{t} r_{\tau,t}\boldsymbol{\Lambda}^{\tau}\right)^2\right\} \tag{A· 29}$$

almost surely in the large system limit, which reduces to (A· 25).

We prove the representation (A· 26)–(A· 28) by induction. For $t = 0$, we use $r_{0,0} = p_{0,0} = 1$ to have $\boldsymbol{x}_0 = \boldsymbol{x} - r_{0,0}\boldsymbol{x} = \boldsymbol{0}$, $\boldsymbol{r}_0 = r_{0,0}\boldsymbol{G}\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{p}_0 = p_{0,0}\boldsymbol{G}\boldsymbol{x} = \boldsymbol{b}$. Thus, the representation (A· 26)–(A· 28) is correct for $t = 0$.

Suppose that (A· 26)–(A· 28) are correct for some non-negative integer t. We shall prove that (A· 26)–(A· 28) provide the correct representation for $\boldsymbol{x}_{t+1}$, $\boldsymbol{r}_{t+1}$ and $\boldsymbol{p}_{t+1}$.

We first confirm the almost sure convergence $a_t \to \bar{a}_t$ in the large system limit. Using the induction hypotheses (A· 27) and (A· 28), and repeating the derivation for the asymptotic MSE (A· 25), we find the almost sure convergence $N^{-1}\|\boldsymbol{r}_t\|^2 \to \bar{r}_t$ and $N^{-1}\boldsymbol{p}_t^{\mathrm{T}}\boldsymbol{G}\boldsymbol{p}_t \to \bar{p}_t$. These results imply the almost sure convergence $a_t \to \bar{a}_t$.

$$\boldsymbol{x}_{t+1} = \boldsymbol{x} - \sum_{\tau=0}^{t} r_{\tau,t}\boldsymbol{G}^{\tau}\boldsymbol{x} + \bar{a}_t \sum_{\tau=0}^{t} p_{\tau,t}\boldsymbol{G}^{\tau+1}\boldsymbol{x} + o(1)$$
$$= \boldsymbol{x} - \sum_{\tau=0}^{t+1} \left(r_{\tau,t} - \bar{a}_t p_{\tau-1,t}\right)\boldsymbol{G}^{\tau}\boldsymbol{x} + o(1)$$
$$= \boldsymbol{x} - \sum_{\tau=0}^{t+1} r_{\tau,t+1}\boldsymbol{G}^{\tau}\boldsymbol{x} + o(1), \tag{A· 30}$$

where the last two equalities follows from $r_{t+1,t} = p_{-1,t} = 0$ and (A· 21), respectively.

Similarly, from (A· 18) we obtain

$$\boldsymbol{r}_{t+1} = \sum_{\tau=0}^{t} r_{\tau,t}\boldsymbol{G}^{\tau+1}\boldsymbol{x} - \bar{a}_t \sum_{\tau=0}^{t} p_{\tau,t}\boldsymbol{G}^{\tau+2}\boldsymbol{x} + o(1)$$
$$= \sum_{\tau=0}^{t+1} \left(r_{\tau,t} - \bar{a}_t p_{\tau-1,t}\right)\boldsymbol{G}^{\tau+1}\boldsymbol{x} + o(1)$$
$$= \sum_{\tau=0}^{t+1} r_{\tau,t+1}\boldsymbol{G}^{\tau+1}\boldsymbol{x} + o(1). \tag{A· 31}$$

This representation implies $N^{-1}\|\boldsymbol{r}_{t+1}\|^2 \to \bar{r}_{t+1}$, so that we have the almost sure convergence $b_t \to \bar{b}_t$. Thus, from (A· 19) we obtain

$$\boldsymbol{p}_{t+1} = \sum_{\tau=0}^{t+1} r_{\tau,t+1}\boldsymbol{G}^{\tau+1}\boldsymbol{x} + \bar{b}_t \sum_{\tau=0}^{t} p_{\tau,t}\boldsymbol{G}^{\tau+1}\boldsymbol{x} + o(1)$$
$$= \sum_{\tau=0}^{t+1} \left(r_{\tau,t+1} + \bar{b}_t p_{\tau,t}\right)\boldsymbol{G}^{\tau+1}\boldsymbol{x} + o(1)$$
$$= \sum_{\tau=0}^{t+1} p_{\tau,t+1}\boldsymbol{G}^{\tau+1}\boldsymbol{x} + o(1). \tag{A· 32}$$

These results imply that the representation (A· 26)–(A· 28) is correct for all t. □

Keigo Takeuchi received the B. Eng., M. Inf., and Ph.D. degrees in informatics from Kyoto University, Kyoto, Japan, in 2004, 2006, and 2009, respectively. From 2009 to 2016, he was an Assistant Professor at Department of Communication Engineering and Informatics, University of Electro-Communications, Japan. He is currently an Associate Professor at Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Japan. He served as an Associate Editor of IEICE Trans. Fundamentals. from 2013 to 2017. He held visiting appointments at Norwegian University of Science and Technology (NTNU), Norway, and at National Sun Yat-sen University (NSYSU), Taiwan. His research interests are in 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.