

LETTER

Finite-Size Correction of Expectation-Propagation DetectionYuki OBA[†], *Nonmember* and Keigo TAKEUCHI^{††a)}, *Member*

SUMMARY Expectation propagation (EP) is a powerful algorithm for signal recovery in compressed sensing. This letter proposes correction of a variance message before denoising to improve the performance of EP in the high signal-to-noise ratio (SNR) regime for finite-sized systems. The variance message is replaced by an observation-dependent consistent estimator of the mean-square error in estimation before denoising. Massive multiple-input multiple-output (MIMO) is considered to verify the effectiveness of the proposed correction. Numerical simulations show that the proposed variance correction improves the high SNR performance of EP for massive MIMO with a few hundred transmit and receive antennas.

key words: *compressed sensing, massive multiple-input multiple-output (MIMO), expectation propagation, state evolution, variance correction*

1. Introduction

Orthogonal approximate message-passing (OAMP) [1] or vector approximate message-passing (VAMP) [2] is a powerful iterative algorithm for signal recovery in compressed sensing. A prototype of OAMP/VAMP was originally proposed by Opper and Winther [3]. Bayes-optimal OAMP/VAMP can be regarded as a large-system approximation [4], [5] of expectation propagation (EP) [6]. In this letter, Bayes-optimal OAMP/VAMP is referred to as EP.

The performance of EP degrades for finite-sized systems in general. This is because the large system limit—both input and output dimensions tend to infinity with their ratio fixed—is assumed in the derivation of EP [5]. This letter improves the performance of EP for finite-sized systems via correction of an update rule in EP.

The main idea is to replace a conventional estimator of the mean-square error (MSE) in estimation before denoising with an observation-dependent consistent estimator. The conventional estimator is equivalent to state evolution (SE) [2], [5] that describes the rigorous dynamics of the MSE in the large system limit. Thus, it is essentially deterministic while a naive approximation of SE is used in practical implementations. To improve the estimation of the MSE for finite-sized systems, we propose a consistent estimator of the MSE that depends on observations explicitly.

As a related work, Donoho et al. [7, Eq. (39)] proposed an observation-dependent consistent estimator of the MSE in approximate message-passing (AMP) [8], which is an iterative algorithm applicable only to a smaller class of systems than EP. The consistent estimator improves the performance of AMP for finite-sized systems. Motivated by [7], we extend the consistent estimator in AMP to the EP case.

The other related methods were proposed in [9], [10]. The initialization of the postulated prior variance was numerically optimized in [9]. Kamilov et al. [10] corrected the conventional estimator of the MSE before denoising in the middle of denoising. This *post*-correction is based on the expectation-maximization (EM) algorithm and applicable to EP. However, the EM iteration requires additional complexity especially for complicated denoisers.

The contributions of this letter are twofold. One is to propose an observation-dependent consistent estimator of the MSE in EP estimation before denoising. We prove the consistency in the large system limit via rigorous SE [5].

The other contribution is numerical simulations for massive multiple-input multiple-output (MIMO) with a few hundred transmit and receive antennas. EP with the proposed correction is numerically shown to outperform conventional EP for high signal-to-noise ratios (SNRs).

Throughout this letter, $CN(\mathbf{0}, \Sigma)$ denotes the circularly symmetric complex Gaussian distribution with covariance Σ . The conjugate transpose and trace operators for a matrix are represented as \cdot^H and $\text{Tr}(\cdot)$, respectively. The notations $\|\cdot\|$, $\stackrel{\text{a.s.}}{=}$, and $\xrightarrow{\text{a.s.}}$ denote the Euclidean norm, almost sure equality, and almost sure convergence.

2. Mathematical Model

The goal of this letter is to estimate an N -dimensional signal vector $\mathbf{x} \in \mathbb{C}^N$ from M -dimensional noisy observations $\mathbf{y} \in \mathbb{C}^M$, given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \quad \mathbf{w} \sim CN(\mathbf{0}, \sigma^2 \mathbf{I}_M). \quad (1)$$

In (1), \mathbf{w} is independent of $\{\mathbf{A}, \mathbf{x}\}$ and the additive white Gaussian noise (AWGN) vector with covariance $\sigma^2 \mathbf{I}_M$. The sensing matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ is assumed known and independent of $\{\mathbf{x}, \mathbf{w}\}$.

Throughout this letter, we postulate the following assumptions:

Assumption 1: \mathbf{x} has independent and identically distributed (i.i.d.) elements with zero mean and unit variance.

Manuscript received June 18, 2021.

Manuscript publicized July 19, 2021.

[†]The author was with the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Aichi, 441-8580 Japan.

^{††}The author is with the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Aichi, 441-8580 Japan.

a) E-mail: takeuchi@ee.tut.ac.jp

DOI: 10.1587/transfun.2021EAL2048

Assumption 2: The sensing matrix \mathbf{A} is assumed right-unitarily invariant [5], i.e. $\mathbf{A}\Psi \sim \mathbf{A}$ for any unitary matrix Ψ independent of \mathbf{A} . Furthermore, the empirical eigenvalue distribution of $\mathbf{A}^H\mathbf{A}$ is assumed to converge almost surely toward a compactly supported deterministic distribution with unit first moment in the large system limit— M and N tend to infinity while the ratio $\delta = M/N$ is kept constant.

Under Assumptions 1 and 2, EP was proved to be Bayes-optimal in the large system limit [2], [5] if the ratio δ is larger than a certain value—called belief-propagation threshold. For example, Assumption 2 holds when \mathbf{A} has zero-mean i.i.d. elements with variance $1/M$. This assumption is called i.i.d. Rayleigh fading in wireless communications.

3. Expectation Propagation

3.1 Conventional EP

EP [5] computes an estimator $\mathbf{x}_{B,t+1} \in \mathbb{C}^N$ of the signal vector \mathbf{x} in iteration $t = 0, 1, \dots$ via message-passing between two modules, called modules A and B. Module A receives an extrinsic estimator $\mathbf{x}_{B \rightarrow A,t} \in \mathbb{C}^N$ of \mathbf{x} and an estimator $\bar{v}_{B \rightarrow A,t} > 0$ of the MSE $N^{-1}\mathbb{E}[\|\mathbf{x} - \mathbf{x}_{B \rightarrow A,t}\|^2]$ from module B. Define a coefficient $\bar{\gamma}_t$ as

$$\bar{\gamma}_t^{-1} = \lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \text{Tr} \left(\Xi_t^{-1} \mathbf{A} \mathbf{A}^H \right), \quad (2)$$

with

$$\Xi_t = \sigma^2 \mathbf{I}_M + \bar{v}_{B \rightarrow A,t} \mathbf{A} \mathbf{A}^H. \quad (3)$$

The messages $\mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t) \in \mathbb{C}^N$ and $\bar{v}_{A \rightarrow B,t} > 0$ sent to module B are computed as follows:

$$\mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t) = \mathbf{x}_{B \rightarrow A,t} + \bar{\gamma}_t \mathbf{A}^H \Xi_t^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}_{B \rightarrow A,t}), \quad (4)$$

$$\bar{v}_{A \rightarrow B,t} = \bar{\gamma}_t - \bar{v}_{B \rightarrow A,t}, \quad (5)$$

with $\mathbf{x}_{B \rightarrow A,0} = \mathbf{0}$ and $\bar{v}_{B \rightarrow A,0} = 1$.

We have set the postulated prior variance $\bar{v}_{B \rightarrow A,0}$ to the true prior variance $N^{-1}\mathbb{E}[\|\mathbf{x}\|^2]$. While this option is intuitively reasonable, optimization of $\bar{v}_{B \rightarrow A,0}$ can improve the performance of EP for finite-sized systems [9].

Module B postulates the virtual AWGN observation model of \mathbf{x} ,

$$\mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t) = \mathbf{x} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{CN}(\mathbf{0}, \bar{v}_{A \rightarrow B,t} \mathbf{I}_N), \quad (6)$$

where \mathbf{w}_t is independent of \mathbf{x} . This postulation implies that module B regards $\bar{v}_{A \rightarrow B,t}$ as an estimator of the MSE $N^{-1}\mathbb{E}[\|\mathbf{x} - \mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t)\|^2]$ in module A.

Module B first computes the posterior mean $\mathbf{x}_{B,t+1} \in \mathbb{C}^N$ of \mathbf{x} given $\mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t)$, $\bar{v}_{A \rightarrow B,t}$, and the corresponding MSE $\bar{v}_{B,t+1}$ as

$$\mathbf{x}_{B,t+1} = \mathbb{E}[\mathbf{x} | \mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t), \bar{v}_{A \rightarrow B,t}], \quad (7)$$

$$\bar{v}_{B,t+1} = \frac{1}{N} \mathbb{E} \left[\|\mathbf{x} - \mathbf{x}_{B,t+1}\|^2 | \bar{v}_{A \rightarrow B,t} \right]. \quad (8)$$

The posterior mean $\mathbf{x}_{B,t+1}$ is used as an estimator of \mathbf{x} in

iteration $t + 1$. To refine the estimator, module B feeds the messages $\mathbf{x}_{B \rightarrow A,t+1}$ and $\bar{v}_{B \rightarrow A,t+1}$ back to module A,

$$\mathbf{x}_{B \rightarrow A,t+1} = \bar{v}_{B \rightarrow A,t+1} \left(\frac{\mathbf{x}_{B,t+1}}{\bar{v}_{B,t+1}} - \frac{\mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t)}{\bar{v}_{A \rightarrow B,t}} \right), \quad (9)$$

$$\frac{1}{\bar{v}_{B \rightarrow A,t+1}} = \frac{1}{\bar{v}_{B,t+1}} - \frac{1}{\bar{v}_{A \rightarrow B,t}}. \quad (10)$$

The variance parameters $\bar{v}_{A \rightarrow B,t}$ and $\bar{v}_{B \rightarrow A,t}$ in EP can be pre-computed via the recursion (2), (3), (5), (8), and (10) with the initial value $\bar{v}_{B \rightarrow A,0} = 1$ —called SE recursion—when $\bar{\gamma}_t$ is represented in closed form. SE analysis [2], [5] proved that the squared errors $N^{-1}\|\mathbf{x} - \mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t)\|^2$, $N^{-1}\|\mathbf{x} - \mathbf{x}_{B,t+1}\|^2$, and $N^{-1}\|\mathbf{x} - \mathbf{x}_{B \rightarrow A,t+1}\|^2$ converge almost surely toward $\bar{v}_{A \rightarrow B,t}$, $\bar{v}_{B,t+1}$, and $\bar{v}_{B \rightarrow A,t+1}$ in the large system limit, respectively. Thus, the AWGN postulation (6) is correct in the large system limit.

In practical implementations, as a naive estimator of $\bar{\gamma}_t$ we use $\gamma_t > 0$ obtained by removing the large system limit in (2),

$$\gamma_t^{-1} = \frac{1}{N} \text{Tr} \left(\Xi_t^{-1} \mathbf{A} \mathbf{A}^H \right). \quad (11)$$

Furthermore, the MSE $\bar{v}_{B,t+1}$ is replaced with the posterior variance $v_{B,t+1}$, given by

$$v_{B,t+1} = \frac{1}{N} \mathbb{E} \left[\|\mathbf{x} - \mathbf{x}_{B,t+1}\|^2 | \mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t), \bar{v}_{A \rightarrow B,t} \right]. \quad (12)$$

EP requires computation of the high-complexity matrix inversion Ξ_t^{-1} . A reduction in the complexity is possible via the singular-value decomposition (SVD) of \mathbf{A} [2] or an approximate implementation based on the conjugate gradient method [11], [12]. Thus, the complexity issue in EP is outside the scope of this letter.

Céspedes et al. [4] proposed EP with vector variance parameters $\mathbf{v}_{A \rightarrow B,t} \in \mathbb{R}^N$ and $\mathbf{v}_{B \rightarrow A,t} \in \mathbb{R}^N$, instead of the scalar variance parameters $\bar{v}_{A \rightarrow B,t}$ and $\bar{v}_{B \rightarrow A,t}$ [5]. The vector variance parameters can improve the performance of EP in small systems. However, the performance gap between the vector and scalar variance parameters shrinks as the system size increases. Furthermore, the low-complexity implementations in [2], [11], [12] are not applicable to EP with the vector variance parameters. Since we are interested in moderately large systems, we focus on EP with the scalar variance parameters in this letter.

3.2 Variance Correction

The three variance parameters $\bar{v}_{A \rightarrow B,t}$, $\bar{v}_{B,t+1}$, and $\bar{v}_{B \rightarrow A,t+1}$ are not accurate for finite-sized systems while they are consistent with the corresponding MSEs in the large system limit. In particular, error propagation occurs for finite-sized systems when $\bar{v}_{A \rightarrow B,t}$ is significantly smaller than the true instantaneous MSE $N^{-1}\|\mathbf{x} - \mathbf{x}_{A \rightarrow B,t}(\bar{\gamma}_t)\|^2$. This is because smaller $\bar{v}_{A \rightarrow B,t}$ lets the posterior mean $\mathbf{x}_{B,t+1}$ be harder decision.

To resolve this finite-size issue, we propose a \mathbf{y} -dependent consistent estimator $\hat{v}_{A \rightarrow B, t}$ of $N^{-1} \|\mathbf{x} - \mathbf{x}_{A \rightarrow B, t}(\bar{\gamma}_t)\|^2$,

$$\hat{v}_{A \rightarrow B, t} = \frac{1}{M} \|\mathbf{A} \mathbf{x}_{A \rightarrow B, t}(\gamma) - \mathbf{y}\|^2, \quad (13)$$

with (4), where $\gamma \geq 0$ is a design parameter. The parameter γ was set straightforwardly to zero in AMP [7, Eq. (39)], i.e. the message itself fed back from the denoiser. In EP, however, the estimator $\hat{v}_{A \rightarrow B, t}$ is not consistent when γ is set to zero or $\bar{\gamma}_t$. In the proposed EP, the conventional update rule (5) is replaced by $\bar{v}_{A \rightarrow B, t} = \hat{v}_{A \rightarrow B, t}$ with an appropriate parameter γ determined in the following theorem.

Theorem 1: Suppose that γ satisfies the quadratic equation

$$\frac{\gamma^2}{\delta \bar{v}_{B \rightarrow A, t}} \left(1 - \frac{\sigma^2}{\bar{\gamma}_t}\right) - \frac{2\gamma}{\delta} + \sigma^2 + \frac{\bar{v}_{B \rightarrow A, t}}{\delta} = \bar{\gamma}_t - \bar{v}_{B \rightarrow A, t}. \quad (14)$$

Then, $\hat{v}_{A \rightarrow B, t}$ is a consistent estimator of $N^{-1} \|\mathbf{x} - \mathbf{x}_{A \rightarrow B, t}(\bar{\gamma}_t)\|^2$ in the large system limit.

Proof: We use [5, Theorem 4] to prove the consistency of $\hat{v}_{A \rightarrow B, t}$ in the large system limit. However, use of [5, Theorem 4] requires the consistency of $\hat{v}_{A \rightarrow B, \tau}$ for all $\tau < t$. To circumvent this dilemma, we prove Theorem 1 by induction.

We omit the proof of Theorem 1 for $t = 0$ since the proof is the same as for the general case. Suppose that $\hat{v}_{A \rightarrow B, \tau}$ is consistent for all $\tau < t$. We prove the consistency of $\hat{v}_{A \rightarrow B, t}$.

Consider the SVD $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$ and let $\mathbf{z}_t = \mathbf{A} \mathbf{x}_{A \rightarrow B, t}(\gamma) - \mathbf{y}$, $\mathbf{b}_t = \mathbf{V}^H \mathbf{q}_t$, and $\mathbf{q}_t = \mathbf{x}_{B \rightarrow A, t} - \mathbf{x}$. We use (1), (3), and (4) to represent \mathbf{z}_t as

$$\mathbf{z}_t = \mathbf{U}(\mathbf{G}_t \mathbf{b}_t + \mathbf{H}_t \mathbf{U}^H \mathbf{w}), \quad (15)$$

with

$$\mathbf{G}_t = (\mathbf{I}_M - \gamma \mathbf{\Sigma} \mathbf{\Sigma}^H \tilde{\mathbf{\Xi}}_t^{-1}) \mathbf{\Sigma}, \quad (16)$$

$$\mathbf{H}_t = \gamma \mathbf{\Sigma} \mathbf{\Sigma}^H \tilde{\mathbf{\Xi}}_t^{-1} - \mathbf{I}_M, \quad (17)$$

$$\tilde{\mathbf{\Xi}}_t = \sigma^2 \mathbf{I}_M + \bar{v}_{B \rightarrow A, t} \mathbf{\Sigma} \mathbf{\Sigma}^H. \quad (18)$$

The induction hypothesis allows us to use [5, Theorem 4 (A2)] to find that (13) reduces to

$$\begin{aligned} \hat{v}_{A \rightarrow B, t} &\stackrel{\text{a.s.}}{=} \frac{\|\mathbf{q}_t\|^2 \text{Tr}(\mathbf{G}_t \mathbf{G}_t^H)}{MN} + \frac{\|\mathbf{w}\|^2 \text{Tr}(\mathbf{H}_t \mathbf{H}_t^H)}{M^2} + o(1) \\ &\stackrel{\text{a.s.}}{=} \frac{\bar{v}_{B \rightarrow A, t} \text{Tr}(\mathbf{G}_t \mathbf{G}_t^H)}{M} + \frac{\sigma^2 \text{Tr}(\mathbf{H}_t \mathbf{H}_t^H)}{M} + o(1). \end{aligned} \quad (19)$$

In the derivation of the last equality, we have used the almost sure convergence $N^{-1} \|\mathbf{q}_t\|^2 \stackrel{\text{a.s.}}{\rightarrow} \bar{v}_{B \rightarrow A, t}$ [5, Theorem 4], obtained from the induction hypothesis, and the strong law of large numbers $M^{-1} \|\mathbf{w}\|^2 \stackrel{\text{a.s.}}{\rightarrow} \sigma^2$.

We next evaluate the traces. Substituting (16) and (17), and using (18), we have

$$\hat{v}_{A \rightarrow B, t} \stackrel{\text{a.s.}}{=} \sigma^2 + \frac{\bar{v}_{B \rightarrow A, t} - 2\gamma}{\delta}$$

$$+ \frac{\gamma^2}{M} \text{Tr}(\mathbf{\Sigma} \mathbf{\Sigma}^H \tilde{\mathbf{\Xi}}_t^{-1} \mathbf{\Sigma} \mathbf{\Sigma}^H) + o(1), \quad (20)$$

where we have used the unit-first-moment assumption $N^{-1} \text{Tr}(\mathbf{\Sigma} \mathbf{\Sigma}^H) \stackrel{\text{a.s.}}{=} 1 + o(1)$ in Assumption 2. Applying the following identity:

$$\frac{\bar{v}_{B \rightarrow A, t}}{N} \text{Tr}(\mathbf{\Sigma} \mathbf{\Sigma}^H \tilde{\mathbf{\Xi}}_t^{-1} \mathbf{\Sigma} \mathbf{\Sigma}^H) \stackrel{\text{a.s.}}{=} 1 - \frac{\sigma^2}{\bar{\gamma}_t} + o(1) \quad (21)$$

obtained from (2) and (18), we arrive at

$$\begin{aligned} \hat{v}_{A \rightarrow B, t} &\stackrel{\text{a.s.}}{\rightarrow} \sigma^2 + \frac{\bar{v}_{B \rightarrow A, t} - 2\gamma}{\delta} + \frac{\gamma^2}{\delta \bar{v}_{B \rightarrow A, t}} \left(1 - \frac{\sigma^2}{\bar{\gamma}_t}\right) \\ &= \bar{\gamma}_t - \bar{v}_{B \rightarrow A, t} = \bar{v}_{A \rightarrow B, t}, \end{aligned} \quad (22)$$

where the last two equalities follow from (14) and (5), respectively. Since $N^{-1} \|\mathbf{x} - \mathbf{x}_{A \rightarrow B, t}\|^2$ was proved to converge almost surely toward $\bar{v}_{A \rightarrow B, t}$ in the large system limit [5], $\hat{v}_{A \rightarrow B, t}$ is a consistent estimator of $N^{-1} \|\mathbf{x} - \mathbf{x}_{A \rightarrow B, t}\|^2$ in the large system limit. \square

Theorem 1 allows us to determine the parameter $\gamma > 0$ as the unique positive solution to the quadratic equation (14) since $\bar{\gamma}_t$ given in (2) is bounded from below by

$$\bar{\gamma}_t \geq \left(\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \text{Tr}(\sigma^{-2} \mathbf{A} \mathbf{A}^H) \right)^{-1} \stackrel{\text{a.s.}}{=} \sigma^2 + o(1). \quad (23)$$

In practical implementations, however, we replace $\bar{\gamma}_t$ with γ_t given in (11), which might be smaller than σ^2 because of the fluctuation of the trace $N^{-1} \text{Tr}(\mathbf{A} \mathbf{A}^H)$. Thus, we select a solution to the quadratic equation (14) such that γ becomes positive and bounded as a function of γ_t .

If γ_t given in (11) is larger than or equal to $\sigma^2 + (1 + \delta^{-1}) \bar{v}_{B \rightarrow A, t}$, we select

$$\gamma = \frac{\bar{v}_{B \rightarrow A, t} + \sqrt{D_t}}{1 - \sigma^2/\gamma_t}, \quad (24)$$

with

$$\begin{aligned} D_t &= \bar{v}_{B \rightarrow A, t}^2 + \left(1 - \frac{\sigma^2}{\gamma_t}\right) \bar{v}_{B \rightarrow A, t} \\ &\quad \cdot \{\delta \gamma_t - \delta \sigma^2 - (1 + \delta) \bar{v}_{B \rightarrow A, t}\}. \end{aligned} \quad (25)$$

If γ_t is smaller than $\sigma^2 + (1 + \delta^{-1}) \bar{v}_{B \rightarrow A, t}$, we select

$$\gamma = \frac{\bar{v}_{B \rightarrow A, t} - \sqrt{D_t}}{1 - \sigma^2/\gamma_t}, \quad (26)$$

while we use the following at $\gamma_t = \sigma^2$:

$$\gamma = \frac{1 + \delta}{2} \bar{v}_{B \rightarrow A, t}. \quad (27)$$

It is straightforward to confirm the positivity and boundedness of γ when γ is given in (24), (26), or (27).

Remark 1: When γ_t is in the interval $(\sigma^2, \sigma^2 + (1 + \delta^{-1}) \bar{v}_{B \rightarrow A, t})$, both (24) and (26) are positive. In this case, we have selected (26), which is continuous at $\gamma_t = \sigma^2$

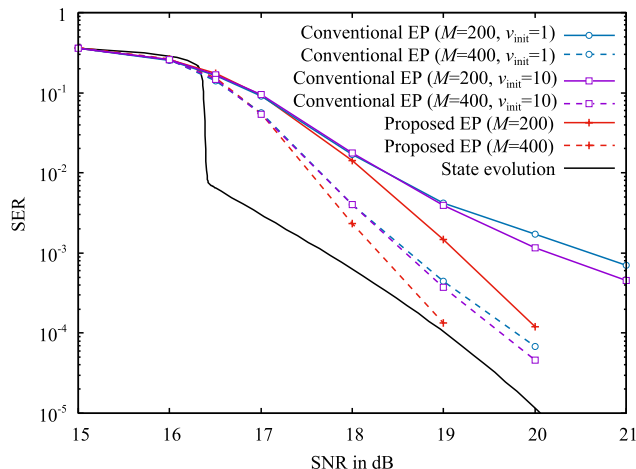


Fig. 1 SER versus SNR $1/\sigma^2$. $\delta = 10/11$ and 40 iterations.

and discontinuous at $\gamma_t = \sigma^2 + (1 + \delta^{-1})\bar{v}_{B \rightarrow A,t}$, while (24) diverges in the limit $\gamma_t \rightarrow \sigma^2$ and is continuous at $\gamma_t = \sigma^2 + (1 + \delta^{-1})\bar{v}_{B \rightarrow A,t}$. Note that the interval is narrow since the event $\gamma_t < \sigma^2 + (1 + \delta^{-1})\bar{v}_{B \rightarrow A,t}$ occurs typically for $\bar{v}_{B \rightarrow A,t} \ll 1$.

4. Numerical Simulations

As a numerical example, we consider massive MIMO with 16 quadrature amplitude modulation (QAM) over i.i.d. Rayleigh fading. In massive MIMO, the signal dimension N corresponds to the number of transmit antennas while the observation dimension M is the number of receive antennas. Note that the variance correction proposed in Sect. 3.2 is applicable to the other problems in signal processing under mild conditions in Assumptions 1 and 2.

The variance parameter $\bar{v}_{A \rightarrow B,t}$ in conventional EP [5] is computed with (5). To make a fair comparison between the improve EP [9] and proposed EP, we simulated the conventional EP with numerical optimization of the initial variance parameter $\bar{v}_{B \rightarrow A,0} = v_{\text{init}}$.

The variance parameter $\bar{v}_{A \rightarrow B,t}$ in the proposed EP is computed with the initial value $\bar{v}_{B \rightarrow A,0} = 1$ and the consistent estimator $\bar{v}_{A \rightarrow B,t} = \hat{v}_{A \rightarrow B,t}$ given in (13). See Sect. 3.2 for how to compute the parameter γ in (13). In these EP algorithms, we replace the parameters $\bar{\gamma}_t$ and $\bar{v}_{B,t+1}$ with the practical alternatives γ_t and $v_{B,t+1}$ given in (11) and (12), respectively.

Figure 1 presents comparisons between the conventional and proposed EP in terms of symbol error rate (SER) for the overloaded case $\delta = 10/11$. The initial value $\bar{v}_{B \rightarrow A,0} = v_{\text{init}} = 10$ in the conventional EP was optimized via numerical simulations at SNR $1/\sigma^2 = 19$ dB. The proposed EP outperforms the two conventional EP algorithms in the high SNR regime. The gap in the SERs between the conventional and proposed EP algorithms is larger for $M = 200$ than for $M = 400$ while their SERs improve toward the rigorous SE prediction [2], [5] in the large system limit as the system size increases. We conclude that the proposed EP

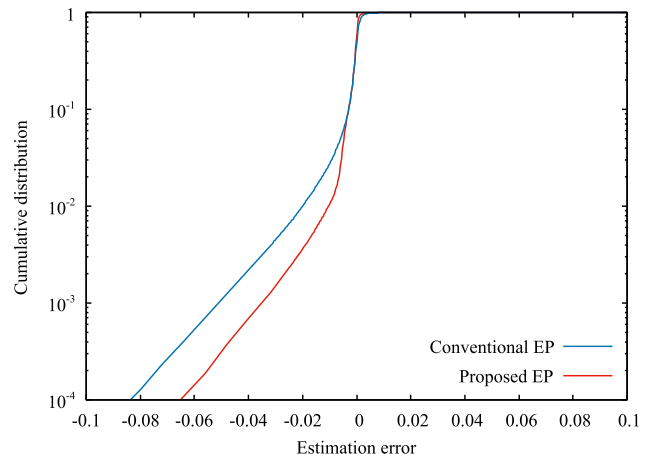


Fig. 2 Cumulative distribution of the estimation error $\bar{v}_{A \rightarrow B,t} - N^{-1} \|\mathbf{x} - \mathbf{x}_{A \rightarrow B,t}\|^2$ for iteration $t = 6$. $M = 200$, $N = 220$, and SNR $1/\sigma^2 = 20$ dB.

improves the high SNR performance of the conventional EP algorithms for finite-sized systems.

To investigate why the proposed EP can improve the high SNR performance, we estimated the cumulative distribution of the estimation error $\bar{v}_{A \rightarrow B,t} - N^{-1} \|\mathbf{x} - \mathbf{x}_{A \rightarrow B,t}\|^2$ from 10^6 independent trials, shown in Fig. 2. The conventional EP with $\bar{v}_{B \rightarrow A,0} = 1$ under-estimates the instantaneous MSE $N^{-1} \|\mathbf{x} - \mathbf{x}_{A \rightarrow B,t}\|^2$ more likely than the proposed EP. When the variance parameter $\bar{v}_{A \rightarrow B,t}$ is under-estimated, the denoiser (7)—called soft decision in communications—is likely to make a harder decision, which implies the occurrence of error propagation. Since error propagation provides more significant impacts in the high SNR regime than in the low SNR regime, the proposed EP can improve the high SNR performance of the conventional EP.

Acknowledgments

K. Takeuchi was in part supported by the Grant-in-Aid for Scientific Research (B) (JSPS KAKENHI Grant Numbers 18H01441 and 21H01326), Japan.

References

- [1] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol.5, pp.2020–2033, Jan. 2017.
- [2] 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.
- [3] M. Opper and O. Winther, "Expectation consistent approximate inference," *J. Mach. Learn. Res.*, vol.6, pp.2177–2204, Dec. 2005.
- [4] 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.
- [5] 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.
- [6] T.P. Minka, "Expectation propagation for approximate Bayesian inference," *Proc. 17th Conf. Uncertainty Artif. Intell.*, Seattle, WA, USA, pp.362–369, Aug. 2001.

- [7] 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.
 - [8] 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.
 - [9] G. Yao, H. Chen, and J. Hu, "An improved expectation propagation based detection scheme for MIMO systems," *IEEE Trans. Commun.*, vol.69, no.4, pp.2163–2175, April 2021.
 - [10] U.S. Kamilov, S. Rangan, A.K. Fletcher, and M. Unser, "Approximate message passing with consistent parameter estimation and applications to sparse learning," *IEEE Trans. Inf. Theory*, vol.60, no.5, pp.2969–2985, May 2014.
 - [11] 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.
 - [12] N. Skuratovs and M. Davies, "Upscaling vector approximate message passing," *Proc. 2020 IEEE Int. Conf. Acoust. Speech Signal Process., virtual*, pp.4757–4761, May 2020.
-