Decentralized Expectation Consistent Signal Recovery for Phase Retrieval

Chang-Jen Wang , Chao-Kai Wen , *Member, IEEE*, Shang-Ho Tsai , *Senior Member, IEEE*,
and Shi Jin , *Senior Member, IEEE*

Abstract—**In this study, we present a phase retrieval solution that aims to recover signals from noisy phaseless measurements. A recently proposed scheme known as generalized expectation consistent signal recovery (GEC-SR), has shown better accuracy, speed, and robustness than many existing methods. However, sensing high-resolution images with large transform matrices presents a computational burden for GEC-SR, thereby limiting its applications to areas, such as real-time implementation. Moreover, GEC-SR does not support distributed computing, which is an important requirement to modern computing. To address these issues, we propose a novel decentralized algorithm called "deGEC-SR" by leveraging the core framework of GEC-SR. deGEC-SR exhibits excellent performance similar to GEC-SR but runs tens to hundreds of times faster than GEC-SR. We derive the theoretical state evolution for deGEC-SR and demonstrate its accuracy using numerical results. Analysis allows quick generation of performance predictions and enriches our understanding on the proposed algorithm.**

Index Terms—**Phase retrieval, Bayes-optimal inference, expectation consistent, decentralized algorithm, distributed processing.**

I. Introduction

Background—Mathematically, phase retrieval (PR) refers to the estimation of an unknown vector $\mathbf{x} \in \mathbb{C}^N$ from a nonlinear function of (noisy) linear measurements in the following form:

$$\mathbf{y} = \mathbb{Q}(\mathbf{z}) = \mathbb{Q}(\mathbf{Ax} + \mathbf{w}), \tag{1}$$

where $\mathbb{Q}(\cdot) = |\cdot|$ or $|\cdot|^2$ takes element-wise, $\mathbf{A} \in \mathbb{C}^{M \times N}$ is known as the transform matrix, and $\mathbf{w}$ represents the noise. Such a problem is of interest in various domains, such as wireless communications [1]–[3] and physical sciences [4]–[8]. The domain of wireless communications is related to PR in terms of a certain equipment of wireless systems, especially those that address economic demands. Such equipment is likely to use passive envelop detectors [1] or have limited accuracy in terms of in-phase and quadrature components [2], [3]. The phase information of such equipment is completely lost or experiences serious degradation. In physical sciences, such as optical imaging [4], [5], electron microscopy [6], astronomical imaging [7], and X-ray diffractive imaging [8], receiving devices usually measure only the amplitude information (brightness or intensity) of electromagnetic waves without measuring their phase information. These applications require the reconstruction of complete transmission signals (including phase information) on the basis of the observed amplitude information.

Despite the ubiquity of applications, the PR problem is non-convex and has been determined to be an NP-hard problem [8]. Several algorithms have been developed to derive solutions in the past decade. PR algorithms were first developed in the 70 s by the optics community, where Fienup [9] and Gerchberg-Saxton [10] proposed a simple interactive projection method to reconstruct received signals with phase information. PR has attracted considerably attention in the optimization community. Most well-established methods, such as PhaseLift [11] and stochastic Wirtinger Flow (WF) [12], are based on semidefinite relaxation (SDR). Such convex optimization methods operate by lifting the original N-dimensional vector space to a high dimensional matrix space, although it results in computational complexity. Other convex relaxation formulations without lifting (i.e., to avoid increasing the dimension), such as PhaseMax [13], PhaseLamp [14], and re-weighted WF [15], have been proposed. The nonlifting-based PR algorithms rely on the so-called spectral initializers [16] that provide accurate initialization vectors. Among the initialization methods, LSPE [17] shows the best performance.

In the past few years, numerous algorithms based on the Bayesian framework have been proposed to solve the inverse problem to a generalized linear model (GLM) in a uniform framework with the emergence of compressive sensing. Approximate message passing (AMP) [18] and its generalization referred to as GAMP [19] are representative of this class, and can be applied to estimate $\mathbf{x}$ from (1) with various nonlinear

functions and statistical priors. AMP has been continuously studied with several updates to enhance the estimation robustness to noise and broaden the class of transform matrices. A Bayesian GAMP algorithm for PR was studied in [20], [21] and rigorously analyzed in [22]. Unfortunately, these AMP-based algorithms are fragile when improper linear transform matrices **A** are used. The performance of AMP-based algorithms decreases when transform matrices are outside the special case of independent and identically distributed (i.i.d.) sub-Gaussian and zero-mean elements. Thus, several new approaches have been proposed to address this problem. Under the partial discrete Fourier transform (DFT) matrix, Ma *et al.* proposed a signal recovery (SR) method from linear observations called Turbo-SR [23]. In [24], Turbo-SR was expanded to nonlinear observations, resulting in the generalized Turbo-SR. Robust alternatives to AMP for right-rotationally invariant matrices, **A**, were independently proposed by two groups of authors [25], [26] and were labeled as OAMP and VAMP, respectively. The ideas of OAMP/VAMP have been extended to many applications, such as unified Bayesian inference [27], matrix completion [28], and bilinear recovery [29]. VAMP was derived from the expectation consistent (EC) approximate inference framework of [30] and applied to the special case of linear transforms. VAMP was then extended to a GLM in [31], [32]. The generalized EC-SR (GEC-SR) algorithm [32] outperforms the initial GEC algorithm [31] because of effective updating. The initial GEC algorithm was applied to solve the PR problem called prVAMP in [33], [34] and exhibited superior accuracy, speed, and robustness to many competing methods.[1] In this context, the GEC-SR for PR is expected to be the best by far.

Despite its remarkable performance, GEC-SR (or the GEC algorithm) requires matrix inversion per-iteration. The singular value decomposition (SVD) form of GEC-SR can reduce its per-iteration cost by pre-computing the SVD of transform matrix **A** [31]. However, the calculation of SVD for high dimensions can be a bottleneck for many applications. For example, consider sensing a $N = 1024^2$ sized image with $M = 4N$ measurements. The calculation of the SVD of such a large transform matrix is usually limited when a personal computer is used. Another limitation of GEC-SR is that measurements of large optical imaging are frequently acquired from distributed sensors [7]. Distributed algorithms for GEC-SR are required to broaden its application. We make three technical contributions to this work utilizing the excellent performance but limited scalability of GEC-SR.

Contributions—First, we propose a novel decentralized algorithm called deGEC-SR to address the scalability issues by leveraging the core framework of GEC-SR [32]. In particular, we partition measurements into several independent clusters with small pieces, perform local inference using the original GEC-SR for each cluster separately, and combine the results to obtain an approximation to the global posterior. Thus, the computational burden in each local computing processor is lower than the original computational burden in centralized processing. The decentralized architecture allows distributed processing. The reduction in the complexity requirement and distributed computing leads to a significant increase in the execution speed of the algorithm. deGEC-SR not only provides the same excellent performance as GEC-SR and reasonable execution time but also presents good robustness to transform matrices.

Second, we derive the theoretical state evolutions (SEs) of GEC-SR and deGEC-SR for PR. We demonstrate their accuracy using the numerical results. Although the decentralization approach benefits the processing speed of large images, deGEC-SR can result in performance degeneration. The theoretical SE allows us to quickly obtain the performance predictions for deGEC-SR with different clusters. Accordingly, we can determine the trade-off between performance and complexity.

Third, we analyze the response of each module of GEC-SR to plot the behavior of the iteration trajectory of GEC-SR in a 2D chart. This analysis provides insight into the performance of GEC-SR and deGEC-SR and enriches our understanding of several relevant questions as follows: 1) Why is the initialization of PR more important than other nonlinear problems (e.g., quantization)? 2) How does decentralization degenerate performance compared with the centralized CGE-SR? Which part of the algorithm considerably degenerates performance? 3) The performance degeneration caused by decentralization differs for different sparsity signals. When does the decentralization result in minor degeneration of performance?

Related work—A decentralization method was proposed by [36] to alleviate the scalability issue for developing DWF. DWF is based on the general perturbed proximal primal-dual and Lagrangian methods. Despite its advantage, DWF inherits the same disadvantages as WF. In particular, the convergence of DWF is excessively slow, and its performance is unacceptable when the transform matrix does not follow a Gaussian distribution. An alternative means to alleviate the scaling issue is to impose some extra constraints on the transform matrix. Research [37] proposed to place the transform matrix in a generalized block diagonal form. However, the entries of the transform matrix might not allow to be designed in several applications.

In the context of algorithm development, belief propagation (BP) [38] is commonly used to approximate the Bayes inference to ensure tractable computation. The Bayesian approach combined with a BP technique leads to the so-called AMP algorithm [18] (or its generalization called GAMP [19]). The first AMP-based algorithm designed for PR is prGAMP [20]. Another approximation is based on a variational Bayesian, which leads to prVBEM [39]. Compared with GEC-SR for PR, prGAMP and prVBEM do not compute the matrix inverse. Thus, the computational complexity of the latter can be immensely reduced. Unfortunately, these algorithms require the transform matrix to be i.i.d. Gaussian (or i.i.d. sub-Gaussian) and encounters convergence problems [33]–[35] for generic transform matrices.

---

[1] Sixteen PR algorithms are compared in [34] under various transform matrices and signal to noise ratios. Among these algorithms, prVAMP and prSAMP [35] exhibit the best performance over various scenarios. However, prSAMP is proposed by making the parallel updates with AMP sequential, indicating that it is extremely slow. prVAMP achieves the same accuracy as prSAMP but runs considerably faster.

Other relevant work are conducted by [40], [41]. In [40], a streaming algorithm based on AMP was proposed to deal with the case where only a small fraction of measurements arrive at a time. Work [41] applied the AMP framework in multi-processor computational systems by processing a subset of measurements at each processor node. Applying the proposed deGEC-SR to the above applications is a promising direction.

Notations—For any matrix $\mathbf{A}$, $\mathbf{A}^H$ is the conjugate transpose of $\mathbf{A}$, and $tr(\mathbf{A})$ denotes the traces of $\mathbf{A}$. The matrix vector product follows the standard definition in linear algebra. For convenience, we also define vector-vector multiplication and division as their component-wise vector multiplication and division, respectively. That is, for any vectors $\mathbf{a} = [a_i]$ and $\mathbf{b} = [b_i]$, we define $\mathbf{ab} = [a_i b_i]$ and $\mathbf{a}/\mathbf{b} = [a_i/b_i]$. In addition, $\mathrm{diag}(\mathbf{Q})$ returns a *vector* containing the diagonal elements of $\mathbf{Q}$, and $\mathrm{Diag}(\mathbf{v})$ returns a diagonal *matrix* containing $\mathbf{v}$ itself. $\mathcal{N}_{\mathbb{C}}(x; \mu, v)$ represents a complex Gaussian distribution of dummy variable x with mean μ and variance v. When a random variable x follows a probability distribution function $f(x)$, we write $x \sim f(x)$.

II. BACKGROUND ON GEC-SR

In this section, we provide an overview of GEC-SR. As GEC-SR is close to Bayes estimation, we first introduce Bayes estimators in the next subsection to aid the succeeding discussion.

A. Preliminary on Bayes Estimation

We first explain the Bayes estimator for (1) under its simple scalar version, which is given by

$$y = \mathtt{Q}(z + w). \tag{2}$$

In the simple model, we assume an identity transformation, that is, $z = x$. In addition, we let z be a random variable with distribution $\mathsf{P}(z)$ to facilitate the interpretation. We define the posterior distribution as

$$\mathsf{P}(z|y) = \frac{f(y|z)\mathsf{P}(z)}{\mathcal{Z}(y)}, \tag{3}$$

where $f(y|z)$ is a likelihood function, and $\mathcal{Z}(y) = \int f(y|z)\mathsf{P}(z)\,\mathrm{d}z$ is the normalization. Then, the Bayes estimate of z is simply the mean of the posterior distribution

$$\hat{\mu}_z = \int z\,\mathsf{P}(z|y)\,\mathrm{d}z. \tag{4}$$

Similarly, the posterior variance of z is given by

$$\hat{v}_z = \int |z|^2\,\mathsf{P}(z|y)\,\mathrm{d}z - |\hat{\mu}_z|^2, \tag{5}$$

which is also the corresponding mean squared error (MSE) of the posterior mean. If $f(y|z)$ is the true conditional distribution of measurement y under (2), then the posterior mean is known as the minimum mean squared error (MMSE) estimator.

When $\mathtt{Q}(\cdot) = |\cdot|$ and w is the circularly symmetric complex Gaussian with mean 0 and variance σ_w^2, we obtain [20, (8)]

$$\mathsf{P}_{\mathrm{out}}(y|z) = \frac{2y}{\sigma_w^2} e^{-\frac{y^2 + |z|^2}{\sigma_w^2}} I_0\left(\frac{2y|z|}{\sigma_w^2}\right), \tag{6}$$

where $I_0(\cdot)$ is the 0th-order modified Bessel function of the first kind. Table I shows the posterior mean and variance when $f(y|z) := \mathsf{P}_{\mathrm{out}}(y|z)$ given above and $\mathsf{P}(z) := \mathcal{N}_{\mathbb{C}}(z; \mu_z, v_z)$ are applied to (3). Additional details are found in [20, Appendix A]. The above model (2) assumes that noise w is added prior to function $\mathtt{Q}$. The post-intensity noise model, i.e., $y = \mathtt{Q}(z) + w$, can also be considered. In this case, the corresponding posterior mean and variance can be found in [20, (48) & (50)]. The relevant information is listed in Table I for convenient.

B. GEC-SR

Now, we return to the inference problem under (vector) PR (1). In this case, the conditional distribution of the measurements is given by

$$\mathsf{P}_{\mathrm{out}}(\mathbf{y}|\mathbf{z}) = \prod_{m=1}^{M} \mathsf{P}_{\mathrm{out}}(y_m|z_m). \tag{7}$$

where $\mathsf{P}_{\mathrm{out}}(y_m|z_m)$ is written as (6). We assume that the components of $\mathbf{x}$ are separable, i.e., $\mathsf{P}(\mathbf{x}) = \prod_{n=1}^{N} \mathsf{P}(x_n)$. The components of $\mathbf{x}$ are chosen to be i.i.d. and modeled by the Gaussian-Bernoulli distribution

$$\mathsf{P}(x) = (1 - \rho)\delta(x) + \rho\mathcal{N}_{\mathbb{C}}(x; 0, \rho^{-1}). \tag{8}$$

Our goal is to estimate the random vector $\mathbf{x}$ from the observed measurements $\mathbf{y}$. The variance of x_n is denoted by P_x, which is 1 if $\mathsf{P}(x)$ is given by (8). In addition, we define

$$P_z = P_x \cdot \mathrm{tr}(\mathbf{A}^H \mathbf{A})/N, \tag{9}$$

which is interpreted as the average power of z_m.

Similar to (3), the posterior probability distribution can be computed as

$$\mathsf{P}(\mathbf{x}|\mathbf{y}) = \frac{\mathsf{P}(\mathbf{x}) \int d\mathbf{z}\,\mathsf{P}_{\mathrm{out}}(\mathbf{y}|\mathbf{z})\delta(\mathbf{z} - \mathbf{A}\mathbf{x})}{\mathcal{Z}(\mathbf{y})}, \tag{10}$$

where $\mathcal{Z}(\mathbf{y})$ is the normalization. Given the posterior probability, an estimator for x_n can be obtained by the posterior mean

$$\hat{x}_n = \int x_n \mathsf{P}(x_n|\mathbf{y})\,dx_n, \tag{11}$$

where $\mathsf{P}(x_n|\mathbf{y}) = \int_{\mathbf{x}\backslash x_n} d\mathbf{x}\mathsf{P}(\mathbf{x}|\mathbf{y})$ denotes the marginal posterior probability of x_n. The notation $\int_{\mathbf{x}\backslash x_n} d\mathbf{x}$ denotes the integration over all the variables in $\mathbf{x}$, except for x_n, i.e., $\int_{\mathbf{x}\backslash x_n} d\mathbf{x} = \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N}$. The posterior mean estimator (11) minimizes the (Bayesian) MSE defined as

$$\mathrm{mse}(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} \mathsf{E}\{|\hat{x}_n - x_n|^2\}, \tag{12}$$

where the expectation operator is w.r.t. $\mathsf{P}(\mathbf{x}, \mathbf{y})$. We refer to (11) as the Bayes-optimal estimator. If $\mathbf{A}$ is an identity matrix, i.e.,

TABLE I
THE POSTERIOR MEAN AND VARIANCE

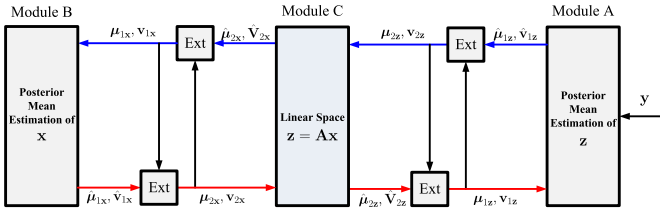| $f(\cdot)$	Posterior Mean	Posterior Variance
Phase Retrieval [20]	$\hat{\mu}_z = \frac{yv_z}{v_z+\sigma_w^2}\frac{\mu_z}{\|\mu_z\|}\frac{I_1\left(\frac{2y\|\mu_z\|}{v_z+\sigma_w^2}\right)}{I_0\left(\frac{2y\|\mu_z\|}{v_z+\sigma_w^2}\right)} + \frac{\mu_z\sigma_w^2}{v_z+\sigma_w^2}$	$\hat{v}_z = \frac{v_z^2 y^2}{(v_z+\sigma_w^2)^2} + \frac{v_z\sigma_w^2}{v_z+\sigma_w^2} - \frac{v_z^2 y^2}{(v_z+\sigma_w^2)^2}\left(\frac{I_1\left(\frac{2y\|\mu_z\|}{v_z+\sigma_w^2}\right)}{I_0\left(\frac{2y\|\mu_z\|}{v_z+\sigma_w^2}\right)}\right)^2$
Phase Retrieval [20] (Post-intensity Noise)	$\hat{\mu}_z = \frac{2e^{j\theta_z}}{\mathcal{Z}_y v_z}\int_0^\infty r^2 e^{-\frac{r^2+\|\mu_z\|^2}{v_z}}\mathcal{N}_{\mathbb{C}}(y;r,\sigma_w^2)I_1\left(\frac{2r\|\mu_z\|^2}{v_z}\right)dr,$ $\mathcal{Z}_y = \frac{2}{v_z}\int_0^\infty r e^{-\frac{r^2+\|\mu_z\|^2}{v_z}}\mathcal{N}_{\mathbb{C}}(y;r,\sigma_w^2)I_0\left(\frac{2r\|\mu_z\|^2}{v_z}\right)dr$	$\hat{v}_z = \frac{2}{\mathcal{Z}_y v_z}\int_0^\infty r^3 e^{-\frac{r^2+\|\mu_z\|^2}{v_z}}\mathcal{N}_{\mathbb{C}}(y;r,\sigma_w^2)I_0\left(\frac{2r\|\mu_z\|^2}{v_z}\right)dr$ $-\|\hat{\mu}_z\|^2$
Gaussian-Bernoulli [42]	$\hat{\mu}_x = \mathcal{Z}_x^{-1}\left(\frac{\mu_x\rho^{-1}}{v_x+\rho^{-1}}\right),$ $\mathcal{Z}_x^{-1} = \frac{\rho\mathcal{N}_{\mathbb{C}}(0;\mu_x,v_x+\rho^{-1})}{(1-\rho)\mathcal{N}_{\mathbb{C}}(0;\mu_x,v_x)+\rho\mathcal{N}_{\mathbb{C}}(0;\mu_x,v_x+\rho^{-1})}$	$\hat{v}_x = \mathcal{Z}_x^{-1}\left(\left\|\frac{\mu_x\rho^{-1}}{v_x+\rho^{-1}}\right\|^2 + \frac{v_x\rho^{-1}}{v_x+\rho^{-1}}\right) - \mathcal{Z}_x^{-2}\left\|\frac{\mu_x\rho^{-1}}{v_x+\rho^{-1}}\right\|^2$

Fig. 1. Block diagram of GEC-SR algorithm.

$\mathbf{A} = \mathbf{I}$, then $\mathbf{z} = \mathbf{x}$ and the inference problem can be reduced to be parallel scalars of the Bayes estimator, as introduced in Section II.A. However, for the general case, the direct computation of (11) is intractable because of the high-dimensional integrals in the marginal posteriors $\mathsf{P}(x_n|\mathbf{y})$.

In our recent study [32], the innovative GEC-SR algorithm was proposed as an iterative method to recover signal $\mathbf{x}$ from nonlinear measurements $\mathbf{y}$. Although GEC-SR is derived based on the sophisticated expectation consistent approximation, its working procedure is straightforward. That is, if $\mathbf{z}$ can be estimated through the *de-nonlinear* process from nonlinear measurements $\mathbf{y}$, then the following problem is to recover $\mathbf{x}$ from the noisy linear transformation of $\mathbf{z} = \mathbf{A}\mathbf{x}$.

The block diagram of GEC-SR is illustrated in Fig. 1. GEC-SR consists of three modules, namely, Modules A, B, and C. Module A computes the posterior mean and variance of $\mathbf{z}$; the process is interpreted as a *de-nonlinear*. Module B computes the posterior mean and variance of $\mathbf{x}$; this process is interpreted as a *de-noising*. Module C provides the posterior means and variances of $\mathbf{x}$ and $\mathbf{z}$, while constraining the estimation problem into the linear space $\mathbf{z} = \mathbf{A}\mathbf{x}$. Except for Module C, which is involved with $\mathbf{A}$, Modules A and B can be performed as parallel scalars of the Bayes estimator. Each module inputs prior mean and variance and calculates the posterior mean and variance. These procedures follow a circular manner, that is,

$$A \underbrace{\rightarrow C \rightarrow}_{\text{forward}} B \underbrace{\rightarrow C \rightarrow}_{\text{backward}} A \underbrace{\rightarrow C \rightarrow}_{\text{forward}} B \rightarrow \cdots \qquad (13)$$

Each module uses the turbo principle in iterative decoding; that is, each module passes the extrinsic messages to the next module. The three modules are executed iteratively until convergence.

Algorithm 1: GEC-SR [32].

Input: Measurements $\mathbf{y}$, linear transformation $\mathbf{A}$, likelihood $\mathsf{P}_{\text{out}}(\mathbf{y}|\mathbf{z})$, and prior distribution $\mathsf{P}(\mathbf{x})$.
Output: Recovered signal $\hat{\mathbf{x}} = \hat{\boldsymbol{\mu}}_{1\mathbf{x}}$.
Initialize: $t \leftarrow 1$, $\boldsymbol{\mu}_{1\mathbf{z}} \leftarrow \mathbf{0}$, $\boldsymbol{\mu}_{2\mathbf{x}} \leftarrow \mathbf{0}$, $\mathbf{v}_{1\mathbf{z}} \leftarrow P_z\mathbf{1}$, and $\mathbf{v}_{2\mathbf{x}} \leftarrow P_x\mathbf{1}$.
for $t = 1, \cdots, t_{\max}$ **do**
  Module A:
  (1) Compute the posterior mean and variance of $\mathbf{z}$
1     $(\hat{\boldsymbol{\mu}}_{1\mathbf{z}}, \hat{\mathbf{v}}_{1\mathbf{z}}) = \mathbb{E}_{\mathsf{P}_{\text{out}}(y|z)}\{\mathbf{z}|\boldsymbol{\mu}_{1\mathbf{z}}, \mathbf{v}_{1\mathbf{z}}\}$.
  (2) Compute the extrinsic information of $\mathbf{z}$
2     $(\boldsymbol{\mu}_{2\mathbf{z}}, \mathbf{v}_{2\mathbf{z}}) = \mathbb{E}\mathrm{xt}\{(\hat{\boldsymbol{\mu}}_{1\mathbf{z}}, \hat{\mathbf{v}}_{1\mathbf{z}}) \setminus (\boldsymbol{\mu}_{1\mathbf{z}}, \mathbf{v}_{1\mathbf{z}})\}$.
  Module C (forward):
  (3) Compute the posterior mean and covariance of $\mathbf{x}$
    from the linear space
3     $(\hat{\boldsymbol{\mu}}_{2\mathbf{x}}, \hat{\mathbf{V}}_{2\mathbf{x}}) =$
    $\mathbb{E}_{\mathcal{N}}\{\mathbf{x}|\boldsymbol{\mu}_{2\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}), \boldsymbol{\mu}_{2\mathbf{z}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}})\}$.
  (4) Compute the extrinsic information of $\mathbf{x}$
4     $(\boldsymbol{\mu}_{1\mathbf{x}}, \mathbf{v}_{1\mathbf{x}}) = \mathbb{E}\mathrm{xt}\{(\hat{\boldsymbol{\mu}}_{2\mathbf{x}}, \mathrm{diag}(\hat{\mathbf{V}}_{2\mathbf{x}})) \setminus (\boldsymbol{\mu}_{2\mathbf{x}}, \mathbf{v}_{2\mathbf{x}})\}$.
  Module B:
  (5) Compute the posterior mean and variance of $\mathbf{x}$
5     $(\hat{\boldsymbol{\mu}}_{1\mathbf{x}}, \hat{\mathbf{v}}_{1\mathbf{x}}) = \mathbb{E}_{\mathsf{P}(x)}\{\mathbf{x}|\boldsymbol{\mu}_{1\mathbf{x}}, \mathbf{v}_{1\mathbf{x}}\}$.
  (6) Compute the extrinsic information of $\mathbf{x}$
6     $(\boldsymbol{\mu}_{2\mathbf{x}}, \mathbf{v}_{2\mathbf{x}}) = \mathbb{E}\mathrm{xt}\{(\hat{\boldsymbol{\mu}}_{1\mathbf{x}}, \hat{\mathbf{v}}_{1\mathbf{x}}) \setminus (\boldsymbol{\mu}_{1\mathbf{x}}, \mathbf{v}_{1\mathbf{x}})\}$.
  Module C (backward):
  (7) Compute the posterior mean and covariance of $\mathbf{z}$
    from the linear space
7     $(\hat{\boldsymbol{\mu}}_{2\mathbf{x}}, \hat{\mathbf{V}}_{2\mathbf{x}}) =$
    $\mathbb{E}_{\mathcal{N}}\{\mathbf{x}|\boldsymbol{\mu}_{2\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}), \boldsymbol{\mu}_{2\mathbf{z}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}})\}$
8     $(\hat{\boldsymbol{\mu}}_{2\mathbf{z}}, \hat{\mathbf{V}}_{2\mathbf{z}}) = (\mathbf{A}\hat{\boldsymbol{\mu}}_{2\mathbf{x}}, \mathbf{A}\hat{\mathbf{V}}_{2\mathbf{x}}\mathbf{A}^H)$
  (8) Compute the extrinsic information of $\mathbf{z}$
9     $(\mathbf{v}_{1\mathbf{z}}, \boldsymbol{\mu}_{1\mathbf{z}}) = \mathbb{E}\mathrm{xt}\left\{(\hat{\boldsymbol{\mu}}_{2\mathbf{z}}, \mathrm{diag}(\hat{\mathbf{V}}_{2\mathbf{z}}) \setminus (\boldsymbol{\mu}_{2\mathbf{z}}, \mathbf{v}_{2\mathbf{z}})\right\}$.

We present the GEC-SR algorithm in Algorithm 1. Before detailing the algorithm, we introduce notations. Given that the extrinsic messages are modeled as Gaussian, we will frequently employ a special version of the posterior mean and variance as in (4)–(5), while modeling $\mathsf{P}(z)$ as Gaussian. Therefore, for ease of notation, we denote by

$$\hat{\mu}_{1a} = \mathtt{E}_f\{a|\mu_{1a}, v_{1a}\}, \quad \hat{v}_{1a} = \mathtt{Var}_f\{a|\mu_{1a}, v_{1a}\} \qquad (14)$$

the posterior mean and variance of $a \sim f(a)$, respectively, where the expectations are taken w.r.t.

$$\frac{f(a)\mathcal{N}_{\mathbb{C}}(a; \mu_{1a}, v_{1a})}{\mathcal{Z}_a} \quad (15)$$

with $\mathcal{Z}_a = \int f(a)\mathcal{N}_{\mathbb{C}}(a; \mu_{1a}, v_{1a})da$ being the normalization. If $f(a) = \mathsf{P}_{\text{out}}(y|a)$ given by (6), then the posterior mean and variance can be obtained explicitly and given in Table I. If $f(a)$ is the Gaussian-Bernoulli distribution given by (8), then the posterior mean and variance are given in the third line of Table I. In addition, we denote by

$$\mu_{2a} = \mathtt{Ext}_\mu\{(\hat{\mu}_{1a}, \hat{v}_{1a}) \setminus (\mu_{1a}, v_{1a})\}, \quad (16a)$$

$$v_{2a} = \mathtt{Ext}_v\{(\hat{\mu}_{1a}, \hat{v}_{1a}) \setminus (\mu_{1a}, v_{1a})\} \quad (16b)$$

the extrinsic mean and variance of a, respectively, which are calculated by excluding the prior mean and variance (μ_{1a}, v_{1a}) given as

$$\mathtt{Ext}_v\{(\hat{\mu}_{1a}, \hat{v}_{1a}) \setminus (\mu_{1a}, v_{1a})\} \quad = \left(\frac{1}{\hat{v}_{1a}} - \frac{1}{v_{1a}}\right)^{-1}, \quad (17a)$$

$$\mathtt{Ext}_\mu\{(\hat{\mu}_{1a}, \hat{v}_{1a}) \setminus (\mu_{1a}, v_{1a})\} \quad = v_{2a}\left(\frac{\hat{\mu}_{1a}}{\hat{v}_{1a}} - \frac{\mu_{1a}}{v_{1a}}\right). \quad (17b)$$

For ease of notation, we often write the extrinsic mean and variance in pairs as

$$(\mu_{2a}, v_{2a}) = \mathbb{Ext}\{(\hat{\mu}_{1a}, \hat{v}_{1a}) \setminus (\mu_{1a}, v_{1a})\}. \quad (18)$$

Similarly,

$$(\hat{\mu}_{1a}, \hat{v}_{1a}) = \mathbb{E}_f\{a|\mu_{1a}, v_{1a}\}. \quad (19)$$

If the argument is a vector, then $\mathtt{E}_f$, $\mathtt{Var}_f$, $\mathtt{Ext}_\mu$, $\mathtt{Ext}_v$, $\mathbb{E}_f$, and $\mathbb{Ext}$ are performed element-wise. In GEC-SR, each module is executed iteratively, and the extrinsic messages serve as the input priors for the next module. For ease of notation, we use the subscripts $(\cdot)_1$ and $(\cdot)_2$ to specify the input priors and extrinsic messages, respectively. That is, if the input priors are denoted by $(\cdot)_1$, then their corresponding extrinsic messages are denoted by $(\cdot)_2$, and vice versa.

In GEC-SR, another set of posterior means and variances of $\mathbf{x}$ and $\mathbf{z}$ are constrained into the linear space $\mathbf{z} = \mathbf{Ax}$, i.e., Module C. Given the priors $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{z}; \boldsymbol{\mu}_{2\mathbf{z}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}}))$, $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{x}; \boldsymbol{\mu}_{2\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}))$, and $\delta(\mathbf{z} - \mathbf{Ax})$, the posterior mean estimate of $\mathbf{x}$ is the LMMSE estimate given by

$$\hat{\boldsymbol{\mu}}_{2\mathbf{x}} = \hat{\mathbf{V}}_{2\mathbf{x}}\left(\mathrm{Diag}(\mathbf{v}_{2\mathbf{x}})^{-1}\boldsymbol{\mu}_{2\mathbf{x}} + \mathbf{A}^H\mathrm{Diag}(\mathbf{v}_{2\mathbf{z}})^{-1}\boldsymbol{\mu}_{2\mathbf{z}}\right), \quad (20)$$

where

$$\hat{\mathbf{V}}_{2\mathbf{x}} = \left[\mathrm{Diag}(\mathbf{v}_{2\mathbf{x}})^{-1} + \mathbf{A}^H\mathrm{Diag}(\mathbf{v}_{2\mathbf{z}})^{-1}\mathbf{A}\right]^{-1} \quad (21)$$

is the posterior variance of $\mathbf{x}$. The posterior mean and variance of $\mathbf{z}$ are respectively given by

$$\hat{\boldsymbol{\mu}}_{2\mathbf{z}} = \mathbf{A}\hat{\boldsymbol{\mu}}_{2\mathbf{x}}, \quad \hat{\mathbf{V}}_{2\mathbf{z}} = \mathbf{A}\hat{\mathbf{V}}_{2\mathbf{x}}\mathbf{A}^H. \quad (22)$$

Similarly, for ease of notation, we write

$$(\hat{\boldsymbol{\mu}}_{2\mathbf{x}}, \hat{\mathbf{V}}_{2\mathbf{x}}) = \mathbb{E}_{\mathcal{N}}\{\mathbf{x}|\boldsymbol{\mu}_{2\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}), \boldsymbol{\mu}_{2\mathbf{z}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}})\}. \quad (23)$$

Here, to distinguish the notation in (19), we use notation $\mathbb{E}_{\mathcal{N}}\{\cdot|\boldsymbol{\mu}_{2\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}), \boldsymbol{\mu}_{2\mathbf{z}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}})\}$ to specify that the expectation is taken w.r.t. Gaussian for $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{x}; \boldsymbol{\mu}_{2\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}))$ and $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{z}; \boldsymbol{\mu}_{2\mathbf{z}}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}}))$.

Now, we detail Algorithm 1 as follows. Line 1 computes the posterior mean and variance of $\mathbf{z}$, which combine the conditional distributions of measurements $\mathbf{y}$ and the priors $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{z}; \boldsymbol{\mu}_{1\mathbf{z}}, \mathrm{Diag}(\mathbf{v}_{1\mathbf{z}}))$. The mean and variance of the priors $\mathbf{z}$ are determined by initialization or by the output of Module C (line 9). Line 2 computes the extrinsic information of $\mathbf{z}$ by excluding its prior mean and variance $(\boldsymbol{\mu}_{1\mathbf{z}}, \mathbf{v}_{1\mathbf{z}})$. The two procedures (i.e., computing the posterior information and calculating the extrinsic information) comprise Module A and are performed in each model. Line 3 (Module C) computes the posterior information of $\mathbf{x}$ under the linear space $\delta(\mathbf{z} - \mathbf{Ax})$, and line 4 calculates the extrinsic information. Line 5 (Module B) computes the posterior information of $\mathbf{x}$ by considering the true prior $\mathsf{P}(\mathbf{x})$ and the priors $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{x}; \boldsymbol{\mu}_{1\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{1\mathbf{x}}))$, and line 6 calculates the extrinsic information. Line 7 performs the same procedures as those in line 3 (Module C). The output of this step is $\mathbf{z}$ rather than $\mathbf{x}$. Thus, we perform line 8 to compute the posterior mean and variance of $\mathbf{z}$. Line 9 computes the extrinsic information of $\mathbf{z}$ and passes it to line 1 (Module A) as prior information.

III. DeGEC-SR

A. Algorithm

Before proceeding, we analyze the computational complexity of GEC-SR. Recall that each module of GEC-SR inputs the prior mean and variance and calculates the posterior and extrinsic mean and variance. The calculations of the extrinsic information, i.e., (17a)–(17b), are element-wise on the input vector. Therefore, their computational complexities are linear with the input size. Except for Module C, the expectation operations $\mathtt{E}\{\cdot\}$ in (14) are performed element-wise and are, thus, simple. Consequently, the computational load is mainly at lines 3 and 7 [precisely (21)], which involve matrix inversions. The calculation of the matrix inversion of an extremely large matrix (e.g., $M \times N$ matrix with $N = 1,024^2$ and $M = 4N$) usually becomes prohibitive for a personal computer.

To address the scalability issue, we partition M measurements into L independent clusters, each of which contains $M_l = M/L$ measurements in the lth cluster. For cluster l, we use superscript $[l]$ to specify the estimated parameters in each cluster and subscript l to represent the subset of parameters. Then, the measurements (1) in cluster l can be written as

$$\mathbf{y}_l = \mathtt{Q}(\mathbf{z}_l) = \mathtt{Q}(\mathbf{A}_l\mathbf{x} + \mathbf{w}_l), \quad (24)$$

where $\mathbf{y}_l$ and $\mathbf{z}_l$ measure M_l and $\mathbf{A}_l \in \mathbb{C}^{M_l \times N}$. Now, the $M \times N$ transform matrix $\mathbf{A}$ is split into L sub-matrices $\mathbf{A}_l$ with sizes $M_l \times N$. If GEC-SR is performed independently on each cluster to estimate $\mathbf{x}^{[l]}$ from the observed measurements $\mathbf{y}_l$, then the computational burden in each local computing processor is lower than the original computational burden in centralized processing. For example, given $M_l < N$, we apply
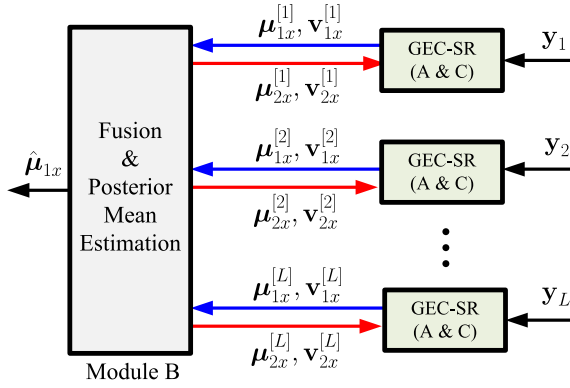
Fig. 2. Block diagram of deGEC-SR algorithm.

the Woodbury matrix identity to (21) and obtain

$$\left[\mathbf{D}_{x^{[l]}}^{-1} + \mathbf{A}_l^H \mathbf{D}_{z^{[l]}}^{-1} \mathbf{A}_l\right]^{-1}$$
$$= \mathbf{D}_{x^{[l]}} - \mathbf{D}_{x^{[l]}} \mathbf{A}_l^H \left[\mathbf{D}_{z^{[l]}} + \mathbf{A}_l \mathbf{D}_{x^{[l]}} \mathbf{A}_l^H\right]^{-1} \mathbf{A}_l \mathbf{D}_{x^{[l]}}.$$
(25)

where $\mathbf{D}_{x^{[l]}} = \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}^{[l]})$ and $\mathbf{D}_{z^{[l]}} = \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}}^{[l]})$. At this point, the complexity order for the matrix inversion in (25) is $N M_l^2$. Considering an extreme case, in which $\mathbf{A}_l$ is a $1 \times N$ matrix, the matrix inversion is only $O(N)$. The decomposition significantly reduces the complexity of GEC-SR.

The block diagram of deGEC-SR is illustrated in Fig. 2, which shows L parallel GEC-SR operations and a fusion center. GEC-SR is performed independently on each cluster to estimate $\mathbf{x}^{[l]}$ from the observed measurements $\mathbf{y}_l$. In deGEC-SR, Module A performs the de-nonlinear process, and Module C computes the posterior means and variances of $\mathbf{x}^{[l]}$, which are identical to those in Algorithm 1. The output of GEC-SR on each cluster is $\mathbf{x}^{[l]}$. We demonstrate a way to deal with the consensus problem in the fusion center, i.e., the solutions for the clusters should be equal.

As the measurements $\mathbf{y}_l$'s go through Modules A and C, we obtain $(\boldsymbol{\mu}_{1\mathbf{x}}^{[l]}, \mathbf{v}_{1\mathbf{x}}^{[l]})$ for $l = 1, \ldots, L$. After feeding them into Module B, we insert a constraint to establish $\mathbf{x} = \mathbf{x}^{[1]} = \mathbf{x}^{[2]} = \cdots = \mathbf{x}^{[L]}$. To this end, we introduce

$$\delta\left(\mathbf{x} - \mathbf{x}^{[1]}\right) \delta\left(\mathbf{x} - \mathbf{x}^{[2]}\right) \cdots \delta\left(\mathbf{x} - \mathbf{x}^{[L]}\right). \quad (26)$$

Given the priors $\mathbf{x}^{[l]} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{x}^{[l]}; \boldsymbol{\mu}_{1\mathbf{x}}^{[l]}, \mathrm{Diag}(\mathbf{v}_{1\mathbf{x}}^{[l]})), l = 1, \ldots, L$ and the constraint (26), we can obtain a new Gaussian distribution

$$\mathcal{N}_{\mathbb{C}}(\mathbf{x}; \boldsymbol{\mu}_{1\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{1\mathbf{x}}))$$
$$\propto \prod_{l=1}^{L} \mathcal{N}_{\mathbb{C}}\left(\mathbf{x}^{[l]}; \boldsymbol{\mu}_{1\mathbf{x}}^{[l]}, \mathrm{Diag}\left(\mathbf{v}_{1\mathbf{x}}^{[l]}\right)\right) \delta\left(\mathbf{x} - \mathbf{x}^{[l]}\right) \quad (27)$$

where

$$\boldsymbol{\mu}_{1\mathbf{x}} = \mathbf{v}_{1\mathbf{x}} \left(\sum_{l=1}^{L} \boldsymbol{\mu}_{1\mathbf{x}}^{[l]} / \mathbf{v}_{1\mathbf{x}}^{[l]}\right), \quad \mathbf{v}_{1\mathbf{x}} = \left(\sum_{l=1}^{L} \mathbf{1} / \mathbf{v}_{1\mathbf{x}}^{[l]}\right)^{-1}. \quad (28)$$

Algorithm 2: deGEC-SR.

Input: $(\mathbf{y}_l, \mathbf{A}_l, \mathsf{P}_{\mathrm{out}}(\mathbf{y}_l | \mathbf{z}_l))$, for $l = 1, \cdots, L$, and prior distribution $\mathsf{P}(\mathbf{x})$.
Output: Recovered signal $\hat{\mathbf{x}} = \hat{\boldsymbol{\mu}}_{1\mathbf{x}}$.
Initialize: $t \leftarrow 1$, $\boldsymbol{\mu}_{1\mathbf{z}}^{[l]} \leftarrow \mathbf{0}$, $\boldsymbol{\mu}_{2\mathbf{x}}^{[l]} \leftarrow \mathbf{0}$, $\mathbf{v}_{1\mathbf{z}}^{[l]} \leftarrow P_z^{[l]} \mathbf{1}$, and $\mathbf{v}_{2\mathbf{x}}^{[l]} \leftarrow P_x \mathbf{1}$.
for $t = 1, \cdots, t_{\max}$ **do**
  Module A:
  For cluster $l = 1, \ldots, L$,
  (1) Compute the posterior mean and variance of $\mathbf{z}_l$
1      $(\hat{\boldsymbol{\mu}}_{1\mathbf{z}}^{[l]}, \hat{\mathbf{v}}_{1\mathbf{z}}^{[l]}) = \mathbb{E}_{\mathsf{P}_{\mathrm{out}}(y|z)}\left\{\mathbf{z}_l \middle| \boldsymbol{\mu}_{1\mathbf{z}}^{[l]}, \mathbf{v}_{1\mathbf{z}}^{[l]}\right\}.$
  (2) Compute the extrinsic information of $\mathbf{z}_l$
2      $(\boldsymbol{\mu}_{2\mathbf{z}}^{[l]}, \mathbf{v}_{2\mathbf{z}}^{[l]}) = \mathbb{E}\mathrm{xt}\left\{(\hat{\boldsymbol{\mu}}_{1\mathbf{z}}^{[l]}, \hat{\mathbf{v}}_{1\mathbf{z}}^{[l]}) \setminus (\boldsymbol{\mu}_{1\mathbf{z}}^{[l]}, \mathbf{v}_{1\mathbf{z}}^{[l]})\right\}.$
  Module C (forward):
  For cluster $l = 1, \ldots, L$,
  (3) Compute the posterior mean and covariance of $\mathbf{x}^{[l]}$
     from the linear space
3      $(\hat{\boldsymbol{\mu}}_{2\mathbf{x}}^{[l]}, \hat{\mathbf{V}}_{2\mathbf{x}}^{[l]}) =$
     $\mathbb{E}_{\mathcal{N}}\left\{\mathbf{x}^{[l]} \middle| \boldsymbol{\mu}_{2\mathbf{x}}^{[l]}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}^{[l]}), \boldsymbol{\mu}_{2\mathbf{z}}^{[l]}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}}^{[l]})\right\}.$
  (4) Compute the extrinsic information of $\mathbf{x}^{[l]}$
4      $(\boldsymbol{\mu}_{1\mathbf{x}}^{[l]}, \mathbf{v}_{1\mathbf{x}}^{[l]}) = \mathbb{E}\mathrm{xt}\left\{(\hat{\boldsymbol{\mu}}_{2\mathbf{x}}^{[l]}, \mathrm{diag}(\hat{\mathbf{V}}_{2\mathbf{x}}^{[l]})) \setminus (\boldsymbol{\mu}_{2\mathbf{x}}^{[l]}, \mathbf{v}_{2\mathbf{x}}^{[l]})\right\}.$
  Module B:
  (5) Compute the fusion of $\mathbf{x}$ with
     $(\boldsymbol{\mu}_{1\mathbf{x}}^{[l]}, \mathbf{v}_{1\mathbf{x}}^{[l]}), l = 1, \ldots, L$
5      $(\boldsymbol{\mu}_{1\mathbf{x}}, \mathbf{v}_{1\mathbf{x}}) = \mathrm{Eq.}\ (28).$
  (6) Compute the posterior mean and variance of $\mathbf{x}$
6      $(\hat{\boldsymbol{\mu}}_{1\mathbf{x}}, \hat{\mathbf{v}}_{1\mathbf{x}}) = \mathbb{E}_{\mathsf{P}(x)}\{\mathbf{x} | \boldsymbol{\mu}_{1\mathbf{x}}, \mathbf{v}_{1\mathbf{x}}\}.$
  (7) Compute the extrinsic information of $\mathbf{x}^{[l]}$ for cluster
     $l = 1, \ldots, L$,
7      $(\boldsymbol{\mu}_{2\mathbf{x}}^{[l]}, \mathbf{v}_{2\mathbf{x}}^{[l]}) = \mathbb{E}\mathrm{xt}\left\{(\hat{\boldsymbol{\mu}}_{1\mathbf{x}}, \hat{\mathbf{v}}_{1\mathbf{x}}) \setminus (\boldsymbol{\mu}_{1\mathbf{x}}^{[l]}, \mathbf{v}_{1\mathbf{x}}^{[l]})\right\}.$
  Module C (backward):
  For cluster $l = 1, \ldots, L$,
  (8) Compute the posterior mean and covariance of $\mathbf{z}_l$
     from the linear space
8      $(\hat{\boldsymbol{\mu}}_{2\mathbf{x}}^{[l]}, \hat{\mathbf{V}}_{2\mathbf{x}}^{[l]}) =$
     $\mathbb{E}_{\mathcal{N}}\left\{\mathbf{x}^{[l]} \middle| \boldsymbol{\mu}_{2\mathbf{x}}^{[l]}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{x}}^{[l]}), \boldsymbol{\mu}_{2\mathbf{z}}^{[l]}, \mathrm{Diag}(\mathbf{v}_{2\mathbf{z}}^{[l]})\right\}$
9      $(\hat{\boldsymbol{\mu}}_{2\mathbf{z}}^{[l]}, \hat{\mathbf{V}}_{2\mathbf{z}}^{[l]}) = (\mathbf{A}_l \hat{\boldsymbol{\mu}}_{2\mathbf{x}}^{[l]}, \mathbf{A}_l \hat{\mathbf{V}}_{2\mathbf{x}}^{[l]} \mathbf{A}_l^H)$
  (9) Compute the extrinsic information of $\mathbf{z}_l$
10     $(\mathbf{v}_{1\mathbf{z}}^{[l]}, \boldsymbol{\mu}_{1\mathbf{z}}^{[l]}) = \mathbb{E}\mathrm{xt}\left\{(\hat{\boldsymbol{\mu}}_{2\mathbf{z}}^{[l]}, \mathrm{diag}(\hat{\mathbf{V}}_{2\mathbf{z}}^{[l]}) \setminus (\boldsymbol{\mu}_{2\mathbf{z}}^{[l]}, \mathbf{v}_{2\mathbf{z}}^{[l]})\right\}.$

Then, similar to that in the original GEC-SR, Module B computes the posterior information $(\hat{\boldsymbol{\mu}}_{1\mathbf{x}}, \hat{\mathbf{v}}_{1\mathbf{x}})$ by considering the true prior $\mathsf{P}(\mathbf{x})$ and the priors $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{x}; \boldsymbol{\mu}_{1\mathbf{x}}, \mathrm{Diag}(\mathbf{v}_{1\mathbf{x}}))$ given by (27). Next, Module B computes the extrinsic information of $\mathbf{x}^{[l]}$ for each cluster by excluding the prior mean and variance $(\boldsymbol{\mu}_{1\mathbf{x}}^{[l]}, \mathbf{v}_{1\mathbf{x}}^{[l]})$, i.e.,

$$\left(\boldsymbol{\mu}_{2\mathbf{x}}^{[l]}, \mathbf{v}_{2\mathbf{x}}^{[l]}\right) = \mathbb{E}\mathrm{xt}\left\{(\hat{\boldsymbol{\mu}}_{1\mathbf{x}}, \hat{\mathbf{v}}_{1\mathbf{x}}) \setminus \left(\boldsymbol{\mu}_{1\mathbf{x}}^{[l]}, \mathbf{v}_{1\mathbf{x}}^{[l]}\right)\right\}. \quad (29)$$

The extrinsic information $(\boldsymbol{\mu}_{2\mathbf{x}}^{[l]}, \mathbf{v}_{2\mathbf{x}}^{[l]})$ is then given as feed-back to Module C of each cluster. Subsequently, the message is sent to Modules C and A, and the procedure is repeated. We present the deGEC-SR algorithm in Algorithm 2.

When deGEC-SR is applied to a distributed network, Modules A and C can be performed in local sensors. Module B in the fusion center collects the posterior information from local sensors and then broadcasts the combined posterior information

(28). With the combined posterior information, each sensor then computes the extrinsic information (29) for the next update. Compared with the centralized GEC-SR, deGEC-SR does not need to store the high-dimensional transform matrix (i.e., $\mathbf{A}$). The computation loading of Module C is also shared by each local sensor.

B. Damping Factor

The extrinsic messages that pass among the modules of deGEC-SR (or GEC-SR) are assumed to be Gaussian distributions. However, these messages might not come close to Gaussian distributions in practice, particularly at the low sampling rate M/N. In this case, deGEC-SR (or GEC-SR) appears to have some unexpected numerical issues and diverges. The same problem is also observed in some message-passing algorithms [30]–[33], [35]. To improve robustness, we use the damping method [20], [31], [33]. Specifically, we smoothen the update of $(\mathbf{x}, \mathbf{z})$ by using a convex combination with the former value:

$$\texttt{Damp}(a(t+1)) = \beta a(t) + (1-\beta)a(t+1), \qquad (30)$$

where $\beta \in [0,1]$ is the damping factor. For ease of notation, we still use $a(t+1)$ as the output of the damping function, i.e., $a(t+1) := \texttt{Damp}(a(t+1))$. For deGEC-SR, we plug the damping function after lines 2 and 7, respectively, as

$$\boldsymbol{\mu}_{2\mathbf{z}}^{[l]}(t+1) := \texttt{Damp}\left(\boldsymbol{\mu}_{2\mathbf{z}}^{[l]}(t+1)\right), \qquad (31a)$$

$$\mathbf{v}_{2\mathbf{z}}^{[l]}(t+1) := \texttt{Damp}\left(\mathbf{v}_{2\mathbf{z}}^{[l]}(t+1)\right), \qquad (31b)$$

and

$$\boldsymbol{\mu}_{2\mathbf{x}}^{[l]}(t+1) := \texttt{Damp}\left(\boldsymbol{\mu}_{2\mathbf{x}}^{[l]}(t+1)\right), \qquad (32a)$$

$$\mathbf{v}_{2\mathbf{x}}^{[l]}(t+1) := \texttt{Damp}\left(\mathbf{v}_{2\mathbf{x}}^{[l]}(t+1)\right). \qquad (32b)$$

C. State Evolution

In this subsection, we demonstrate that the iteration performance of deGEC-SR can be characterized by a set of SE equations. Our derivation is based on a large system limit, i.e., $M, M_l, N \to \infty$ with fixed sampling ratios

$$\gamma = M/N, \quad \gamma_l = M_l/N. \qquad (33)$$

In the large system limit, each component of the variance vectors (e.g., $\mathbf{v}_{1\mathbf{z}}, \hat{\mathbf{v}}_{1\mathbf{z}}, \cdots$) in deGEC-SR has a similar value to their self-average (or expectation) results. With the self-average property in mind, we are able to obtain explicit expressions for the SE equations. As we will use a self-average value to represent each variance vector, we list all the concerned notations below

$$\mathbf{v}_{1\mathbf{z}}^{[l]} \to v_{1z}^{[l]}, \; \hat{\mathbf{v}}_{1\mathbf{z}}^{[l]} \to \hat{v}_{1z}^{[l]}, \; \mathbf{v}_{1\mathbf{x}}^{[l]} \to v_{1x}^{[l]}, \; \hat{\mathbf{v}}_{1\mathbf{x}}^{[l]} \to \hat{v}_{1x}^{[l]},$$

$$\mathbf{v}_{2\mathbf{z}}^{[l]} \to v_{2z}^{[l]}, \; \text{diag}\left(\hat{\mathbf{V}}_{2\mathbf{z}}^{[l]}\right) \to \hat{v}_{2z}^{[l]}, \; \mathbf{v}_{2\mathbf{x}}^{[l]} \to v_{2x}^{[l]},$$

$$\text{diag}\left(\hat{\mathbf{V}}_{2\mathbf{x}}^{[l]}\right) \to \hat{v}_{2x}^{[l]}.$$

Here, the left-hand side of the arrow represents the original variance vector, and the right-hand side represents the corresponding self-average value in the large system limit.

Let us start by considering line 1 of Algorithm 2. The average value of the posterior variance $\hat{v}_{1z}^{[l]}$ given prior information $(\boldsymbol{\mu}_{1z}^{[l]}(t), \mathbf{v}_{1z}^{[l]}(t))$ can be characterized by

$$\hat{v}_{1z}^{[l]}(t+1) = \mathbb{E}\left\{\texttt{Var}_{\mathsf{P}_{\text{out}}(y|z)}\left\{z\Big|\mu_{1z}^{[l]}(t), v_{1z}^{[l]}(t)\right\}\right\}, \qquad (34)$$

where $\texttt{Var}_{\mathsf{P}_{\text{out}}(y|z)}$ is calculated by following its definition in (14) and the outer expectation is taken w.r.t. y and $\mu_{1z}^{[l]}$. Note that $\mu_{1z}^{[l]}(t)$ is the linear reconstruction from the previous iteration and is a random variable. We model it as $\mu_{1z}^{[l]}(t) \sim \mathcal{N}_{\mathbb{C}}(\mu_{1z}^{[l]}(t); 0, P_z - v_{1z}^{[l]}(t))$. The explicit expression of (34) is provided later. Line 2 computes the extrinsic variance, and the result can be characterized by

$$v_{2z}^{[l]}(t+1) = \left(\frac{1}{\hat{v}_{1z}^{[l]}(t+1)} - \frac{1}{v_{1z}^{[l]}(t)}\right)^{-1}. \qquad (35)$$

Then, we move to line 3. The posterior variance is shown in the left-hand side of (25). Let $\lambda_i^{[l]}$ be the ith eigenvalue of $\mathbf{A}_l^H \mathbf{A}_l$, which has a total of N eigenvalues.[2] Then, we obtain

$$\hat{v}_{2x}^{[l]}(t+1) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{v_{2x}^{[l]}(t)} + \frac{\lambda_i^{[l]}}{v_{2z}^{[l]}(t+1)}\right)^{-1}. $$

Note that if $M_l < N$, then $N - M_l$ of eigenvalues $\lambda_i^{[l]}$ are zero. For brevity, we denote the above expression as

$$\hat{v}_{2x}^{[l]}(t+1) = \left\langle\frac{1}{\frac{1}{v_{2x}^{[l]}(t)} + \frac{\lambda^{[l]}}{v_{2z}^{[l]}(t+1)}}\right\rangle. \qquad (36)$$

Line 4 computes the extrinsic variance, and the result is

$$v_{1x}^{[l]}(t+1) = \left(\frac{1}{\hat{v}_{2x}^{[l]}(t+1)} - \frac{1}{v_{2x}^{[l]}(t)}\right)^{-1}. \qquad (37)$$

Line 5 performs the combining, and yields

$$v_{1x}(t+1) = \left(\sum_{l=1}^{L}\frac{1}{v_{1x}^{[l]}(t+1)}\right)^{-1}. \qquad (38)$$

Line 6 computes the posterior variance, and we obtain

$$\hat{v}_{1x}(t+1) = \mathbb{E}\left\{\texttt{Var}_{\mathsf{P}(x)}\left\{x\Big|\mu_{1x}, v_{1x}(t+1)\right\}\right\}, \qquad (39)$$

where the outer expectation is taken w.r.t. μ_{1x}. We will explicitly explain the expression of (39) later. Line 7 computes the extrinsic variance of each cluster, and the result is given by

$$v_{2x}^{[l]}(t+1) = \left(\frac{1}{\hat{v}_{1x}(t+1)} - \frac{1}{v_{1x}^{[l]}(t+1)}\right)^{-1}. \qquad (40)$$

---

[2]The eigenvalues of $\mathbf{A}_l$ are required only in the SE analysis rather than on the algorithm implementation.

Line 8 performs similarly as line 3. Combining lines 8 and 9, we obtain

$$\hat{v}_{2z}^{[l]}(t+1) = \frac{1}{\gamma_l} \left\langle \frac{\lambda^{[l]}}{\frac{1}{v_{2x}^{[l]}(t+1)} + \frac{\lambda^{[l]}}{v_{2z}^{[l]}(t+1)}} \right\rangle. \tag{41}$$

Finally, line 10 computes the extrinsic variance, and we obtain

$$v_{1z}^{[l]}(t+1) = \left(\frac{1}{\hat{v}_{2z}^{[l]}(t+1)} - \frac{1}{v_{2z}^{[l]}(t+1)} \right)^{-1}. \tag{42}$$

In the absence of an advanced process during initialization, we could set $v_{1z}^{[l]}(0) = P_z$ and $v_{1x}^{[l]}(0) = P_x$. We summarize the above equations in Proposition 1.

Proposition 1: In the large-system limit, the SE equations of deGEC-SR, given the initialization $v_{1z}^{[l]}(0)$ and $v_{1x}^{[l]}(0)$, can be characterized by

$$\hat{v}_{1z}^{[l]}(t+1) = \mathtt{E}\left\{ \mathtt{Var}_{\mathrm{P}_{\mathrm{out}}(y|z)}\left\{ z \Big| \mu_{1z}^{[l]}(t), v_{1z}^{[l]}(t) \right\} \right\}, \tag{43a}$$

$$v_{2z}^{[l]}(t+1) = \left(\frac{1}{\hat{v}_{1z}^{[l]}(t+1)} - \frac{1}{v_{1z}^{[l]}(t)} \right)^{-1}, \tag{43b}$$

$$\hat{v}_{2x}^{[l]}(t+1) = \left\langle \frac{1}{\frac{1}{v_{2x}^{[l]}(t)} + \frac{\lambda^{[l]}}{v_{2z}^{[l]}(t+1)}} \right\rangle, \tag{43c}$$

$$v_{1x}^{[l]}(t+1) = \left(\frac{1}{\hat{v}_{2x}^{[l]}(t+1)} - \frac{1}{v_{2x}^{[l]}(t)} \right)^{-1}, \tag{43d}$$

$$v_{1x}(t+1) = \left(\sum_{l=1}^{L} \frac{1}{v_{1x}^{[l]}(t+1)} \right)^{-1}, \tag{43e}$$

$$\hat{v}_{1x}(t+1) = \mathtt{E}\left\{ \mathtt{Var}_{\mathrm{P}(x)}\left\{ x \Big| \mu_{1x}(t+1), v_{1x}(t+1) \right\} \right\}, \tag{43f}$$

$$v_{2x}^{[l]}(t+1) = \left(\frac{1}{\hat{v}_{1x}(t+1)} - \frac{1}{v_{1x}^{[l]}(t+1)} \right)^{-1}, \tag{43g}$$

$$\hat{v}_{2z}^{[l]}(t+1) = \frac{1}{\gamma_l} \left\langle \frac{\lambda^{[l]}}{\frac{1}{v_{2x}^{[l]}(t+1)} + \frac{\lambda^{[l]}}{v_{2z}^{[l]}(t+1)}} \right\rangle, \tag{43h}$$

$$v_{1z}^{[l]}(t+1) = \left(\frac{1}{\hat{v}_{2z}^{[l]}(t+1)} - \frac{1}{v_{2z}^{[l]}(t+1)} \right)^{-1}, \tag{43i}$$

where $t = 0, 1, 2 \ldots$ denotes the iteration index.

In the SE equations (43), all the expressions are explicit except for (43a) and (43f). We first consider (43a). For the PR model, the posterior variance $\mathtt{Var}_{\mathrm{P}_{\mathrm{out}}(y|z)}\{ z | \mu_{1z}^{[l]}(t), v_{1z}^{[l]}(t) \}$ is given by; here, the posterior variance is a function of y and $\mu_{1z}^{[l]}$. Note that y and $\mu_{1z}^{[l]}$ are random variables. By taking the expectation of the posterior variance over y and $\mu_{1z}^{[l]}$ (which is detailed in Appendix

A), we can obtain the following explicit expression of (43a):

$$\hat{v}_{1z}^{[l]}(t+1) = \frac{v_{1z}^{[l]}(t)\sigma_w^2}{v_{1z}^{[l]}(t) + \sigma_w^2} + \frac{v_{1z}^{[l]}(t)^2(P_{1z} + \sigma_w^2)}{\left(v_{1z}^{[l]}(t) + \sigma_w^2 \right)^2}$$
$$- \frac{4v_{1z}^{[l]}(t)^2}{\left(v_{1z}^{[l]}(t) + \sigma_w^2 \right)^3 \left(P_{1z} - v_{1z}^{[l]}(t) \right)}$$
$$\times F\left(v_{1z}^{[l]}(t), P_{1z}, \sigma_w^2 \right), \tag{44}$$

where the F-function is defined in (55). For the post-intensity noise model, $\mathtt{Var}_{\mathrm{P}_{\mathrm{out}}(y|z)}\{z|\mu_z, v_z\}$ in (48) should be replaced by line 2 in Table I. The expectation of $\mathtt{Var}_{\mathrm{P}_{\mathrm{out}}(y|z)}\{z|\mu_z, v_z\}$ can be derived by using the similar technique in Appendix A.

Next, we consider (43f). From line 2 of Table I, we can obtain the posterior variance $\mathtt{Var}_{\mathrm{P}(x)}\{x|\mu_{x1}, v_{x1}\}$, which is a function of μ_{x1}. By taking expectation of the posterior variance over μ_{x1} (which is detailed in Appendix B), we can obtain the following explicit expression of (43f):

$$\hat{v}_{1x}(t+1) = 1 - \frac{1}{v_{1x} + \rho^{-1}}$$
$$\times \int_0^\infty \frac{re^{-r}}{\rho + (1-\rho)(1 + (\rho v_{1x})^{-1})e^{-\frac{r}{\rho v_{1x}}}} dr, \tag{45}$$

where we denote $v_{1x} = v_{1x}(t+1)$ for brevity.

Finally, to reflect the damping factor in the SE, we can introduce it after (43b) and (43g) as

$$v_{2z}^{[l]}(t+1) := \mathtt{Damp}\left(v_{2z}^{[l]}(t+1) \right),$$
$$v_{2x}^{[l]}(t+1) := \mathtt{Damp}\left(v_{2x}^{[l]}(t+1) \right). \tag{46}$$

All the equations in Proposition 1 involve a simple calculation and can be used to evaluate the evolution trajectory of deGEC-SR. Proposition 1 provides a quick and efficient evaluation of the performance of the deGEC-SR algorithm.

IV. DISCUSSION AND NUMERICAL RESULTS

A. Damping Factor

Before exhaustive comparisons with other state-of-the-art PR algorithms can be conducted, we first perform a quick experiment to show the influence of the damping factor on GEC-SR. We consider an i.i.d. Gaussian transform matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ with measurement dimension $M = 4,000$ and signal dimension $N = 1,000$. Signals $\mathbf{x}$ are generated in accordance to the Gaussian-Bernoulli distribution (8) with sparse ratio $\rho = 0.5$. The elements of noise $\mathbf{w}$ are i.i.d. and follow $\mathcal{N}_{\mathbb{C}}(w; 0, \sigma_w^2)$, and the SNR of the simulation is defined as $\|\mathbf{A}\mathbf{x}\|^2/(M\sigma_w^2)$.

Fig. 3 shows the MSE as a function of iteration for GEC-SR under four damping factors, including no damping (i.e., $\beta = 0$), constant damping with $\beta = 0.1$, 0.9, and exponentially decreasing damping $\beta = 0.9^t$. GEC-SR shows instabilities under the settings with fast updates of the extrinsic information (i.e., $\beta = 0$ and $\beta = 0.1$). The slow update by setting $\beta = 0.9$ can prevent
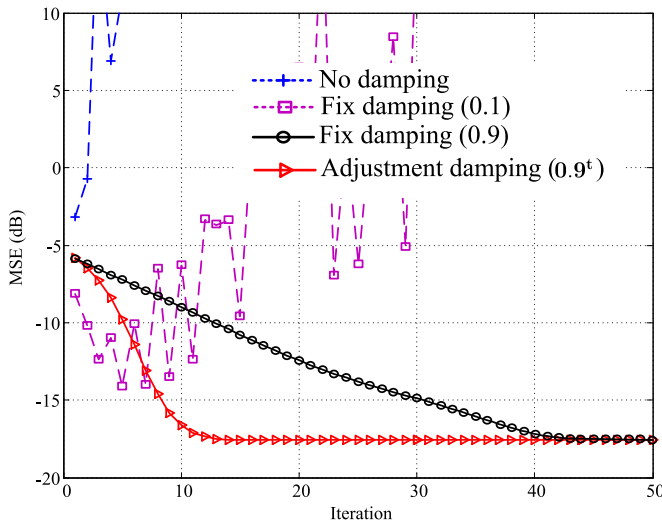
Fig. 3. MSE as a function of iteration for GEC-SR under four damping factors for $M = 4,000$, $N = 1,000$, and SNR = 20 dB. Transform matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ is an i.i.d. complex Gaussian matrix, and signal $\mathbf{x}$ follows the i.i.d. Gaussian-Bernoulli distribution with $\rho = 0.5$.

instabilities while slowing the algorithm to obtaining a stationary solution. Exponentially decreasing damping $\beta = 0.9^t$, which starts with a conservative value of β and decreases it exponentially with the number of iterations, shows better trade-off between instabilities and speed. Consequently, the damping factor determines the speed of the algorithm and control instabilities.

B. Performance Comparisons

Following the same settings as those in [34, Appendix A], we compare the performance and running time of two state-of-the-art PR algorithms, namely, prGAMP [20] and prVAMP [33], [34] with GEC-SR and deGEC-SR. Other algorithms (such as PhaseMax and PhaseLift) show worse performance compared with prGAMP and prVAMP in [33] and are excluded in the comparisons. All these algorithms require parameter tuning and initialization. For prGAMP and prVAMP, the parameter settings (e.g., signal priors and damping factors) are the same as those listed in [34, Table III]. For GEC-SR, the damping factor of 0.9 is used. All the algorithms are initialized the same as those used in [34].

First, we consider the case where signals $\mathbf{x}$ with dimension $N = 128$ are generated on the basis of the Gaussian distribution [i.e., sparse ratio $\rho = 1$ in (8)]. Two transform matrices with entries, namely, 1) i.i.d. standard Gaussian distribution and 2) i.i.d. $(0,1)$-binary distribution, are used. The SNR is set to 50 dB. Fig. 4 shows the reconstruction performance of the concerned algorithms in terms of MSE.[3] The simulation results are obtained by averaging over 100 realizations. We run all algorithms for 200 iterations. For Gaussian distributed transform matrices, all the algorithms operate closely well. GEC-SR can reconstruct the signals using slightly fewer measurements than that of competing methods. deGEC-SR slightly degenerates
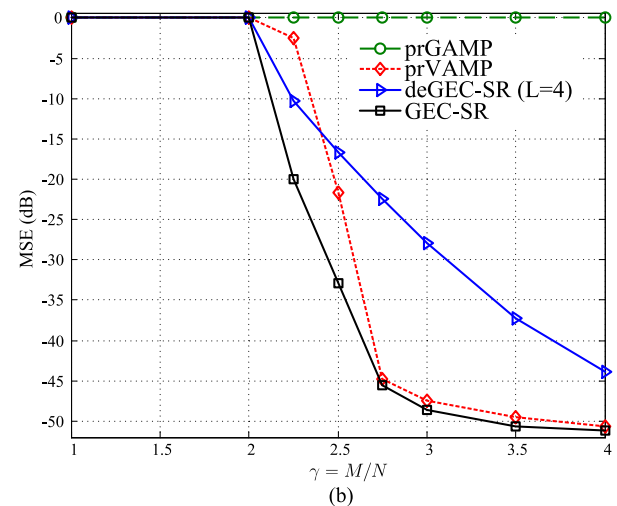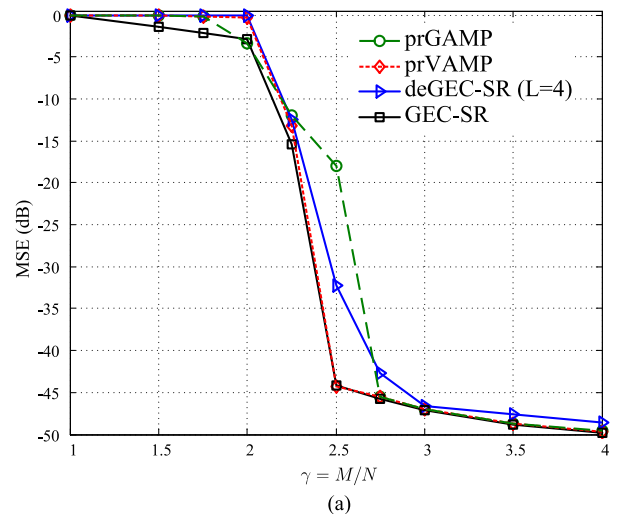
Fig. 4. MSE versus sampling rates γ for (a) Gaussian transform matrices and (b) $(0,1)$-binary transform matrices.

because of decomposition. For $(0,1)$-binary transform matrices, GEC-SR significantly requires fewer measurements than that of competing methods, prVAMP is the second best, and deGEC-SR requires a large sampling rate to achieve accuracy reconstruction. prGAMP fails in this regime.

Next, we consider the case in which the signals are obtained from a practical image. The image resolution is $N = 50 \times 50$. For the tests, we set $M = 4N$. We run all algorithms for 50 iterations. The reconstruction results are shown in Fig. 5 and 6 for the Gaussian and binary distributed transform matrices, respectively. Similar to the results in the first experiment, GEC-SR performs the best in terms of reconstruction performance, and deGEC-SR slightly degenerates. In the case of $(0,1)$-binary measurements, prGAMP fails because the transform matrix is not zero-mean.

Table II shows the average running time (referred to as time complexity) for 50 iterations. Compared with GEC-SR and prVAMP, deGEC-SR immensely reduces the computation time by processing multiple clusters simultaneously. The advantage is attributed to two reasons as follows: 1) The processing of SVD is disassembled into several SVD operations with small
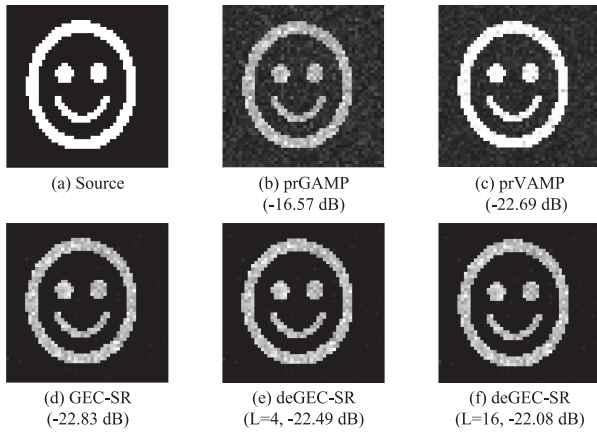
Fig. 5. Reconstructions (50×50) with various PR algorithms from i.i.d. standard Gaussian measurements for SNR = 10 dB.
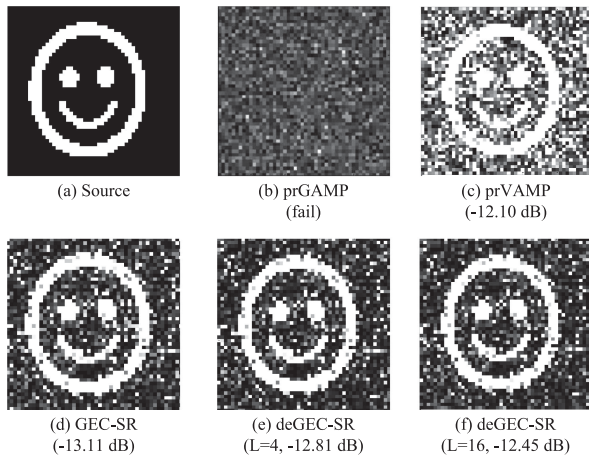

Fig. 6. Reconstructions (50×50) with various PR algorithms from $(0,1)$-binary measurements for SNR = 20 dB.

TABLE II
AVERAGE RUNNING TIMES OF DIFFERENT ALGORITHMS IN SECONDS

$N = 50 \times 50$ $M = 4N$	prVAMP	prGAMP	GEC-SR	deGEC-SR	
				$L = 4$	$L = 16$
SVD	32.14	–	32.14	20.70	1.56
Algorithm	8.02	5.10	12.06	16.83	6.01

dimensions, and the SVD operations for all clusters are simultaneously performed. 2) Matrix multiplication in the iteration is broken down into multiple matrix multiplications that can be simultaneously processed with small dimensions. Although the time complexity of prGAMP is extremely low, prGAMP is strictly limited by its transform matrix. Therefore, deGEC-SR possesses a low computation time and is robust to the type of transform matrix.

C. Accuracy of SE

Next, we verify the SE equations of deGEC-SR given in (43). We consider the case with $M = 4,000$ and $N = 1,000$. In particular, the transform matrix is constructed from SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$ are drawn from the normalized DFT matrices and all the singular values
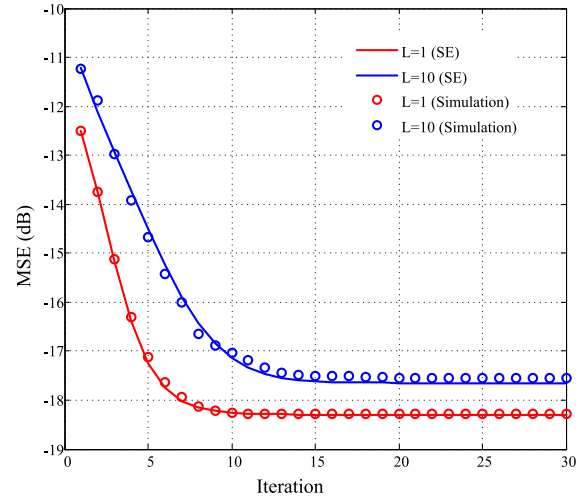

Fig. 7. Simulated and predicted MSEs for deGEC-SR under different clusters L. $\rho = 0.5$, $M = 4,000$, $N = 1,000$, and SNR = 20 dB. The simulated MSEs are averaged over 5,000 realizations.

are equal to one. Signal $\mathbf{x}$ follows i.i.d. Gaussian-Bernoulli distribution (8) with sparse ratio $\rho = 0.5$. The SNR is set to 20 dB, and the damping factor of 0.9^t is used.

Fig. 7 plots the simulation performance of deGEC-SR and the predicted performance of SE. The simulation results are obtained by averaging over 5,000 realizations. Given that GEC-SR diverges when the initials are improperly set,[4] we carefully select the initials as follows: $\boldsymbol{\mu}_{2\mathbf{x}} \leftarrow \mathbf{x}$ and $\boldsymbol{\mu}_{1\mathbf{z}} \leftarrow \sqrt{(P_z - v_{1z})/P_z}\mathbf{A}\mathbf{x} + \sqrt{v_{1z}(P_z - v_{1z})/P_z}\tilde{\mathbf{w}}$, where $\tilde{\mathbf{w}}$ is a standard Gaussian vector. With the initialization, the covariance of $(\mathbf{z}, \boldsymbol{\mu}_{1\mathbf{z}})$ can consistently follow with the theoretical derivation in (49). In addition, we set $v_{1z} = 0.05$ to ensure that GEC-SR is on its way to a convergence path. The damping factors for $\mathbf{x}$ and $\mathbf{z}$ are 0.9^t. As shown in the figure, the SE of deGEC-SR is accurate. The slight divergence between the simulations and SE predictions is because of the damping factor of $\mathbf{z}$, which introduces correlations among iterations.[5] The SE predictions are useful because performance predictions can be obtained without time-consuming Monte Carlo simulations.

D. Performance Degenerations due to Decentralization

The SE equations of deGEC-SR in (43) converge toward a fixed point, which determines the final performance of deGEC-SR. Although the computational complexity of deGEC-SR decreases with the number of clusters L, the performance of the algorithm degenerates. To understand how decentralization leads to performance degeneration, we then detail the evolution trajectory of deGEC-SR through SE. A similar technique called the extrinsic information transfer (EXIT) chart was first developed in the Turbo coding community [43]. Recently, the

---

[4]As surveyed in the Introduction, most non-lifting-based PR methods rely on accurate initializations.

[5]If without introducing the damping factors for $\mathbf{x}$ and $\mathbf{z}$, the GEC-SR to the PR problem cannot converge. To verify the accuracy of the SE equations, we also verify them under other nonlinear function $Q(\cdot)$ such as a complex-valued quantizer. In this case, the GEC-SR can work very well even without introducing the damping factors.
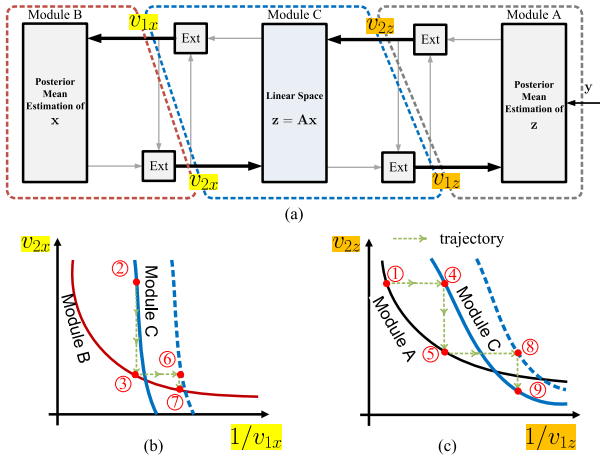
Fig. 8. (a) Relevant parameters (i.e., extrinsic information) in GEC-SR for analyzing evolution trajectory. (b) Transfer chart of Modules C and B. (c) Transfer chart of Modules C and A.

technique of EXIT chart has been used to design an optimal code for AMP [44].

Toward this end, we reproduce block diagrams of GEC-SR while removing the details and only highlighting the relevant parameters (i.e., extrinsic information) in Fig. 8(a). We first delineate the transfer charts of Modules A and C. For Module A, the input and output are characterized by v_{1z} and v_{2z}, respectively. For the right-hand side of Module C, the input and output are characterized by v_{2z} and v_{1z}, respectively, which are reversion of Module A. Therefore, if we plot the response curves of Modules A and C together, then the extrinsic messages exchanged between two curves (i.e., evolution trajectory) can be characterized, as shown in Fig. 8(c). Analogously, we can plot the transfer charts of Modules B and C. For Module B, the input and output are characterized by v_{1x} and v_{2x}, respectively. For the left-hand side of Module C, the input and output are reversed. Therefore, the evolution trajectory between Modules B and C can also be characterized, as shown in Fig. 8(b). If the response curve of Module C can be determined, then the evolution trajectory of v_{1x} can be sketched. Thus, the MSE of $\mathbf{x}$ for GEC-SR can be obtained. However, if we take a closer look at Module C, then we shall notice that output v_{1x} is determined not only by v_{2x}, but also by v_{2z}. Therefore, the response curve of Module C shown in Fig. 8(b) is obtained by the given fixed value of v_{2z}. Similarly, the response curve of Module C shown in Fig. 8(c) is obtained by the given value of v_{2x}.

On top of the response curves, we now detail the evolution trajectory. Recall from (13) that the iteration procedures of GEC-SR follow a circular manner. Therefore, the evolution trajectory shall start from a given v_{1z}. With a given v_{1z}, we can obtain v_{2z} on the basis of the response curve of Module A in Fig. 8(c), i.e., the point marked by ① . With v_{2z}, we can determine the response curve of Module C in Fig. 8(b). Then, given an initial value of v_{2x}, we can obtain v_{1x}, as marked by ② in Fig. 8(b). Thereafter, we obtain v_{2x} through the transfer function of Module B, as marked by ③ in Fig. 8(b). Given the new v_{2x}, we determine the response curve of Module C in Fig. 8(c). Next, we obtain v_{1z} through the response curve of Module C, as marked by ④ in Fig. 8(c). We also obtain v_{1z} through the response curve of

Module A, as marked by ⑤ in Fig. 8(c). Note that with the new v_{1z}, the response curve of Module C in Fig. 8(b) changes. Specifically, with the decrease of v_{2z}, the response curve of Module C shall move to the right, as shown by the dashed line in Fig. 8(b). We then obtain ⑥ and ⑦ in Fig. 8(b). The above processes are repeated (i.e., ⑧ , ⑨ , ...) until a fixed point is reached.

Fig. 9 shows a practical example of the evolution trajectory using the same setting as that in Fig. 7. Fig. 9(a) illustrates that $\hat{v}_{2x}$ starts from -12.51 dB and eventually converges to a fixed point at -18.29 dB; this result is consistent with that in Fig. 7. The response curve of Module C continuously moves to the right with the number of iterations and eventually reaches the saturation point. Other interesting observations are as follows. First, we plot another response curve of Module B with a different sparse ratio ρ in Fig. 9(a). We see that the response curve becomes lower as signal $\mathbf{x}$ becomes sparse. This characteristic implies that the MSE of $\mathbf{x}$ can converge to a small value as the signal becomes sparse. Second, we plot another response curve of Module A, while removing the nonlinear effect in Fig. 9(b). In the special linear case, the response curve of Module A is a horizontal line at -20 dB, which implies that the iteration between Modules A and C is unnecessary because v_{2z} is always at -20 dB. However, in the non-linear case, the response curve of Module A is eventually saturated at around -16 dB and cannot reach -20 dB. Hence, 4 dB represents an inherent performance degeneration due to non-linearity. To improve our understanding, we further provide other examples of the response curves of Module A for different nonlinear functions $\mathbb{Q}(\cdot)$ such as the quantizers in Fig. 10. Compared with that for the one- or two-bit quantizer, the response curve for the PR becomes saturated at a low MSE. However, we find that if $1/v_{1z}$ is close to zero (i.e., poor initialization), v_{2z} for PR is worse than that in the one-bit case. This observation implies that good initialization for PR is more critical than that for the one-bit case when employing the GEC-SR algorithm.

Next, we consider the evolution trajectory for the decentralized case. Comparing GEC-SR with deGEC-SR, we find that Modules A and B are identical and that the only difference is noted in Module C. Thus, the response curves for Modules A and B in deGEC-SR are identical to those in GEC-SR. Recall that Module C solves the linear inverse problem. Therefore, one can expect that the response curve of Module C highly depends on the ratio between the dimensions of measurements M_l and unknown signal N, i.e., $\gamma_l = M_l/N$. We show examples of the response curves with fixed $N = 1,000$, $M = 4,000$ but varying $M_l = \{400, 2000, 4000\}$ in Fig. 11, i.e., the cases with $L = 10, 2, 1$, respectively. For the cases *without* fusion, we find that the whole response curve of Module C move close to the left as the measurement dimension decreases. The smaller the measurement dimensions are, the closer to the left the response curve is. As the response curve of Modules A is fixed, the response curves of Modules C shall intersect at high MSE values as it is close to the left. This fact implies poor reconstruction performance when the dimensions of measurements are small. Therefore, the reconstruction performance of each cluster is degenerated with the number of clusters. However, if the fusion center is applied, we find that the response curve of Module C
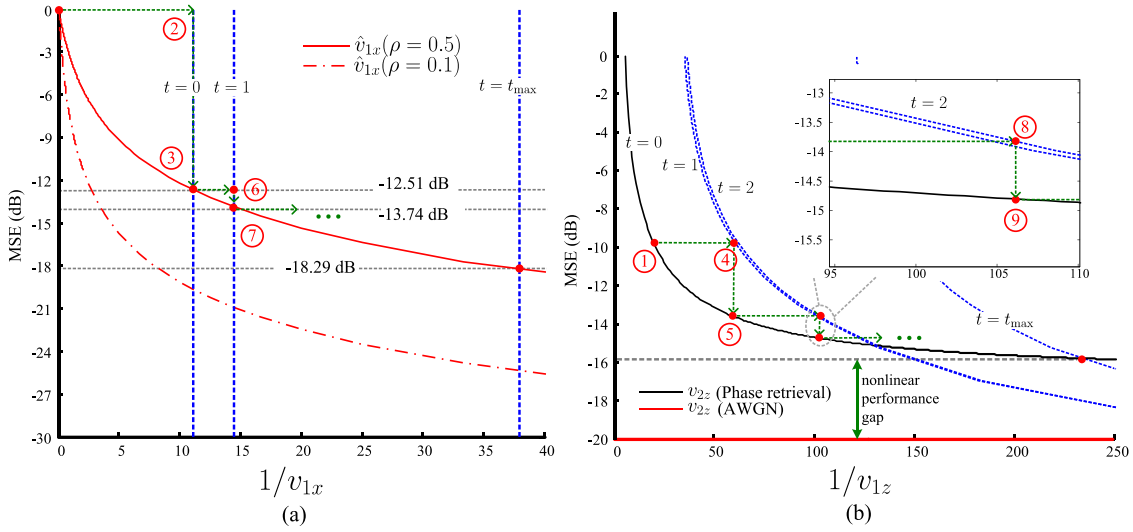
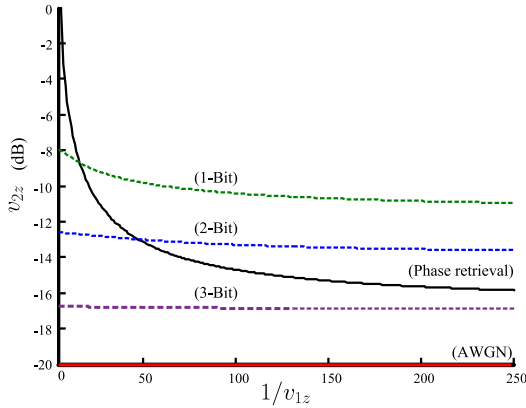Fig. 9. Examples of the evolution trajectory for GEC-SR.

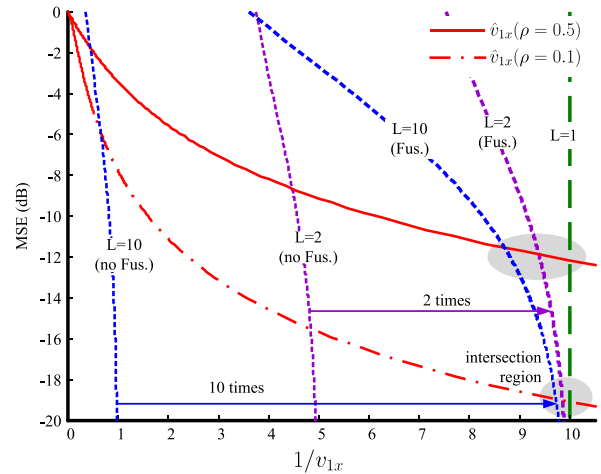Fig. 10. Examples of response curves of Module A for different nonlinear functions $\mathtt{Q}(\cdot)$.

moves to the right. Specifically, because of the combining formulation (38), the response curve with the combination uniformly shifts to the right by L times. From the intersections of the response curves of Modules B and C, we can determine the degeneration due to decentralization. Given the shapes of the response curves of Module C (i.e., their shapes are similar at the intersection region), we can realize that the degeneration is insignificant as long as signal $\mathbf{x}$ becomes sparse. For example, we find that the degeneration for the case with $\rho = 0.1$ is more negligible than that for the case with $\rho = 0.5$.

E. Performance and Time Complexity

SE can be used to quickly obtain the performance predictions for deGEC-SR with different clusters because of its accuracy. The predicted performance provide useful information to determine the trade-off between performance and time complexity. To understand this trade-off, we compare the performance and time complexity with a fixed unknown signal dimension $N = 1,000$, while varying measurement dimensions
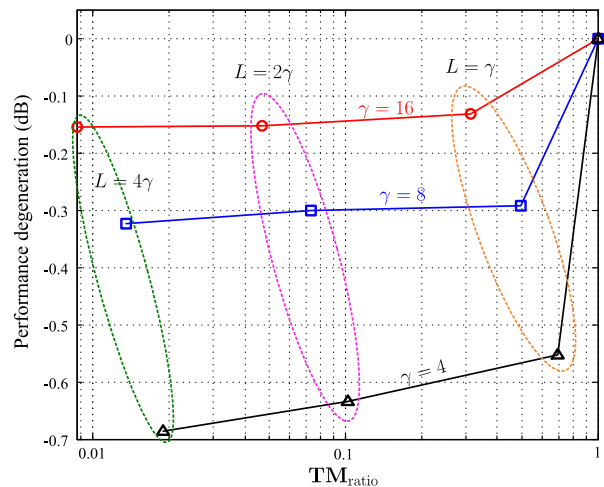

Fig. 11. Examples of the response curves of Module C with different number of decentralization clusters.

Fig. 12. $\mathrm{MSE_{ratio}}$ versus $\mathrm{TM_{ratio}}$ for fixed unknown signal dimension $N = 1,000$ with varying measurement dimensions $M = \{4000, 8000, 16000\}$ and clusters L. The ovals represent the same number of clusters $L = \{\gamma, 2\gamma, 4\gamma\}$.

$M = \{4000, 8000, 16000\}$ and clusters L. In particular, we consider $\gamma = M/N = \{4, 8, 16\}$ and $L = \gamma, 2\gamma, 4\gamma$. Here, performance is predicted using SE, and time complexity is determined through computer simulations.

To clearly present the comparisons, we define the time complexity reduction ratio as

$$TM_{ratio} = \frac{Time_{deGEC-SR}}{Time_{GEC-SR}}. \tag{47}$$

Fig. 12 compares the performance and time complexity in terms of performance degeneration and TM_{ratio} for $\rho = 0.5$ and SNR = 20 dB. We find that for most of the cases, deGEC-SR leads to minor performance degeneration, while reducing time complexity by tens to hundreds of times. For the case with sufficiently large measurement dimensions, that is, sampling rate γ is large, performance degeneration caused by decentralization can be ignored.

V. CONCLUSION

We developed a novel decentralized algorithm called deGEC-SR by leveraging the core framework of GEC-SR. In most cases of interest, performance degeneration due to decentralization can be ignored. deGEC-SR offers excellent performance as GEC-SR and runs tens to hundreds of times faster than GEC-SR. We derived the theoretical SE for deGEC-SR and demonstrated its accuracy by using numerical results. With our analysis, we highlight the following facts about our algorithm. i) A good initialization is more critical for signal recovery in the phase-less stage than other nonlinearity when employing the GEC-SR algorithm. ii) Given the shaping of the response curves of Module C (i.e., their shapes are similar at the intersection region), the degeneration is ignored as long as the signal becomes sparse.

APPENDIX A
EXPLICIT EXPRESSION OF (43A)

In this appendix, we show the explicit expression of (43a) for the PR model. For brevity, we omit the cluster index $[l]$, iteration index (t), and prior/extrinsic index $_{1/2}$ throughout the following derivation.

From Table I, we can obtain the posterior variance $\mathtt{Var}_{P_{out}(y|z)}\{z|\mu_z, v_z\}$, which is given by

$$\mathtt{Var}_{P_{out}(y|z)}\{z|\mu_z, v_z\}$$

$$= \underbrace{\frac{v_z^2 y^2}{(v_z + \sigma_w^2)^2}}_{(a)} + \underbrace{\frac{v_z \sigma_w^2}{v_z + \sigma_w^2}}_{(b)} - \underbrace{\frac{v_z^2 y^2}{(v_z + \sigma_w^2)^2} \left(\frac{I_1\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2}\right)}{I_0\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2}\right)} \right)^2}_{(c)}. \tag{48}$$

Notice that y and μ_z are random variables. To take the expectation on (48) w.r.t. y and μ_z, we need their joint distribution $P(y, \mu_z)$.

In GEC-SR algorithm, z is modeled by its priors (μ_z, v_z). The orthogonality principle of the LMMSE implies that $E\{\mu_z^*(\mu_z - z)\} = 0$. Combining this expression and $E\{|\mu_z - z|^2\} = v_z$, we

can obtain the covariance of (z, μ_z) as [19]

$$\begin{bmatrix} P_z & P_z - v_z \\ P_z - v_z & P_z - v_z \end{bmatrix}. \tag{49}$$

With this covariance matrix, then the conditional distribution of z given μ_z is $\mathcal{N}_{\mathbb{C}}(z; \mu_z, v_z)$, which agrees with our desired formula. In particular, the bivariate Gaussian distribution can be expressed by

$$P(z, \mu_z) = P(z|\mu_z)P(\mu_z)$$
$$= \mathcal{N}_{\mathbb{C}}(z; \mu_z, v_z)\mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z). \tag{50}$$

In the PR model, y is related to z via $P_{out}(y|z)$. With the Markov process, we obtain

$$P(y, \mu_z) = \int P(y, z, \mu_z)dz = \int P(y|z, \mu_z)P(z, \mu_z)dz$$

$$\overset{(a)}{=} \int \underbrace{1_{y \geq 0} y \int_0^{2\pi} \mathcal{N}_{\mathbb{C}}(ye^{j\theta}; z, \sigma_w^2)d\theta}_{=P_{out}(y|z)}$$

$$\underbrace{\mathcal{N}_{\mathbb{C}}(z; \mu_z, v_z)}_{=P(z|\mu_z)} \underbrace{\mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z)}_{=P(\mu_z)} dz$$

$$\overset{(b)}{=} 1_{y \geq 0} y \int_0^{2\pi} \left(\int \mathcal{N}_{\mathbb{C}}\left(z; \frac{v_z ye^{j\theta} + \sigma_w^2 \mu_z}{v_z + \sigma_w^2}, \frac{v_z \sigma_w^2}{v_z + \sigma_w^2} \right) dz \right)$$

$$\times \mathcal{N}_{\mathbb{C}}(ye^{j\theta}; \mu_z, v_z + \sigma_w^2)d\theta \mathcal{N}(\mu_z; 0, P_z - v_z)$$

$$= 1_{y \geq 0} y \int_0^{2\pi} \mathcal{N}_{\mathbb{C}}(ye^{j\theta}; \mu_z, v_z + \sigma_w^2)d\theta \mathcal{N}(\mu_z; 0, P_z - v_z)$$

$$\overset{(c)}{=} \frac{2y}{v_z + \sigma_w^2} e^{-\frac{y^2 + |\mu_z|^2}{v_z + \sigma_w^2}} I_0\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2} \right)$$

$$\times \mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z)1_{y \geq 0}, \tag{51}$$

where (a) follows (65), (b) follows the Gaussian production lemma, and (c) follows (66) to calculate the integral over θ.

Now, we are able to compute the expectation of (48) w.r.t. (y, μ_z). As (48) consists of three terms, we address each separately. First, we consider the second term. Taking the expectation of (48)(b) with $P(y, \mu_z)$ given in (51),

$$\int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{v_z \sigma_w^2}{v_z + \sigma_w^2} \right) \frac{2y}{v_z + \sigma_w^2} e^{-\frac{y^2 + |\mu_z|^2}{v_z + \sigma_w^2}} I_0\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2} \right)$$

$$\times \mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z) \, dy \, d\mu_z$$

$$= \frac{2v_z \sigma_w^2}{(v_z + \sigma_w^2)^2} \int_{-\infty}^{\infty} \left[\int_0^{\infty} ye^{-\frac{y^2}{v_z + \sigma_w^2}} I_0\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2} \right) dy \right]$$

$$\times e^{-\frac{|\mu_z|^2}{v_z + \sigma_w^2}} \mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z)d\mu_z$$

$$\overset{(a)}{=} \frac{2v_z \sigma_w^2}{(v_z + \sigma_w^2)^2} \int_{-\infty}^{\infty} \left[\frac{v_z + \sigma_w^2}{2} e^{\frac{|\mu_z|^2}{v_z + \sigma_w^2}} \right]$$

$$\times e^{-\frac{|\mu_z|^2}{v_z + \sigma_w^2}} \mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z) \, d\mu_z$$

$$= \frac{v_z \sigma_w^2}{(v_z + \sigma_w^2)} \int_{-\infty}^{\infty} \mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z)d\mu_z = \frac{v_z \sigma_w^2}{v_z + \sigma_w^2}, \tag{52}$$

where (a) follows the expression in [45, Eq. 10.43.23] to calculate the integral over y.

Next, we consider the first term. Taking the expectation of (48)(a) with $\mathsf{P}(y, \mu_z)$ given in (51), we obtain

$$
\frac{2v_z^2}{(v_z + \sigma_w^2)^3} \int_0^\infty \int_{-\infty}^\infty y^3 e^{-\frac{y^2 + |\mu_z|^2}{v_z + \sigma_w^2}} I_0\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2}\right)
$$
$$
\times \mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z) \, d\mu_z \, dy.
$$

To continue, we use the rectangular-to-polar transformation $\mu_z = u_z e^{j\theta_z}$ (65) and obtain

$$
\frac{2v_z^2}{(v_z + \sigma_w^2)^3} \int_0^\infty \int_0^\infty \int_0^{2\pi} y^3 u_z e^{-\frac{y^2 + u_z^2}{v_z + \sigma_w^2}} I_0\left(\frac{2yu_z}{v_z + \sigma_w^2}\right)
$$
$$
\times \mathcal{N}_{\mathbb{C}}(u_z e^{j\theta_z}; 0, P_z - v_z) \, d\theta_z \, du_z dy
$$
$$
\overset{(a)}{=} \frac{2v_z^2}{(v_z + \sigma_w^2)^3} \int_0^\infty \int_0^\infty y^3 u_z e^{-\frac{y^2 + u_z^2}{v_z + \sigma_w^2}} I_0\left(\frac{2yu_z}{v_z + \sigma_w^2}\right)
$$
$$
\times \frac{2\pi}{\pi(P_z - v_z)} e^{-\frac{u_z^2}{P_z - v_z}} \, du_z \, dy
$$
$$
= \frac{4v_z^2}{(v_z + \sigma_w^2)^3 (P_z - v_z)} \int_0^\infty y^3 e^{-\frac{y^2}{v_z + \sigma_w^2}}
$$
$$
\times \left[\int_0^\infty u_z e^{-\frac{u_z^2}{P_z - v_z} - \frac{u_z^2}{v_z + \sigma_w^2}} I_0\left(\frac{2yu_z}{v_z + \sigma_w^2}\right) du_z\right] dy
$$
$$
\overset{(b)}{=} \frac{4v_z^2}{(v_z + \sigma_w^2)^3 (P_z - v_z)} \int_0^\infty y^3 e^{-\frac{y^2}{v_z + \sigma_w^2}}
$$
$$
\times \left[\frac{(P_z - v_z)(v_z + \sigma_w^2)}{2(P_z + \sigma_w^2)} e^{\frac{(P_z - v_z)y^2}{(v_z + \sigma_w^2)(P_z + \sigma_w^2)}}\right] dy
$$
$$
\overset{(c)}{=} \frac{v_z^2}{(v_z + \sigma_w^2)^2 (P_z + \sigma_w^2)} \int_0^\infty u e^{-\frac{1}{P_z + \sigma_w^2} u} \, du
$$
$$
= \frac{v_z^2(P_z + \sigma_w^2)}{(v_z + \sigma_w^2)^2}. \tag{53}
$$

where (a) follows (67), (b) follows the expression in (63) to calculate the integral over u_z, and for (c), we change the variable by letting $u = y^2$.

Next, we consider the third term. Putting (48)(c) and $\mathsf{P}(y, \mu_z)$ together, we obtain

$$
-\frac{2v_z^2}{(v_z + \sigma_w^2)^3} \int_0^\infty \int_{-\infty}^\infty y^3 \left(\frac{I_1\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2}\right)}{I_0\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2}\right)}\right)^2
$$
$$
\times e^{-\frac{y^2 + |\mu_z|^2}{v_z + \sigma_w^2}} I_0\left(\frac{2y|\mu_z|}{v_z + \sigma_w^2}\right) \mathcal{N}_{\mathbb{C}}(\mu_z; 0, P_z - v_z) d\mu_z dy.
$$

Applying the rectangular-to-polar transformation $\mu_z = u_z e^{j\theta_z}$, the above equation becomes

$$
-\frac{2v_z^2}{(v_z + \sigma_w^2)^3} \int_0^\infty \int_0^\infty \int_0^{2\pi} y^3 u_z \left(\frac{I_1\left(\frac{2yu_z}{v_z + \sigma_w^2}\right)}{I_0\left(\frac{2yu_z}{v_z + \sigma_w^2}\right)}\right)^2
$$
$$
\times e^{-\frac{y^2 + u_z^2}{v_z + \sigma_w^2}} I_0\left(\frac{2y u_z}{v_z + \sigma_w^2}\right) \mathcal{N}_{\mathbb{C}}(re^{j\theta_z}; 0, P_z - v_z) d\theta_z \, du_z \, dy,
$$

$$
= -\frac{4v_z^2}{(v_z + \sigma_w^2)^3 (P_z - v_z)} \int_0^\infty \int_0^\infty y^3 u_z \left(\frac{I_1\left(\frac{2yu_z}{v_z + \sigma_w^2}\right)}{I_0\left(\frac{2yu_z}{v_z + \sigma_w^2}\right)}\right)^2
$$
$$
\times e^{-\frac{y^2 + u_z^2}{v_z + \sigma_w^2} - \frac{u_z^2}{P_z - v_z}} I_0\left(\frac{2yu_z}{v_z + \sigma_w^2}\right) du_z dy, \tag{54}
$$

where the equality follows (67). Defining

$$
F(a, b, c) = \int_0^\infty \int_0^\infty y^3 u \left(\frac{I_1\left(\frac{2yu}{a+c}\right)}{I_0\left(\frac{2yu}{a+c}\right)}\right)^2
$$
$$
\times e^{-\frac{y^2 + u^2}{a+c} - \frac{u^2}{b-a}} I_0\left(\frac{2yu}{a+c}\right) du dy \tag{55}
$$

and collecting (52)–(54), we finally obtain

$$
\hat{v}_z = \frac{v_z \sigma_w^2}{v_z + \sigma_w^2} + \frac{v_z^2(P_z + \sigma_w^2)}{(v_z + \sigma_w^2)^2}
$$
$$
- \frac{4v_z^2}{(v_z + \sigma_w^2)^3 (P_z - v_z)} F\left(v_z, P_z, \sigma_w^2\right). \tag{56}
$$

APPENDIX B

In this appendix, we show that the explicit expression of (43f), as x follows the Gaussian-Bernoulli distribution. For brevity, we omit the cluster index $[l]$, iteration index (t), and prior/extrinsic index $_{1/2}$ throughout the following derivation.

From line 3 of Table I, we can obtain the posterior variance $\mathrm{Var}_{\mathsf{P}(x)}\{x|\mu_x, v_x\}$, which is given by

$$
\underbrace{\mathcal{Z}_x^{-1}\left(\left|\frac{\mu_x \rho^{-1}}{v_x + \rho^{-1}}\right|^2 + \frac{v_x \rho^{-1}}{v_x + \rho^{-1}}\right)}_{(a)} - \underbrace{\mathcal{Z}_x^{-2}\left|\frac{\mu_x \rho^{-1}}{v_x + \rho^{-1}}\right|^2}_{(b)}. \tag{57}
$$

Recalling from (14), the calculation of $\mathrm{Var}_{\mathsf{P}(x)}\{x|\mu_x, v_x\}$ involves $\mathcal{N}_{\mathbb{C}}(x; \mu_x, v_x)$ and the true prior $\mathsf{P}(x)$ (8). Therefore, the joint distribution of interest is given by

$$
\mathsf{P}(x, \mu_x) = \mathsf{P}(\mu_x|x)\mathsf{P}(x)
$$
$$
= \mathcal{N}_{\mathbb{C}}(x; \mu_x, v_x)\left[(1-\rho)\delta(x) + \rho\mathcal{N}_{\mathbb{C}}(x; 0, \rho^{-1})\right]. \tag{58}
$$

Integrating out x, we obtain the marginal distribution

$$
\mathsf{P}(\mu_x) = (1-\rho)\mathcal{N}_{\mathbb{C}}(\mu_x; 0, v_x) + \rho\mathcal{N}_{\mathbb{C}}(\mu_x; 0, v_x + \rho^{-1}), \tag{59}
$$

where the second term can be easily obtained by the Gaussian production lemma.

We address the two terms of (57) separately. First, we consider the first term. Taking the expectation of (57)(a) with $\mathsf{P}(\mu_x)$ given in (59),

$$
\int \mathcal{Z}_x^{-1}\left(\left|\frac{\mu_x \rho^{-1}}{v_x + \rho^{-1}}\right|^2 + \frac{v_x \rho^{-1}}{v_x + \rho^{-1}}\right) \mathsf{P}(\mu_x) d\mu_x
$$
$$
= \frac{v_x \rho^{-1}}{v_x + \rho^{-1}} \rho \int \mathcal{N}_{\mathbb{C}}(\mu_x; 0, v_x + \rho^{-1}) d\mu_x
$$
$$
+ \rho \int \left(\frac{\rho^{-1}}{v_x + \rho^{-1}}\right)^2 |\mu_x|^2 \mathcal{N}_{\mathbb{C}}(\mu_x; 0, v_x + \rho^{-1}) d\mu_x = 1, \tag{60}
$$

where the first equality can be easy obtained by noticing that the denominator of $\mathcal{Z}_x^{-1}$ is identical to $\mathsf{P}(\mu_x)$. Next, taking the expectation of (57)(b) with $\mathsf{P}(\mu_x)$ given in (59),

$$
\frac{1}{(v_x + \rho^{-1})^2}
$$

$$
\times \int_{-\infty}^{\infty} \frac{\left| \mu_x \mathcal{N}_{\mathbb{C}}(\mu_x; 0, v_x + \rho^{-1}) \right|^2}{(1 - \rho)\mathcal{N}_{\mathbb{C}}(\mu_x; 0, v_x) + \rho \mathcal{N}_{\mathbb{C}}(\mu_x; 0, v_x + \rho^{-1})} d\mu_x
$$

$$
= \frac{1}{v_x + \rho^{-1}} \int_0^{\infty} \frac{r e^{-r}}{\rho + (1 - \rho)(1 + (\rho v_x)^{-1}) e^{-\frac{r}{\rho v_x}}} dr,
$$
(61)

where the equality can be easily obtained by using the rectangular-to-polar transformation and applying (67). Combining (60) and (61), we obtain the expression given in (45).

APPENDIX C

In this appendix, we provide useful functions and integrals.
Bessel functions: The modified Bessel function of the first kind is defined as

$$
I_n(t) = \frac{1}{\pi} \int_0^{\pi} e^{t \cos \theta} \cos(n\theta) d\theta,
$$
(62)

where n is an integral. Using it, we have [45, Eq. 10.43.23]

$$
\int_0^{\infty} t^{n+1} I_n(b\,t) e^{-p^2 t^2} dt = \frac{b^n}{(2p^2)^{n+1}} e^{\frac{b^2}{4p^2}}.
$$
(63)

Rectangular-to-polar transformation: Considering the conditionally distribute $\mathsf{P}_{\text{out}}(u|z) = \mathcal{N}_{\mathbb{C}}(u; z, v)$, we transform $u = y e^{j\theta}$ from rectangular to polar coordinates to obtain [20, (22)]

$$
\mathsf{P}_{\text{out}}(y, \theta|z) = 1_{y \geq 0} 1_{\theta \in [0, 2\pi)} \mathcal{N}_{\mathbb{C}}(y e^{j\theta}; z, v)\, y.
$$
(64)

Integrating the unobserved phase θ, we can obtain [20, (23)]

$$
\mathsf{P}_{\text{out}}(y|z) = 1_{y \geq 0}\, y \int_0^{2\pi} \mathcal{N}_{\mathbb{C}}(y e^{j\theta}; z, v) d\theta
$$
(65)

$$
= \frac{2y}{\sigma_w^2} e^{-\frac{y^2 + |z|^2}{\sigma_w^2}} I_0\left(\frac{2y|z|}{\sigma_w^2} \right) 1_{y \geq 0}.
$$
(66)

In addition, we find the following integral to be useful

$$
\int_0^{2\pi} \mathcal{N}_{\mathbb{C}}(y e^{j\theta}; 0, v)\, d\theta = \int_0^{2\pi} \frac{1}{\pi v} e^{-\frac{|y e^{j\theta}|^2}{v}} d\theta = \frac{2}{v} e^{-\frac{y^2}{v}}.
$$
(67)

REFERENCES

[1] S. Wang, L. Zhang, and X. Jing, "Phase retrieval motivated nonlinear MIMO communication with magnitude measurements," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5452–5466, Aug. 2017.
[2] R. Combes and S. Yang, "An approximate ML detector for MIMO channels corrupted by phase noise," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1176–1189, Mar. 2018.
[3] X. Yang, S. Jin, and C. Wen, "Symbol detection of phase noise-impaired massive MIMO using approximate Bayesian inference," *IEEE Signal Process. Lett.*, vol. 26, no. 4, pp. 607–611, Apr. 2019.
[4] L. Bian, J. Suo, G. Zheng, K. Guo, F. Chen, and Q Dai, "Fourier ptychographic reconstruction using Wirtinger flow optimization," *Opt. Express*, vol. 23, no. 4, pp. 4856–4866, 2015.
[5] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Amer.*, vol. 7, no. 3, pp. 394–411, Mar. 1990.

[6] DL Misell, "A method for the solution of the phase problem in electron microscopy," *J. Phys. D: Appl. Phys.*, vol. 6, no. 1, pp. L6–L9, 1973.
[7] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.
[8] P. M. Pardalos and S. A. Vavasis, "Quadratic programming with one negative eigenvalue is NP-hard," *J. Global Optim.*, vol. 1, no. 1, pp. 15–22, Feb. 1991.
[9] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.*, vol. 3, no. 1, pp. 27–29, 1978.
[10] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237–246, 1972.
[11] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commu. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, Nov. 2013.
[12] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Jan. 2015.
[13] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2675–2689, Apr. 2018.
[14] C. T. O. Dhifallah and Y. M. Lu, "Phase retrieval via linear programming: Fundamental limits and algorithmic improvements," in *Proc. 55th Annual Allerton Conf. Commun., Control, Comput*, Monticello, USA, Oct. 2017, pp. 1071–1077.
[15] Z. Yuan and H. Wang, "Phase retrieval via reweighted wirtinger flow," *Appl. Opt.*, vol. 56, no. 9, pp. 2418–2427, 2017.
[16] P. J. P. Netrapalli and S. Sanghavi, "Phase retrieval using alternating minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4814–4826, Sep. 2015.
[17] R. Ghods, A. S. Lan, T. Goldstein, and C. Studer, "Linear spectral estimators and an application to phase retrieval," in *Proc. 35th Int. Conf. Mach. Learn.*, vol. 80, Stockholm, Sweden, Jul. 2018, pp. 1734–1743.
[18] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Nat. Acad. Sci.*, vol. 106, pp. 18 914–18 919, Nov. 2009.
[19] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Jul./Aug. 2011, pp. 2168–2172.
[20] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1043–1055, Feb. 2015.
[21] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," *Nat. Acad. Sci.*, vol. 116, pp. 5541–5460, Dec. 2019.
[22] J. Ma, J. Xu, and A. Maleki, "Optimization-based AMP for phase retrieval: The impact of initialization and $\ell 2$-regularization," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3600–3629, Jun. 2019.
[23] J. Ma, X. Yuan, and L. Ping, "Turbo compressed sensing with partial DFT sensing matrix," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 158–161, Feb. 2015.
[24] T. Liu, C. K. Wen, S. Jin, and X. You, "Generalized turbo signal recovery for nonlinear measurements and orthogonal sensing matrices," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 2883–2887.
[25] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, no. 14, pp. 2020–2033, 2017.
[26] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, 25–30 Jun. 2017, pp. 1588–1592.
[27] X. Meng, S. Wu, and J. Zhu, "A unified Bayesian inference framework for generalized linear models," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 398–402, Mar. 2018.
[28] Z. Xue, X. Yuan, J. Ma, and Y. Ma, "TARM: A turbo-type algorithm for affine rank minimization," *IEEE Trans. Signal Process.*, vol. 67, no. 22, pp. 5730–5745, Nov. 2019.
[29] X. Meng and J. Zhu, "Bilinear adaptive generalized vector approximate message passing," *IEEE Access*, vol. 7, pp. 4807–4815, 2019.
[30] A. Fletcher, M. Sahraee-Ardakan, S. Rangan, and P. Schniter, "Expectation consistent approximate inference: Generalizations and convergence," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 190–194.

[31] P. Schniter, S. Rangan, and A. Fletcher, "Vector approximate message passing for the generalized linear model," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 1525–1529.

[32] H. He, C.-K. Wen, and S. Jin, "Generalized expectation consistent signal recovery for nonlinear measurements," in *Proc. IEEE Int. Symp. Inf. Theory*, Germany: Aachen, Jun. 2017, pp. 2333–2337.

[33] C. A. Metzler, M. K. Sharma, S. Nagesh, R. G. Baraniuk, O. Cossairt, and A. Veeraraghavan, "Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Stanford, CA, USA, May 2017, pp. 1–16.

[34] M. Sharma, C. A. Metzler, S. Nagesh, O. Cossairt, R. G. Baraniuk, and A. Veeraraghavan, "Inverse scattering via transmission matrices: Broadband illumination and fast phase retrieval algorithms," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 95–108, 2020.

[35] B. Rajaei, S. Gigan, F. Krzakala, and L. Daudet, "Robust phase retrieval with the swept approximate message passing (prSAMP) algorithm," *Image Process. Line*, vol. 7, pp. 43–55, Jan. 2017.

[36] Z. Zhao, S. Lu, M. Hong, and D. P. Palomar, "Distributed optimization for generalized phase retrieval over networks," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, USA, Oct. 2018, pp. 48–52.

[37] B. Rajaei, S. Gigan, F. Krzakala, and L. Daudet, "Fast phase retrieval for high dimensions: A block-based approach," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1179–1182, Sep. 2016.

[38] J. Pearl, "Reverend bayes on inference engines: A distributed hierarchical approach," in *Proc. Second AAAI Conf. Artif. Intell.*, Pittsburgh, USA, Aug. 1982, pp. 133–136.

[39] A. Dremeau and F. Krzakala, "Phase recovery from a Bayesian point of view: the variational approach," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brisbane, Australia, Apr. 2015, pp. 3661–3665.

[40] A. Manoel, F. Krzakala, E. W. Tramel, and L. Zdeborová, "Streaming Bayesian inference: Theoretical limits and mini-batch approximate message-passing," in *Proc. 55th Annual Allerton Conf. Commun., Control, Comput.*, Monticello, USA, Oct. 2017, pp. 1048–1055.

[41] J. Zhu, A. Beirami, and D. Baron, "Performance trade-offs in multiprocessor approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 680–684.

[42] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *J. Stat. Mech-Theory E*, vol. P08009, pp. 1–57, Aug. 2012.

[43] S. ten Brink, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, no. 10, pp. 806–808, May 1999.

[44] L. Liu, C. Liang, J. Ma, and L. Ping, "Capacity optimality of AMP in coded systems," preprint, 2019. [Online]. Available: https://arxiv.org/abs/1901.09559

[45] F. W. J. Olver *et al.*, "NIST Digital Library of Mathematical Functions," Release 1.0.21 of 2018-12-15. [Online]. Available: http://dlmf.nist.gov/

Chao-Kai Wen (Member, IEEE) received the Ph.D. degree from the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2004. He was with Industrial Technology Research Institute, Hsinchu, Taiwan and MediaTek Inc., Hsinchu, Taiwan, from 2004 to 2009, where he was engaged in broadband digital transceiver design. In 2009, he joined the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, where he is currently Professor. His research interests include optimization in wireless multimedia networks.

Shang-Ho (Lawrence) Tsai (Senior Member, IEEE) was born in Kaohsiung, Taiwan. He received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in August 2005. From June 1999 to July 2002, he was with Silicon Integrated Systems Corporation (SiS), where he engaged in VLSI design for DMT-ADSL systems. From September 2005 to January 2007, he was with the MediaTek Inc. (MTK) participating in the VLSI design for MIMO-OFDM systems and standard specifications for IEEE 802.11n. In February 2007, he joined the Department of Electrical Engineering, National Chiao Tung University where he is currently a Professor. His research interests include signal processing for communications, statistical signal processing, and signal processing for VLSI designs. Dr. Tsai is in the editorial board of IEEE SIGPORT since 2018. He was awarded a government scholarship for overseas study from the Ministry of Education, Taiwan, in 2002–2005. He was a Visiting Fellow with the Department of Electrical Engineering, Princeton University in June 2013–December 2013.
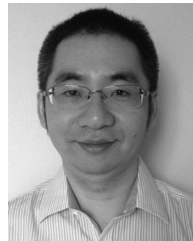
Shi Jin (Senior Member, IEEE) received the B.S. degree in communications engineering from the Guilin University of Electronic Technology, Guilin, China, in 1996, the M.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2003, and the Ph.D. degree in information and communications engineering from the Southeast University, Nanjing, China, in 2007. From June 2007 to October 2009, he was a Research Fellow with the Adastral Park Research Campus, University College London, London, U.K. He is currently with the faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include space time wireless communications, random matrix theory, and information theory. He serves as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE COMMUNICATIONS LETTERS, and *IET Communications*. He and his co-authors have been awarded the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory and a 2010 Young Author Best Paper Award by the IEEE Signal Processing Society.

Chang-Jen Wang received the B.S. degree from the Department of Electrical Engineering, and the M.S. degree from the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, in 2013 and 2015, respectively. He is currently working toward the Ph.D. degree with the Institute of Electrical and Control Engineerin, National Chiao Tung University, Hsinchu, Taiwan. His current research interests include signal processing and wireless multimedia networks.