

Compressive Phase Retrieval via Generalized Approximate Message Passing

Philip Schniter, *Fellow, IEEE*, and Sundeep Rangan, *Senior Member, IEEE*

Abstract—In phase retrieval, the goal is to recover a signal $\mathbf{x} \in \mathbb{C}^N$ from the magnitudes of linear measurements $\mathbf{Ax} \in \mathbb{C}^M$. While recent theory has established that $M \approx 4N$ intensity measurements are necessary and sufficient to recover generic \mathbf{x} , there is great interest in reducing the number of measurements through the exploitation of sparse \mathbf{x} , which is known as compressive phase retrieval. In this work, we detail a novel, probabilistic approach to compressive phase retrieval based on the generalized approximate message passing (GAMP) algorithm. We then present a numerical study of the proposed PR-GAMP algorithm, demonstrating its excellent phase-transition behavior, robustness to noise, and runtime. Our experiments suggest that approximately $M \geq 2K \log_2(N/K)$ intensity measurements suffice to recover K -sparse Bernoulli-Gaussian signals for \mathbf{A} with i.i.d Gaussian entries and $K \ll N$. Meanwhile, when recovering a 6678-sparse 65536-pixel grayscale image from 32768 randomly masked and blurred Fourier intensity measurements at 30 dB measurement SNR, PR-GAMP achieved an output SNR of no less than 28 dB in all of 100 random trials, with a median runtime of only 7.3 seconds. Compared to the recently proposed CPRL, sparse-Fienup, and GESPAR algorithms, our experiments suggest that PR-GAMP has a superior phase transition and orders-of-magnitude faster runtimes as the sparsity and problem dimensions increase.

Index Terms—Belief propagation, compressed sensing, estimation, phase retrieval.

I. INTRODUCTION

A. Phase Retrieval

In phase retrieval, the goal is to recover a signal $\mathbf{x} \in \mathbb{C}^N$ from the magnitudes $y_m = |u_m|$ of possibly noisy linear measurements $\mathbf{u} = [u_1, \dots, u_M]^T = \mathbf{Ax} + \mathbf{w} \in \mathbb{C}^M$. This problem is motivated by the fact that it is often easier to build detectors (e.g., photographic plates or CCDs) that measure intensity rather than phase [3], [4]. Imaging applications of phase retrieval include X-ray diffraction imaging [5], X-ray crystallography [6], [7], array imaging [8], optics [9], speckle imaging

in astronomy [10], and microscopy [11]. Non-imaging applications include acoustics [12], interferometry [13], and quantum mechanics [14].

To reconstruct $\mathbf{x} \in \mathbb{C}^N$ (up to a global phase uncertainty), it has been recently established that $M \geq 4N - o(N)$ intensity measurements are necessary [15] and $M \geq 4N - 4$ are sufficient [16] through appropriate design of the linear transform \mathbf{A} . Meanwhile, to reconstruct $\mathbf{x} \in \mathbb{R}^N$ (up to a global sign uncertainty), it has been shown that $M \geq 2N - 1$ measurements are both necessary and sufficient [12]. However, there exist applications where far fewer measurements are available, such as sub-wavelength imaging [17], [18], Bragg sampling from periodic crystalline structures [19], and waveguide-based photonic devices [20]. To facilitate these *compressive* phase retrieval tasks, it has been proposed to exploit *sparsity*¹ in \mathbf{x} . In fact, very recent theory confirms the potential of this approach: to reconstruct K -sparse N -length \mathbf{x} using a generic (e.g., i.i.d Gaussian) \mathbf{A} , only $M \geq 4K - 2$ intensity measurements suffice in the complex case and $M \geq 2K$ suffice in the real case (where $M \geq 2K$ is also necessary) when $K < N$ [21]. While these bounds are extremely encouraging, achieving them with a practical algorithm remains elusive.

To our knowledge, the first algorithm for compressive phase retrieval was proposed by Moravec, Romberg, and Baraniuk in [22] and worked by incorporating an ℓ_1 -norm *constraint* into a traditional Fienup-style [3] iterative algorithm. However, this approach requires that the ℓ_1 norm of the true signal is known, which is rarely the case in practice. Recently, a more practical sparse-Fienup algorithm was proposed by Mukherjee and Seelamantula [23], which requires knowledge of only the signal sparsity K but is applicable only to measurement matrices \mathbf{A} for which $\mathbf{A}^H \mathbf{A} = \mathbf{I}$. Although this algorithm guarantees that the residual error $\|\mathbf{y} - |\mathbf{A}\hat{\mathbf{x}}(t)|\|_2^2$ is non-increasing over the iterations t , it succumbs to local minima and, as we show in Section IV-D, is competitive only in the highly sparse regime.

To circumvent the local minima problem, Ohlsson, Yang, Dong, and Sastry proposed the *convex relaxation* known as Compressive Phase Retrieval via Lifting (CPRL) [24], which adds ℓ_1 regularization to the well-known PhaseLift algorithm [8], [25]. Both CPRL and PhaseLift “lift” the unknown vector $\mathbf{x} \in \mathbb{C}^N$ into the space of $N \times N$ rank-one matrices and solve a semidefinite program in the lifted space, requiring $O(N^3)$ complexity, which is impractical for practical image sizes N . Subsequent theoretical analysis [21] revealed that, while $M \gtrsim O(K^2 \log N)$ intensity measurements suffice for

¹ \mathbf{x} may represent the sparse transform coefficients of a non-sparse signal-of-interest $\mathbf{s} = \mathbf{\Psi}\mathbf{x}$ in a sparsifying basis (or frame) $\mathbf{\Psi}$, in which case the intensity measurements would be $\mathbf{y} = |\mathbf{\Phi}\mathbf{s} + \mathbf{w}|$ and $\mathbf{A} \triangleq \mathbf{\Phi}\mathbf{\Psi}$.

Manuscript received June 26, 2014; revised October 17, 2014; accepted December 08, 2014. Date of publication December 25, 2014; date of current version January 26, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ami Wiesel. This work was supported in part by NSF grant CCF-1018368, by DARPA/ONR grant N66001-10-1-4090, and by an allocation of computing time from the Ohio Supercomputer Center. Portions of this work were presented at the Allerton Conference, October 2012, and FFT Workshop, February 2013.

P. Schniter is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: schniter@ece.osu.edu).

S. Rangan is with the Department of Electrical and Computer Engineering, New York University, Brooklyn, NY 10003 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2014.2386294

CPRL when $\mathbf{x} \in \mathbb{R}^N$, $M \gtrsim O(K^2/\log^2 N)$ measurements are *necessary*, which is disappointing because this greatly exceeds the $2K$ measurements that suffice for the optimal solver [21]. That said, the noise-robustness of PhaseLift-type algorithms and the sufficiency of $M \gtrsim O(K \log^2 N)$ samples for K -sparse signals with power-law decay has been recently established [26, Thm. 4]. Also, a cleverly initialized alternating minimization (AltMin) approach was recently proposed by Natrapalli, Jain, and Sanghavi in [27] that gives CPRL-like guarantees/performance with only $O(NK^3)$ complexity. However, this is still too complex for practical sparsities K , which tend to grow linearly with image size N .

Recently, Shechtman, Beck, and Eldar proposed the GrEedy Sparse PhAse Retrieval (GESPAR) algorithm [28], which applies fast 2-opt local search [29] to a sparsity constrained non-linear optimization formulation of the phase-retrieval problem. Numerical experiments (see Section IV-D) suggest that GESPAR handles higher sparsities K than the sparse-Fienup technique from [23], but at the cost of significantly increased runtime. In fact, due to the combinatorial nature of GESPAR's support optimization, its complexity scales very rapidly in K , making it impractical for many problems of interest.

In this work, we describe a novel² approach to compressive retrieval that is based on loopy belief propagation and, in particular, the *generalized approximate message passing* (GAMP) algorithm from [30]. In addition to describing and deriving our phase-retrieval GAMP (PR-GAMP) algorithm, we present a detailed numerical study of its performance. For i.i.d Gaussian, Fourier, and masked-Fourier matrices \mathbf{A} , we demonstrate that PR-GAMP performs far better than existing compressive phase-retrieval algorithms in terms of both success rate and runtime for large values K and N . Our experiments suggest that PR-GAMP requires approximately $4\times$ the number of measurements as phase-oracle GAMP (i.e., GAMP given the magnitude-and-phase measurements $\mathbf{u} = \mathbf{A}\mathbf{x} + \mathbf{w}$). Interestingly, for *non*-sparse signals in \mathbb{C}^N , the ratio of magnitude-only to magnitude-and-phase measurements necessary and sufficient for perfect recovery is also known to be $4\times$ (as $N \rightarrow \infty$) [15], [16]. Our experiments also suggest that PR-GAMP is robust to additive noise, giving mean-squared error that is only 3 dB worse than phase-oracle GAMP over a wide SNR range.

Notation

For matrices, we use boldface capital letters like \mathbf{A} , and we use \mathbf{A}^\top , \mathbf{A}^H , and $\|\mathbf{A}\|_F$ to denote the transpose, Hermitian transpose, and Frobenius norm, respectively. For vectors, we use boldface small letters like \mathbf{x} , and we use $\|\mathbf{x}\|_p = (\sum_n |x_n|^p)^{1/p}$ to denote the ℓ_p norm, with $x_n = [\mathbf{x}]_n$ representing the n^{th} element of \mathbf{x} . For random variable X , we write the pdf as $p_X(x)$, the expectation as $\mathbb{E}\{X\}$, and the variance as $\text{var}\{X\}$. In some cases where it does not cause confusion, we drop the subscript on $p_X(x)$ and write the pdf simply as $p(x)$. For a ‘‘circular Gaussian’’ random variable $X \in \mathbb{C}$ with mean m and variance v , we write the pdf as $p_X(x) = \mathcal{N}(x; m, v) \triangleq \frac{1}{\pi v} \exp(-|x -$

²We described an earlier version of PR-GAMP in the conference paper [1] and the workshop presentation [2].

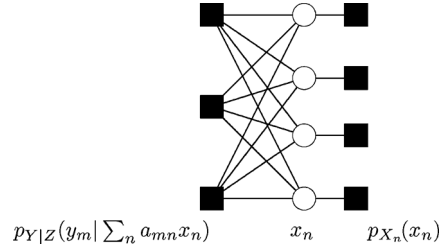


Fig. 1. GAMP factor graph, with white circles denoting random variables and black squares denoting pdf factors, for the case $M = 3$ and $N = 4$.

$m|v)$. Note that $X \sim \mathcal{N}(m, v)$ has real and imaginary components that are jointly Gaussian, uncorrelated, and of equal variance $v/2$. For the point mass at $x = 0$, we use the Dirac delta distribution $\delta(x)$. Finally, we use \mathbb{R} for the real field, \mathbb{C} for the complex field, $\text{Re}\{x\}$ and $\text{Im}\{x\}$ for the real and imaginary parts of x , and x^* for the complex conjugate of x .

II. BACKGROUND ON GAMP

The approximate message passing (AMP) algorithm was recently proposed by Donoho, Maleki, and Montanari [31], [32] for the task estimating a signal vector $\mathbf{x} \in \mathbb{R}^N$ from linearly transformed and additive-Gaussian-noise corrupted measurements³

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \in \mathbb{R}^M. \quad (1)$$

The Generalized-AMP (GAMP) algorithm proposed by Rangan [30] then extends the methodology of AMP to the generalized linear measurement model

$$\mathbf{y} = q(\mathbf{A}\mathbf{x} + \mathbf{w}) \in \mathbb{R}^M, \quad (2)$$

where $q(\cdot)$ is a component-wise nonlinearity. This nonlinearity facilitates the application of AMP to phase retrieval.

Both AMP and GAMP can be derived from the perspective of *belief propagation* [33], a Bayesian inference strategy that is based on a factorization of the signal posterior pdf $p(\mathbf{x}|\mathbf{y})$ into a product of simpler pdfs that, together, reveal the probabilistic structure in the problem. Concretely, if we model the signal coefficients in \mathbf{x} and noise samples in \mathbf{w} from (1)–(2) as statistically independent, so that $p(\mathbf{x}) = \prod_{n=1}^N p_{X_n}(x_n)$ and $p(\mathbf{y}|\mathbf{z}) = \prod_{m=1}^M p_{Y|Z}(y_m|z_m)$ for $\mathbf{z} \triangleq \mathbf{A}\mathbf{x}$, then we can factor the posterior pdf as

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \quad (3)$$

$$= \prod_{m=1}^M p_{Y|Z}(y_m | [\mathbf{A}\mathbf{x}]_m) \prod_{n=1}^N p_{X_n}(x_n), \quad (4)$$

yielding the factor graph in Fig. 1.

In belief propagation [33], beliefs about the unknown variables are passed among the nodes of the factor graph until all agree on a common set of beliefs. The set of beliefs passed

³Here and elsewhere, we use \mathbf{y} when referring to the M measurements that are available for signal reconstruction. In the canonical (noisy) compressive sensing problem, the measurements take the form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$, but in the (noisy) compressive phase retrieval problem, the measurements instead take the form $\mathbf{y} = |\mathbf{A}\mathbf{x} + \mathbf{w}|$.

TABLE I
THE GAMP ALGORITHM FROM [30] WITH T_{\max} ITERATIONS

input	$\mathbf{A}, \{p_{X_n}(\cdot), \hat{x}_n(1), \nu_n^x(1)\}_{n=1}^N, \{p_{Y Z}(y_m \cdot), \hat{s}_m(0)\}_{m=1}^M$
define	
	$p_{Z Y,P}(z y, \hat{p}; \nu^p) = \frac{p_{Y Z}(y z) \mathcal{N}(z; \hat{p}, \nu^p)}{\int_{\mathbb{C}} p_{Y Z}(y z') \mathcal{N}(z'; \hat{p}, \nu^p) dz'}$ (D1)
	$g_{\text{out},m}(\hat{p}, \nu^p) = \frac{1}{\nu^p} (\mathbb{E}_{Z Y,P}\{Z y_m, \hat{p}; \nu^p\} - \hat{p})$ (D2)
	$g'_{\text{out},m}(\hat{p}, \nu^p) = \frac{1}{\nu^p} \left(\frac{\text{var}_{Z Y,P}\{Z y_m, \hat{p}; \nu^p\}}{\nu^p} - 1 \right)$ (D3)
	$p_{X_n R_n}(x \hat{r}; \nu^r) = \frac{p_{X_n}(x) \mathcal{N}(x; \hat{r}, \nu^r)}{\int_{\mathbb{C}} p_{X_n}(x') \mathcal{N}(x'; \hat{r}, \nu^r) dx'}$ (D4)
	$g_{\text{in},n}(\hat{r}, \nu^r) = \mathbb{E}_{X_n R_n}\{X_n \hat{r}; \nu^r\}$ (D5)
	$g'_{\text{in},n}(\hat{r}, \nu^r) = \text{var}_{X_n R_n}\{X_n \hat{r}; \nu^r\}$ (D6)
for	$t = 1, 2, 3, \dots, T_{\max}$
	$\forall m: \nu_m^p(t) = \sum_{n=1}^N a_{mn} ^2 \nu_n^x(t)$ (R1)
	$\forall m: \hat{p}_m(t) = \sum_{n=1}^N a_{mn} \hat{x}_n(t) - \nu_m^p(t) \hat{s}_m(t-1)$ (R2)
	$\forall m: \hat{s}_m(t) = g_{\text{out},m}(\hat{p}_m(t), \nu_m^p(t))$ (R3)
	$\forall m: \nu_m^s(t) = -g_{\text{out},m}(\hat{p}_m(t), \nu_m^p(t))$ (R4)
	$\forall n: \nu_n^r(t) = \left(\sum_{m=1}^M a_{mn} ^2 \nu_m^s(t) \right)^{-1}$ (R5)
	$\forall n: \hat{r}_n(t) = \hat{x}_n(t) + \nu_n^r(t) \sum_{m=1}^M a_{mn}^* \hat{s}_m(t)$ (R6)
	$\forall n: \nu_n^x(t+1) = \nu_n^r(t) g'_{\text{in},n}(\hat{r}_n(t), \nu_n^r(t))$ (R7)
	$\forall n: \hat{x}_n(t+1) = g_{\text{in},n}(\hat{r}_n(t), \nu_n^r(t))$ (R8)
end	
output	$\{\hat{x}_n(T_{\max}+1), \nu_n^x(T_{\max}+1)\}_{n=1}^N, \{\hat{s}_m(T_{\max})\}_{m=1}^M$

into a given variable node are then used to determine the posterior pdf of that variable, or an approximation thereof. The sum-product algorithm [34] is perhaps the most well-known incarnation of belief propagation, wherein the messages take the form of pdfs and exact posteriors are guaranteed whenever the graph does not have loops. For graphs with loops, exact inference is known to be NP hard, and so loopy belief propagation (LBP) is not guaranteed to produce correct posteriors. Still, LBP has shown state-of-the-art performance on many problems in, e.g., decoding, computer vision, and compressive sensing [35].

The conventional wisdom surrounding LBP says that accurate inference is possible only when the circumference of the loops are relatively large. With (1)–(2), this would require that \mathbf{A} is a sparse matrix, which precludes most interesting cases of compressive inference, including compressive phase retrieval. Hence, the recent realization by Donoho, Maleki, Montanari, and Bayati that LBP-based compressive sensing is not only feasible [31], [32] for *dense* matrices \mathbf{A} , but provably accurate [36], [37], was a breakthrough. In particular, they established that, in the large-system limit (i.e., as $M, N \rightarrow \infty$ with M/N fixed) and under i.i.d sub-Gaussian \mathbf{A} , the iterations of AMP are governed by a state-evolution whose fixed points describe the algorithm's performance. To derive the AMP algorithm, [31], [32] proposed an ingenious set of message-passing approximations that become exact in the limit of large sub-Gaussian \mathbf{A} .

Remarkably, the “approximate message passing” (AMP) principles in [31], [32]—including the state evolution—can be extended from the linear model (1) to the generalized linear model in (2), as established in [30]. The GAMP algorithm from [30] is summarized in Table I. It is possible to recover the Bayesian AMP algorithm [32] from Table I by considering the special case of $g_{\text{out},m}(\hat{p}, \nu^p) = (y_m - \hat{p})/(\nu^p + \nu^w)$ in line (D2) and by replacing all terms $|a_{mn}|^2$ in lines (R1) and (R5) with the constant value M^{-1} (assuming that $\|\mathbf{A}\|_F^2 = N$). In the AMP literature, the term $\nu_m^p(t) \hat{s}_m(t-1)$ in (R2) is often referred to as the “momentum” or “Onsager correction” term.

As in [30], we state the GAMP algorithm in a way that facilitates the use of complex-valued quantities, which is the case of interest in phase retrieval. However, we note that the GAMP algorithm as stated in Table I is fully justified only in the case that all Gaussian random variables are circular (i.e., having independent real and imaginary components with identical variances), and we use $\mathcal{N}(z; \hat{z}, \nu^z)$ to denote the circular-Gaussian pdf in variable z with mean \hat{z} and variance ν^z . In the sequel, we detail how GAMP allows us to tackle the compressive phase retrieval problem.

III. PHASE RETRIEVAL GAMP

To apply the GAMP algorithm outlined in Table I to compressive phase retrieval, we specify a measurement likelihood function $p_{Y|Z}(y_m|\cdot)$ that models the lack of phase information in the observations y_m and a signal prior pdf $p_{X_n}(\cdot)$ that facilitates measurement compression, e.g., a sparsity-inducing pdf. In addition, we propose several extensions to the GAMP algorithm that aim to improve its robustness, and we propose an expectation-maximization method to learn the noise variance that parameterizes $p_{Y|Z}(y_m|\cdot)$.

A. Likelihood Function

Before deriving the likelihood function $p_{Y|Z}(y_m|\cdot)$, we introduce some notation. First, we will denote the noiseless transform outputs by

$$z_m \triangleq \mathbf{a}_m^H \mathbf{x} = |z_m| e^{j\phi_m} \text{ with } \phi_m \in [0, 2\pi), \quad (5)$$

where \mathbf{a}_m^H is the m th row of \mathbf{A} and $j \triangleq \sqrt{-1}$. Next, we will assume the presence of additive noise w_m and denote the noisy transform outputs by

$$u_m \triangleq z_m + w_m = |u_m| e^{j\theta_m} \text{ with } \theta_m \in [0, 2\pi). \quad (6)$$

Our (noisy) intensity measurements are then

$$y_m = |u_m| \text{ for } m = 1, \dots, M, \quad (7)$$

Henceforth, we assume additive white circular-Gaussian noise (AWGN) $w_m \sim \mathcal{N}(0, \nu^w)$. Thus, if we condition on z_m , then u_m is circular Gaussian with mean z_m and variance ν^w , and y_m is Rician with pdf [38]

$$p_{Y|Z}(y_m|z_m; \nu^w) = \frac{2y_m}{\nu^w} \exp\left(-\frac{y_m^2 + |z_m|^2}{\nu^w}\right) I_0\left(\frac{2y_m|z_m|}{\nu^w}\right) 1_{y_m \geq 0}, \quad (8)$$

where $I_0(\cdot)$ is the 0th-order modified Bessel function of the first kind.

The functions $g_{\text{out},m}(\cdot, \cdot)$ and $g'_{\text{out},m}(\cdot, \cdot)$ defined in lines (D1)–(D3) of Table I can be computed using the expressions

$$\mathbb{E}_{Z|Y,P}\{Z|y_m, \hat{p}_m; \nu_m^p\} = \frac{\int_{\mathbb{C}} z p_{Y|Z}(y_m|z; \nu^w) \mathcal{N}(z; \hat{p}_m, \nu_m^p) dz}{\int_{\mathbb{C}} p_{Y|Z}(y_m|z'; \nu^w) \mathcal{N}(z'; \hat{p}_m, \nu_m^p) dz'} \quad (9)$$

$$= \left(\frac{y_m}{1 + \nu^w/\nu_m^p} R_0(\varrho_m) + \frac{|\hat{p}_m|}{\nu_m^p/\nu^w + 1} \right) \frac{\hat{p}_m}{|\hat{p}_m|} \quad (10)$$

and

$$\begin{aligned} & \text{var}_{Z|Y,P}\{Z|y_m, \hat{p}_m; \nu_m^p\} \\ &= \frac{\int_{\mathbb{C}} |z|^2 p_{Y|Z}(y_m|z; \nu^w) \mathcal{N}(z; \hat{p}_m, \nu_m^p) dz}{\int_{\mathbb{C}} p_{Y|Z}(y_m|z'; \nu^w) \mathcal{N}(z'; \hat{p}_m, \nu_m^p) dz'} \\ & \quad - |\mathbb{E}_{Z|Y,P}\{Z|y_m, \hat{p}_m; \nu_m^p\}|^2 \end{aligned} \quad (11)$$

$$\begin{aligned} &= \frac{y_m^2}{(1 + \nu^w/\nu_m^p)^2} + \frac{|\hat{p}_m|^2}{(\nu_m^p/\nu^w + 1)^2} + \frac{1 + \varrho_m R_0(\varrho_m)}{1/\nu^w + 1/\nu_m^p} \\ & \quad - |\mathbb{E}_{Z|Y,P}\{Z|y_m, \hat{p}_m; \nu_m^p\}|^2, \end{aligned} \quad (12)$$

where

$$R_0(\varrho_m) \triangleq \frac{I_1(\varrho_m)}{I_0(\varrho_m)} \quad \text{and} \quad \varrho_m \triangleq \frac{2y_m |\hat{p}_m|}{\nu^w + \nu_m^p}, \quad (13)$$

as shown in Appendix A.

Whereas the above assumes that AWGN is added prior to the intensity step (7), it is also possible to consider post-intensity noise models, i.e.,

$$y_m = q(|z_m|) + w_m, \quad (14)$$

where common examples of $q(\cdot)$ include $q(|z|) = |z|$ and $q(|z|) = |z|^2$ (see, e.g., [25]) and where $w_m \sim p_W$ for a specified noise distribution p_W . The likelihood would then become

$$p_{Y|Z}(y_m|z_m) = p_W(y_m - q(|z_m|)), \quad (15)$$

and the functions $g_{\text{out},m}(\cdot, \cdot)$ and $g'_{\text{out},m}(\cdot, \cdot)$ defined in lines (D1)–(D3) of Table I would be computed as described in Appendix B. Note that, to assign zero likelihood to negative intensity measurements y_m , the assumed noise density $p_W(w)$ must have zero measure on the negative reals.

B. EM Update of the Noise Variance

Until now we have treated the noise variance ν^w as a known parameter. In practice, however, ν^w may be unknown, in which case it is not clear what value to use in (10) and (12). To address this problem, we now describe how ν^w can be learned using an expectation-maximization (EM) [39] procedure. The methodology is similar to that proposed in [40] for the case of a Gaussian $p_{Y|Z}(y_m|\cdot)$, but is more involved due to the fact that the $p_{Y|Z}(y_m|\cdot)$ used for phase-retrieval (recall (8)) is non-Gaussian.

Choosing \mathbf{x} as the hidden data, the standard form of the i th EM update is [39]

$$\widehat{\nu^w}[i+1] = \arg \max_{\nu^w > 0} \mathbb{E}\{\ln p(\mathbf{y}, \mathbf{x}; \nu^w) | \mathbf{y}; \widehat{\nu^w}[i]\}, \quad (16)$$

where square brackets are used to distinguish EM iterations from GAMP iterations (recall Table I). Because the true posterior pdf $p(\mathbf{x}|\mathbf{y})$ needed for (16) is generally NP-hard to compute [41], Appendix C describes an approximate EM update of the form

$$\widehat{\nu^w}(t+1) = \arg \min_{\nu^w > 0} \tilde{J}(\nu^w; \widehat{\mathbf{r}}(t), \nu^r(t), \widehat{\nu^w}(t)), \quad (17)$$

which performs one EM iteration i for every GAMP iteration t , allowing us to state the EM update (17) using GAMP iterations. In (17), \tilde{J} is a certain Bethe free entropy and $(\widehat{\mathbf{r}}(t), \nu^r(t))$ are the results of lines (R5)–(R6) in Table I when GAMP is run under the noise variance $\widehat{\nu^w}(t)$. (See Appendix C for details.)

C. Signal Prior Distribution

GAMP offers great flexibility with respect to the choice of prior distribution on the signal vector \mathbf{x} . In this work, we focus on separable priors, which have the form $p(\mathbf{x}) = \prod_{n=1}^N p_{X_n}(x_n)$ with arbitrary $p_{X_n}(\cdot)$ (recalling (4)), but we note that various forms of non-separable priors can be supported using the “turbo GAMP” formulation proposed in [42] or the “analysis GAMP” formulation proposed in [43].

For separable priors, $p_{X_n}(\cdot)$ should be chosen to reflect whatever form of probabilistic structure is known about coefficient x_n . For example, if $\mathbf{x} \in \mathbb{C}^N$ is known to be K -sparse, but nothing is known about the support, then it is typical to choose the Bernoulli-Gaussian (BG) model

$$p_{X_n}(x_n) = (1 - \lambda)\delta(x_n) + \lambda\mathcal{N}(x_n; 0, \varphi), \quad (18)$$

with sparsity rate $\lambda = \frac{K}{N}$ and non-zero-coefficient variance φ that, if unknown, can be estimated from the observations via [44, eqn. (71)]

$$\varphi = \frac{\|\mathbf{y}\|_2^2 - M\nu^w}{\lambda\|\mathbf{A}\|_F^2}, \quad (19)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. For this BG prior, expressions for the thresholding functions $g_{\text{in},n}(\cdot, \cdot)$ and $g'_{\text{in},n}(\cdot, \cdot)$ defined in lines (D5)–(D6) of Table I were given in [42]. When the sparsity rate λ in (18) is unknown, it can be learned using the EM-BG procedure described in [44]. In most cases, improved performance is obtained when a Gaussian mixture (GM) pdf is used in place of the Gaussian pdf in (18) [44].

Various extensions of the above are possible. For example, when all coefficients x_n are known to be real-valued or positive, the circular-Gaussian pdf in (18) should be replaced by a real-Gaussian or truncated-Gaussian pdf, respectively, or even a truncated-GM [45]. Furthermore, when certain coefficient subsets are known to be more or less sparse than others, a non-uniform sparsity [46] rate λ_n can be used in (18).

D. GAMP Normalization and Damping

To increase the numerical robustness of GAMP, we propose to normalize certain internal GAMP variables. To do this, we define $\alpha(t) \triangleq \frac{1}{M} \sum_{m=1}^M \nu_m^p(t)$ (which tends to grow very small with t at high SNR), normalize both $\widehat{s}_m(t)$ and $\nu_m^s(t)$ (which tend to grow very large) by $1/\alpha(t)$, and normalize $\nu_n^r(t)$ (which tends to grow very small) by $\alpha(t)$. This prevents the normalized variables \widehat{s}_m , $\underline{\nu}_m^s(t)$, and $\underline{\nu}_n^r(t)$ from growing very large and causing numerical precision issues in Matlab. We note that, under infinite precision, these normalizations would cancel each other out and have absolutely no effect. The resulting normalized GAMP iterations are shown in Table II.

To reduce the chance of GAMP divergence, we propose to “damp” certain variable updates. Damping is a technique commonly used in loopy belief propagation (see, e.g., [47])

TABLE II
GAMP STEPS WITH VARIANCE NORMALIZATION $\alpha(t)$ AND DAMPING
PARAMETER $\beta \in (0, 1]$

for $t = 1, 2, 3, \dots, T_{\max}$	
$\forall m : \nu_m^p(t) = \beta \sum_{n=1}^N a_{mn} ^2 \nu_n^x(t) + (1-\beta) \nu_m^p(t-1)$	(S1)
$\alpha(t) = \frac{1}{M} \sum_{m=1}^M \nu_m^p(t)$	(S2)
$\forall m : \hat{p}_m(t) = \sum_{n=1}^N a_{mn} \hat{x}_n(t) - \frac{\nu_m^p(t)}{\alpha(t)} \hat{s}_m(t-1)$	(S3)
$\forall m : \hat{s}_m(t) = \beta \alpha(t) g_{\text{out},m}(\hat{p}_m(t), \nu_m^p(t)) + (1-\beta) \hat{s}_m(t-1)$	(S4)
$\forall m : \nu_m^s(t) = -\beta \alpha(t) g_{\text{out},m}(\hat{p}_m(t), \nu_m^p(t)) + (1-\beta) \nu_m^s(t-1)$	(S5)
$\forall n : \nu_n^r(t) = (\sum_{m=1}^M a_{mn} ^2 \nu_m^s(t))^{-1}$	(S6)
$\forall n : \bar{x}_n(t) = \beta \hat{x}_n(t) + (1-\beta) \bar{x}_n(t-1)$	(S7)
$\forall n : \hat{r}_n(t) = \bar{x}_n(t) + \nu_n^r(t) \sum_{m=1}^M a_{mn}^* \hat{s}_m(t)$	(S8)
$\forall n : \nu_n^x(t+1) = \alpha(t) \nu_n^r(t) g'_{\text{in},n}(\hat{r}_n(t), \alpha(t) \nu_n^r(t))$	(S9)
$\forall n : \hat{x}_n(t+1) = g_{\text{in},n}(\hat{r}_n(t), \alpha(t) \nu_n^r(t))$	(S10)
end	

to reduce the chance of divergence, although at the cost of convergence speed. For GAMP, it was established in [48] that damping is both necessary and sufficient to guarantee *global* convergence under arbitrary \mathbf{A} in the case of Gaussian $p_{X_n}(\cdot)$ and $p_{Y|Z}(y_m|\cdot)$. Similarly, [48] established that damping is both necessary and sufficient to guarantee the *local* convergence of GAMP under arbitrary \mathbf{A} in the case of strictly log-concave $p_{X_n}(\cdot)$ and $p_{Y|Z}(y_m|\cdot)$. For general $p_{X_n}(\cdot)$ and $p_{Y|Z}(y_m|\cdot)$, theory is (to the authors' knowledge) lacking, but empirical results (see, e.g., [49]) suggest that the use of damping in GAMP can be very effective. Table II presents a version of damped GAMP that uses a common damping parameter $\beta \in (0, 1]$ throughout the algorithm: when $\beta = 1$, the algorithm reduces to the original GAMP algorithm, but when $\beta < 1$, the updates in lines (S1), (S4), (S5), and (S7) are slowed. Our numerical experiments suggest that the value $\beta = 0.25$ works well for phase retrieval. One consequence of the proposed damping implementation is the existence of additional state variables like $\bar{x}_n(t)$. To avoid the need to initialize these variables, we use $\beta = 1$ during the first iteration. We note that the damping modification described here is the one included in the public domain GAMPmatlab implementation,⁴ which differs slightly from the one described in [48].

E. Avoiding Bad Local Minima

As is well known [22], [23], [27], [28], the compressive phase retrieval problem is plagued by bad local minima. We now propose methods to randomly initialize and restart PR-GAMP that aim to avoid these bad local minima. Our empirical experience (see Section IV) suggests that the existence of bad local minima is a more serious issue with Fourier \mathbf{A} than with randomized (e.g., i.i.d Gaussian or masked-Fourier) \mathbf{A} .

1) *GAMP Initialization*: The GAMP algorithm in Table I requires an initialization of the signal coefficient estimates $\{\hat{x}_n(1)\}_{n=1}^N$, their variances $\{\nu_n^x(1)\}_{n=1}^N$, and the state variables $\{\hat{s}_m(0)\}_{m=1}^M$ (which can be interpreted as Lagrange multipliers [48]). The standard procedure outlined in [30] uses the fixed initialization $\hat{x}_n(1) = \mathbb{E}\{X_n\}$, $\nu_n^x(1) = \text{var}\{X_n\}$, $\hat{s}_m(0) = 0$. But, from this fixed initialization, GAMP may converge to a bad local minimum. To allow the possibility of avoiding this bad local minima, we propose to *randomly* initialize and restart GAMP multiple times if needed. For the

random initializations, we propose to draw each $\hat{x}_n(1)$ as an independent realization of the random variable X_n . This way, the empirical mean of $\{\hat{x}_n(1)\}_{n=1}^N$ matches that of the standard initialization from [30]. Likewise, we propose to initialize $\nu_n^x(1)$, for all n , at the empirical variance of $\{\hat{x}_n(1)\}_{n=1}^N$.

2) *EM Initialization*: For the EM algorithm described in Section III-B, we must choose the initial noise-variance estimate $\widehat{\nu}^w(1)$. Even when accurate knowledge of ν^w is available, our numerical experience leads us to believe that setting $\widehat{\nu}^w(1)$ at a relatively large value can help to avoid bad local minima. In particular, our empirical experience leads us to suggest setting $\widehat{\nu}^w(1)$ in correspondence with an initial SNR estimate of 10, i.e., $\widehat{\nu}^w(1) = \frac{\|\mathbf{y}\|_2^2}{M(\text{SNR}_{\text{init}}+1)}$ with $\text{SNR}_{\text{init}} = 10$.

3) *Multiple Restarts*: To further facilitate the avoidance of bad local minima, we propose to run multiple attempts of EM-GAMP, each using a different random GAMP initialization (constructed as above). The attempt that yields the lowest normalized residual ($\text{NR} \triangleq \|\mathbf{y} - |\mathbf{A}\hat{\mathbf{x}}\|_2 / \|\mathbf{y}\|_2$) is then selected as the algorithm output. The efficacy of multiple attempts is numerically investigated in Section IV.

Furthermore, to avoid unnecessary restarts, we allow the algorithm to be stopped as soon as the NR drops below a user-defined stopping tolerance of NR_{stop} . When the true SNR is known, we suggest setting $\text{NR}_{\text{stop}} \text{dB} = -(\text{SNR}_{\text{true}} \text{dB} + 2)$.

Algorithm Summary: The PR-GAMP algorithm is summarized in Table III, where A_{\max} controls the number of attempts, SNR_{init} controls the initial SNR, and NR_{stop} controls the stopping tolerance.

IV. NUMERICAL RESULTS

In this section we numerically investigate the performance of PR-GAMP⁵ under various scenarios and in comparison to several existing algorithms: Compressive Phase Retrieval via Lifting (CPRL) [24], GrEedy Sparse PhASE Retrieval (GESPAR) from [28], and the sparse Fienup technique from [23]. As a benchmark, we also compare to ‘‘phase oracle’’ (PO) GAMP, i.e., GAMP operating on the magnitude-and-phase measurements $\mathbf{u} = \mathbf{A}\mathbf{x} + \mathbf{w}$ rather than on the intensity measurements $\mathbf{y} = |\mathbf{u}|$.

Unless otherwise noted, we generated random realizations of the true signal vector \mathbf{x} as K -sparse length- N with support chosen uniformly at random and with nonzero coefficients drawn i.i.d zero-mean circular-Gaussian. Then, for a given matrix \mathbf{A} , we generated M noisy intensity measurements $\mathbf{y} = |\mathbf{A}\mathbf{x} + \mathbf{w}|$, where \mathbf{w} was i.i.d circular-Gaussian with variance selected to achieve a target signal-to-noise ratio of $\text{SNR} \triangleq \|\mathbf{A}\mathbf{x}\|_2^2 / \mathbb{E}\{\|\mathbf{w}\|_2^2\}$. Finally, each algorithm computed an estimate $\hat{\mathbf{x}}$ from (\mathbf{y}, \mathbf{A}) in an attempt to best match \mathbf{x} up to the inherent level of ambiguity. We recall that, for any \mathbf{A} , the magnitude $|\mathbf{A}\mathbf{x}|$ is invariant to global phase rotations in \mathbf{x} . For Fourier \mathbf{A} and real-valued \mathbf{x} , the magnitudes of $\mathbf{A}\mathbf{x}$ are also invariant to flips and circular shifts of \mathbf{x} . Performance was then assessed using normalized mean-squared error on the disambiguated estimate:

$$\text{NMSE}(\hat{\mathbf{x}}) \triangleq \min_{\Theta} \frac{\|\mathbf{x} - \text{disambig}(\hat{\mathbf{x}}, \Theta)\|_2^2}{\|\mathbf{x}\|_2^2}, \quad (20)$$

⁵PR-GAMP is part of the GAMPmatlab package at <http://sourceforge.net/projects/gampmatlab/>.

⁴<http://sourceforge.net/projects/gampmatlab/>

TABLE III
PROPOSED PR-GAMP ALGORITHM WITH A_{\max} ATTEMPTS, SNR
INITIALIZATION SNR_{init} , AND STOPPING RESIDUAL NR_{stop}

```

input  $\mathbf{y}, \mathbf{A}, \{p_{X_n}(\cdot)\}_{n=1}^N, \text{SNR}_{\text{init}}, \text{NR}_{\text{stop}}, A_{\max}, T_{\max}$ 
 $\widehat{\nu}^w(1) = \frac{\|\mathbf{y}\|_2^2}{M(\text{SNR}_{\text{init}} + 1)}$ 
 $\forall m : \widehat{s}_m(0) = 0$ 
 $\text{NR}_{\text{best}} = \infty$ 
for  $a = 1, 2, 3, \dots, A_{\max}$ ,
  draw random  $\widehat{\mathbf{x}}(1)$ 
   $\forall n : \nu_n^x(1) = \|\widehat{\mathbf{x}}(1) - \mathbb{E}\{\mathbf{X}\}\|_2^2/N$ 
  for  $t = 1, 2, 3, \dots, T_{\max}$ 
     $(\widehat{\mathbf{x}}(t+1), \nu^x(t+1), \widehat{\mathbf{s}}(t), \widehat{\mathbf{r}}(t), \nu^r(t))$ 
    = GAMP( $\mathbf{A}, \{p_{X_n}(\cdot)\}_{n=1}^N, \{p_{Y|Z}(y_m|z; \widehat{\nu}^w(t))\}_{m=1}^M,$ 
       $\widehat{\mathbf{x}}(t), \nu^x(t), \widehat{\mathbf{s}}(t-1)$ )
     $\widehat{\nu}^w(t+1) = \arg \min_{\nu^w > 0} \widetilde{J}(\nu^w; \widehat{\mathbf{r}}(t), \nu^r(t), \widehat{\nu}^w(t))$ 
  end
   $\text{NR} = \|\mathbf{y} - \mathbf{A}\widehat{\mathbf{x}}(T_{\max}+1)\|_2^2 / \|\mathbf{y}\|_2^2$ 
  if  $\text{NR} < \text{NR}_{\text{best}}$ 
     $\widehat{\mathbf{x}}_{\text{best}} = \widehat{\mathbf{x}}(T_{\max}+1)$ 
     $\text{NR}_{\text{best}} = \text{NR}$ 
  end
  if  $\text{NR} < \text{NR}_{\text{stop}}$ 
    stop
  end
end
output  $\widehat{\mathbf{x}}_{\text{best}}$ 

```

where Θ are the ambiguity parameters. When computing empirical phase-transition curves, we defined a “successful” recovery as one that produced $\text{NMSE} < 10^{-6}$.

A. Empirical Phase Transitions: I.I.D Gaussian \mathbf{A}

First we investigated the phase-transition performance of PR-GAMP with i.i.d circular-Gaussian sensing matrices \mathbf{A} . Fig. 2 plots the empirical success rate (averaged over 100 independent problem realizations) as a function of signal sparsity K and measurement length M for a fixed signal length of $N = 512$. Here we used $\text{SNR} = 100$ dB, which makes the observations essentially “noiseless,” and we allowed PR-GAMP up to 10 attempts from random initializations (i.e., $A_{\max} = 10$ in Table III). The figure shows a “phase transition” behavior that separates the (K, M) plane into two regions: perfect recovery in the top-left and failure in the bottom-right. Moreover, the figure suggests that, to recover K -sparse Bernoulli-Gaussian signals with $K \ll N$, approximately $M \geq 2K \log_2(N/K)$ intensity measurements suffice for PR-GAMP.

To investigate how well (versus how often) PR-GAMP recovers the signal, we plot the median NMSE achieved over the same problem realizations in Fig. 3. There we see that the signal estimates were extremely accurate on the good side of the phase transition.

To investigate the effect of number-of-attempts A_{\max} , we extracted the 50%-success contour (i.e., the phase-transition curve) from Fig. 2 and plotted it in Fig. 4, along with the corresponding contours obtained under different choices of A_{\max} . Fig. 4 suggests that, in the case of i.i.d \mathbf{A} , there is relatively little to gain from multiple restarts from random realizations. With Fourier \mathbf{A} , however, we will see in the sequel that multiple restarts are indeed important.

Fig. 4 also plots the phase-transition curve of phase-oracle (PO)-GAMP calculated from the same problem realizations.

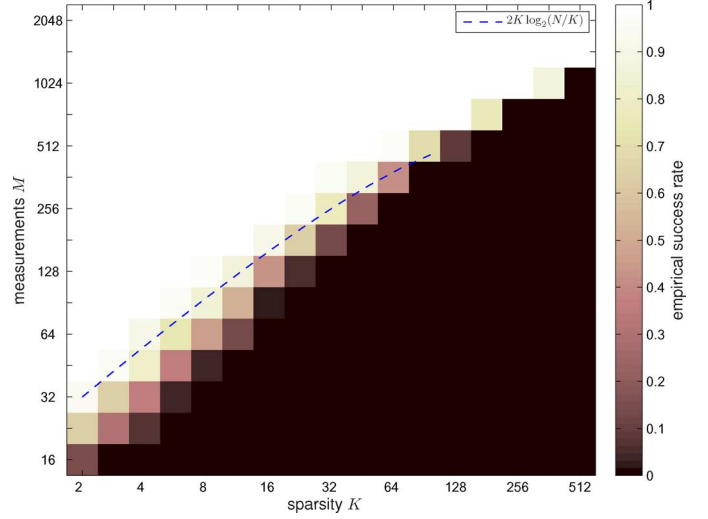


Fig. 2. Empirical probability of successful PR-GAMP recovery of an $N = 512$ -length signal, versus signal sparsity K and number of intensity measurements M , using i.i.d Gaussian \mathbf{A} at $\text{SNR} = 100$ dB. Here, PR-GAMP was allowed up to 10 attempts from different random initializations.

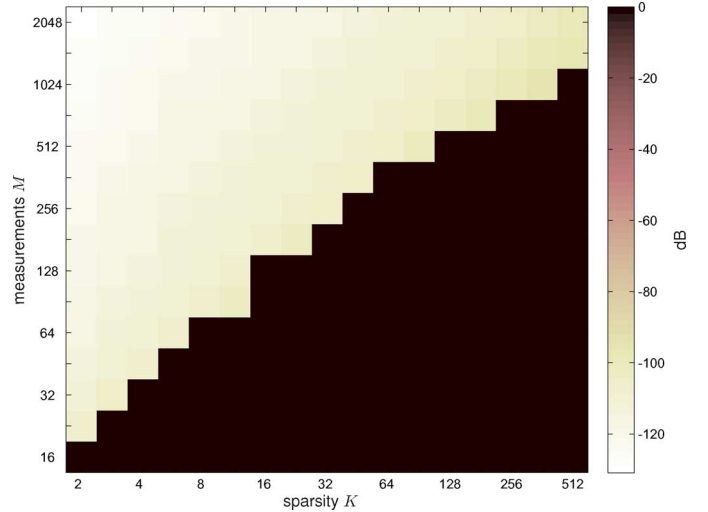


Fig. 3. Median **NMSE** for PR-GAMP recovery of an $N = 512$ -length signal, versus signal sparsity K and number of intensity measurements M , using i.i.d Gaussian \mathbf{A} at $\text{SNR} = 100$ dB. Here, PR-GAMP was allowed up to 10 attempts from different random initializations.

A comparison of the PO-GAMP phase transition to the PR-GAMP phase transition suggests that PR-GAMP requires approximately $4\times$ the number of measurements as PO-GAMP, regardless of sparsity rate K , for Bernoulli-Gaussian signals. Remarkably, this “ $4\times$ ” rule generalizes what is known about the recovery of *non*-sparse signals in \mathbb{C}^N , where the ratio of (necessary and sufficient) magnitude-only to magnitude-and-phase measurements is also $4\times$ (as $N \rightarrow \infty$) [15], [16].

Overall, Figs. 2–4 demonstrate that PR-GAMP is indeed capable of *compressive* phase retrieval, i.e., successful \mathbb{C}^N -signal recovery from $M \ll 4N$ intensity measurements, when the signal is sufficiently sparse. Moreover, to our knowledge, these phase transitions are far better than those reported for existing algorithms in the literature.

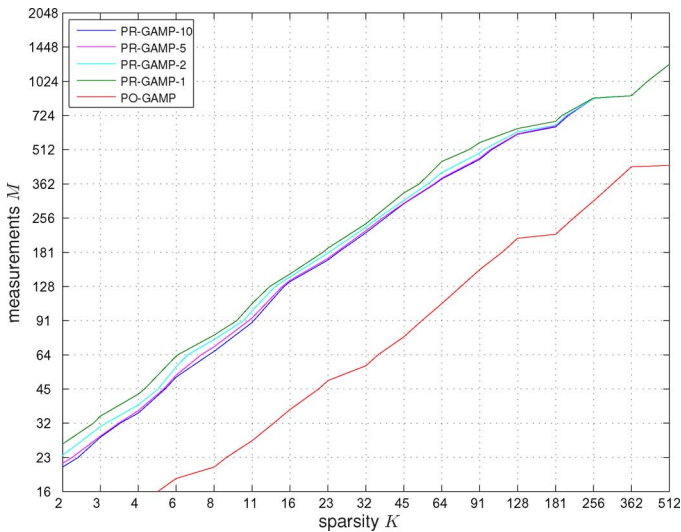


Fig. 4. 50%-success contours for PR-GAMP and phase-oracle GAMP recovery of an $N = 512$ -length signal, versus signal sparsity K and number of intensity measurements M , using i.i.d Gaussian \mathbf{A} at $\text{SNR} = 100$ dB. PR-GAMP- A_{\max} denotes PR-GAMP under a maximum of A_{\max} attempts.

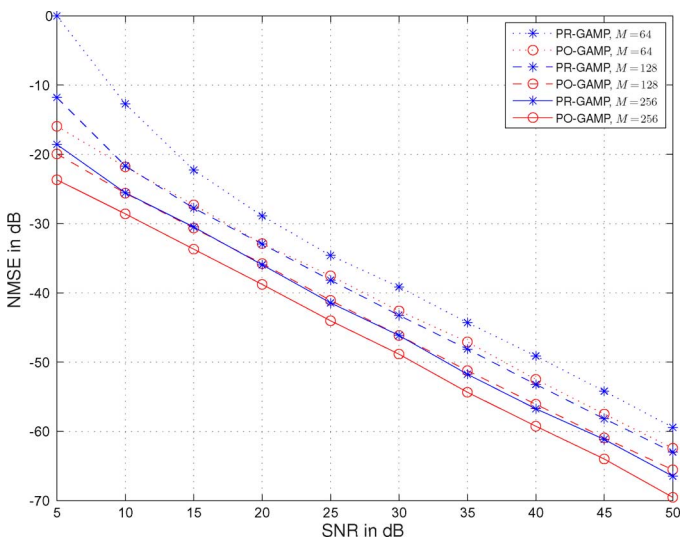


Fig. 5. Median **NMSE** for PR-GAMP and phase-oracle GAMP recovery of an $N = 512$ -length $K = 4$ -sparse signal, versus **SNR**, from $M \in \{64, 128, 256\}$ measurements and i.i.d Gaussian \mathbf{A} .

B. Robustness to Noise

We now demonstrate the robustness of PR-GAMP to non-trivial levels of additive white circular-Gaussian noise \mathbf{w} in the M intensity measurements $\mathbf{y} = |\mathbf{Ax} + \mathbf{w}|$. As before, we use $N = 512$ -length K -sparse Bernoulli-Gaussian signals and i.i.d Gaussian \mathbf{A} , but now we focus on sparsity $K = 4$ and number of measurements $M \in \{64, 128, 256\}$. We note that these (K, M) pairs are all on the good side of the phase-transition in Fig. 2, although $(K, M) = (4, 64)$ is near the boundary. Fig. 5 shows median NMSE performance over 200 independent problem realizations as a function of $\text{SNR} \triangleq \|\mathbf{Ax}\|_2^2 / \|\mathbf{w}\|_2^2$. There we see that, for most of the tested (M, SNR) pairs, PR-GAMP performs only about 3 dB worse than PO-GAMP. This 3 dB gap can be explained by the fact that PO-GAMP

TABLE IV
EMPIRICAL SUCCESS RATE AND MEDIAN RUNTIME OVER 100 PROBLEM REALIZATIONS FOR SEVERAL COMBINATIONS OF SIGNAL LENGTH N , MEASUREMENT LENGTH M , AND SIGNAL SPARSITY $K = 1$

	$(M, N) = (20, 32)$	$(M, N) = (30, 48)$	$(M, N) = (40, 64)$
CPRL	1.00 (3.4 sec)	1.00 (37 sec)	1.00 (434 sec)
PR-GAMP	1.00 (0.18 sec)	1.00 (0.17 sec)	1.00 (0.16 sec)

TABLE V
EMPIRICAL SUCCESS RATE AND MEDIAN RUNTIME OVER 100 PROBLEM REALIZATIONS FOR SEVERAL COMBINATIONS OF SIGNAL LENGTH N , MEASUREMENT LENGTH M , AND SIGNAL SPARSITY $K = 2$

	$(M, N) = (20, 32)$	$(M, N) = (30, 48)$	$(M, N) = (40, 64)$
CPRL	0.55 (4.1 sec)	0.65 (42 sec)	0.66 (496 sec)
PR-GAMP	0.93 (0.25 sec)	1.00 (0.21 sec)	1.00 (0.19 sec)

is able to average the noise over twice as many real-valued measurements as PR-GAMP (i.e., $\{\text{Re}\{u_m\}, \text{Im}\{u_m\}\}_{m=1}^M$ versus $\{|u_m|\}_{m=1}^M$). Fig. 5 shows that the performance gap grows beyond 3 dB when both the SNR is very low and the measurements are very few. But this may reflect a fundamental performance limitation rather than a weakness in PR-GAMP.

C. Comparison to CPRL

In this section, we present compare PR-GAMP to the state-of-the-art convex-relaxation approach to compressive phase retrieval, CPRL [24]. To implement CPRL, we used the authors' CVX-based matlab code⁶ under default algorithmic settings. We also tried the authors' ADMM implementation, but found that it gave significantly worse performance. As before, we examine the recovery of a K -sparse signal in \mathbb{C}^N from M intensity measurements $\mathbf{y} = |\mathbf{Ax} + \mathbf{w}|$, but now we use $\mathbf{A} = \Phi\mathbf{F}$ with i.i.d circular-Gaussian Φ and discrete Fourier transform (DFT) \mathbf{F} , to be consistent with the setup assumed in [24].

Table IV shows empirical success⁷ rate and runtime (on a standard personal computer) for a problem with sparsity $K = 1$, signal lengths $N \in \{32, 48, 64\}$, and compressive measurement lengths $M \in \{20, 30, 40\}$. The table shows that, over 100 problem realizations, both algorithms were 100% successful in recovering the signal at all tested combinations of (M, N) . But the table also shows that CPRL's runtime increased rapidly with the signal dimensions, whereas that of PR-GAMP remained orders-of-magnitude smaller and relatively independent of (M, N) over the tested range.⁸

Table V repeats the experiment carried out in Table IV, but at the sparsity $K = 2$. For this more difficult problem, the table shows that CPRL was much less successful at recovering the signal than PR-GAMP. Meanwhile, the runtimes reported in Table V again show that CPRL's complexity scaled rapidly with the problem dimensions, whereas GAMP's complexity stayed orders-of-magnitude smaller and relatively constant over the tested problem dimensions. In fact, the comparisons conducted

⁶<http://users.isy.liu.se/rt/ohlsson/code/CPRL.zip>

⁷Since CPRL rarely gave $\text{NMSE} < 10^{-6}$, we reduced the definition of "success" to $\text{NMSE} < 10^{-4}$ for this subsection only.

⁸Although the complexity of GAMP is known to scale as $O(MN)$ for this type of \mathbf{A} , the values of M and N in Tables IV and V are too small for this scaling law to manifest. Instead, the runtime values in these tables are biased by the overhead computations associated with Matlab's object-oriented programming environment.

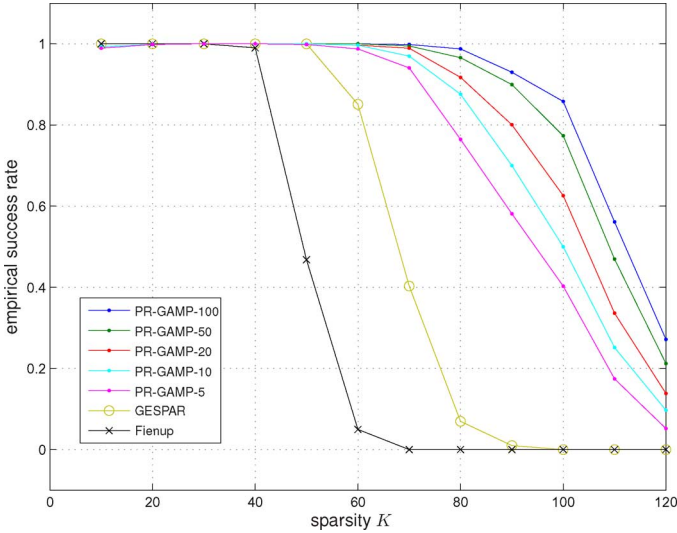


Fig. 6. Empirical success rate versus sparsity K in the recovery of an $N = 1024$ -length real-valued signal from $M = 1024$ 2D-Fourier intensities at $\text{SNR} = 100$ dB. PR-GAMP- A denotes PR-GAMP under a maximum of A attempts.

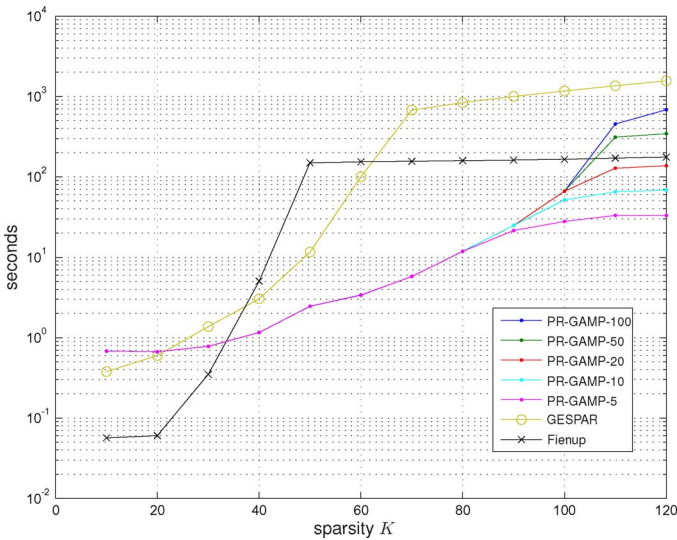


Fig. 7. Median runtime versus sparsity K in the recovery of an $N = 1024$ -length real-valued signal from $M = 1024$ 2D-Fourier intensities at $\text{SNR} = 100$ dB. PR-GAMP- A denotes PR-GAMP under a maximum of A attempts.

in this section were restricted to very small problem dimensions precisely due to the poor complexity scaling of CPRL.

D. Comparison to Sparse-Fienup and GESPAR: Fourier \mathbf{A}

In this section, we compare PR-GAMP to the sparse-Fienup [23] and GESPAR⁹[28] algorithms. This comparison requires¹⁰ that we restrict our attention to Fourier-based \mathbf{A} and real-valued

⁹For GESPAR, we used the November 2013 version of the Matlab code provided by the authors at <https://sites.google.com/site/yoavshechtman/resources/software>.

¹⁰The sparse Fienup from [23] requires $\mathbf{A}^H \mathbf{A}$ to be a (scaled) identity matrix. Although GESPAR can in principle handle generic \mathbf{A} , the implementation provided by the authors is based on 1D and 2D Fourier \mathbf{A} and is not easily modified.

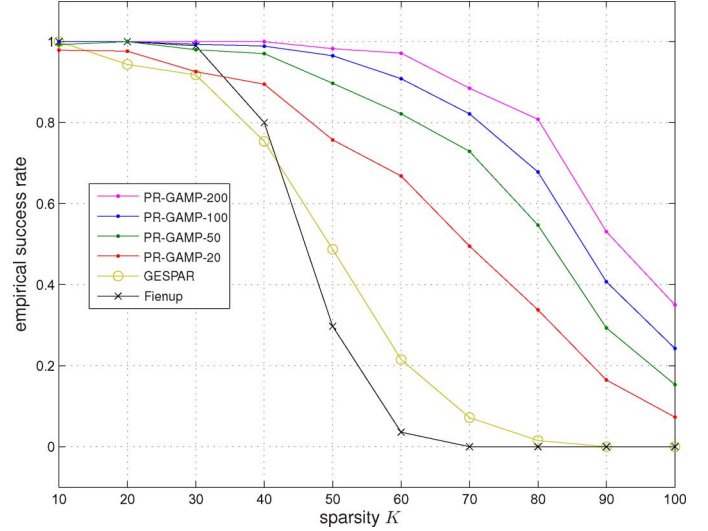


Fig. 8. Empirical success rate versus sparsity K in the recovery of an $N = 512$ -length real-valued signal from $M = 1024$ 1D-Fourier intensities at $\text{SNR} = 100$ dB. PR-GAMP- A denotes PR-GAMP under a maximum of A attempts.

sparse vectors \mathbf{x} . For the experiments below, we generated realizations of \mathbf{x} as described earlier, but now with the non-zero elements drawn from a real-Gaussian distribution. Also, we used $ITER = 6400$ in GESPAR as recommended by the authors in [28], and we allowed sparse-Fienup 1000 attempts from random initializations.

We first consider 2D Fourier \mathbf{A} , which is especially important for imaging applications. In particular, we repeat an experiment from [28], where the measurement and signal lengths were fixed at $M = N$ and the signal sparsity K was varied. For $N = 1024$, Fig. 6 shows the empirical success rate (over 200 realizations) for PR-GAMP, GESPAR, and sparse Fienup. Meanwhile, Fig. 7 shows the corresponding median runtime for each algorithm, where all algorithms leveraged fast Fourier transform (FFT) implementations of \mathbf{A} . From Fig. 6, we can see that PR-GAMP produced a significantly better phase-transition than GESPAR and sparse Fienup. Meanwhile, from Fig. 7 we see that, for the challenging case of $K \geq 40$, PR-GAMP-10 had uniformly better runtime *and* success rate than GESPAR and sparse Fienup.

Next we consider 1D Fourier \mathbf{A} . Again, we repeat an experiment from [28], where the measurement and signal lengths were fixed at $M = 2N$ and the signal sparsity K was varied. For $N = 1024$, Fig. 8 shows the empirical success rate (over 200 realizations) for PR-GAMP, GESPAR, and sparse Fienup, and Fig. 7 shows the corresponding median runtimes. From Fig. 8, we can see that PR-GAMP produced a significantly better phase-transition than GESPAR and sparse Fienup. Meanwhile, from Fig. 9 we see that, for the challenging case of $K \geq 40$, PR-GAMP-20 had uniformly better runtime *and* success rate than GESPAR and sparse Fienup.

Comparing the results in this section to those in Section IV-A, we observe that the PR-GAMP, GESPAR, and Fienup algorithms had a much more difficult time with Fourier matrices \mathbf{A} than with i.i.d. matrices \mathbf{A} . Similar observations were made in previous studies, leading to proposals of randomized Fourier-based phase retrieval, e.g., using “coded” binary masks [50].

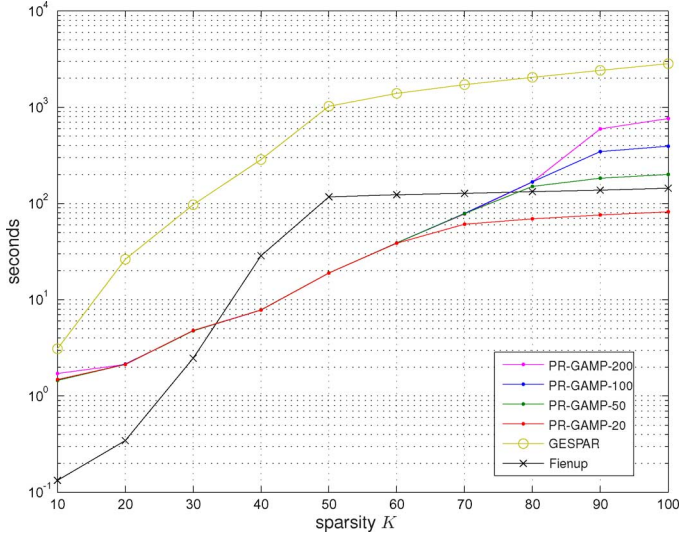


Fig. 9. Median runtime versus sparsity K in the recovery of an $N = 512$ -length real-valued signal from $M = 1024$ 1D-Fourier intensities at $\text{SNR} = 100$ dB. PR-GAMP- A denotes PR-GAMP under a maximum of A attempts.

Also, we notice that the use of multiple restarts in PR-GAMP was much more important with Fourier \mathbf{A} than it was with i.i.d \mathbf{A} .

E. Practical Image Recovery With Masked Fourier \mathbf{A}

Finally, we demonstrate practical image recovery from compressed intensity measurements. For this experiment, the signal \mathbf{x} was the $N = 65536$ -pixel grayscale image shown on the left of Fig. 10, which has a sparsity of $K = 6678$. Since this image is real and non-negative, we ran PR-GAMP with a non-negative-real-BG prior [45], as opposed to the BG prior (18) used in previous experiments.

For the first set of experiments, we used a “masked” Fourier transformation $\mathbf{A} \in \mathbb{C}^{M \times N}$ of the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{J}_1 \mathbf{F} \mathbf{D}_1 \\ \mathbf{J}_2 \mathbf{F} \mathbf{D}_2 \\ \mathbf{J}_3 \mathbf{F} \mathbf{D}_3 \\ \mathbf{J}_4 \mathbf{F} \mathbf{D}_4 \end{bmatrix}, \quad (21)$$

where \mathbf{F} was a 2D DFT matrix of size $N \times N$, \mathbf{D}_i were diagonal “masking” matrices of size $N \times N$ with diagonal entries drawn uniformly at random from $\{0, 1\}$, and \mathbf{J}_i were “selection” matrices of size $\frac{M}{4} \times N$ constructed from rows of the identity matrix drawn uniformly at random. The matrices \mathbf{D}_i and \mathbf{J}_i help to “randomize” the DFT, and they circumvent unicity issues such as shift and flip ambiguities. For phase retrieval, the use of image masks was discussed in [50]. Note that, because \mathbf{D}_i and \mathbf{J}_i are sparse and \mathbf{F} has a fast FFT-based implementation, the overall matrix \mathbf{A} has a fast implementation.

To eliminate the need for the expensive matrix multiplications with the elementwise-squared versions of \mathbf{A} and \mathbf{A}^H , as specified in lines (S1) and (S6) of Table II, GAMP was run in “uniform variance” mode, meaning that $\{\nu_m^p(t)\}_{m=1}^M$ were approximated by $\nu^p(t) \triangleq \frac{1}{M} \sum_{m'=1}^M \nu_{m'}^p(t)$; similar was done with $\{\nu_m^s(t)\}_{m=1}^M$, $\{\nu_n^r(t)\}_{n=1}^N$, and $\{\nu_n^x(t)\}_{n=1}^N$. The result is that lines (S1)-(S2) in Table II become

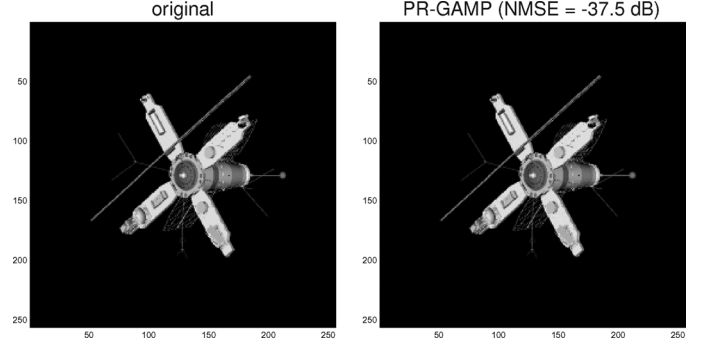


Fig. 10. Original image (left) and a typical PR-GAMP-recovery (right) from $M = N$ masked-Fourier intensity measurements at $\text{SNR} = 30$ dB, which took 1.8 seconds.

$$\nu^p(t) = \beta \|\mathbf{A}\|_F^2 \nu^x(t)/M + (1 - \beta) \nu^p(t - 1) = \alpha(t) \text{ and line (S6) becomes } \underline{\nu}^r(t) = (\|\mathbf{A}\|_F^2 \nu^s(t)/N)^{-1}.$$

As before, the observations took the form $\mathbf{y} = |\mathbf{A}\mathbf{x} + \mathbf{w}|$, but now the noise variance was adjusted to yield a nontrivial $\text{SNR} = 30$ dB. To demonstrate *compressive* phase retrieval, only $M = N = 65536$ intensity measurements were used. Running PR-GAMP on 100 problem realizations (each with different random \mathbf{A} and \mathbf{w} , and allowing at most 10 restarts per realization), we observed $\text{NMSE} < -36$ dB for all 100 realizations and a median runtime of only **5.9 seconds**. The right subplot in Fig. 10 shows a typical PR-GAMP recovery.

For the second set of experiments, we “blurred” the masked-Fourier outputs to further randomize \mathbf{A} , which allowed us to achieve similar recovery performance using *half* the intensity measurements, i.e., $M = \frac{N}{2} = 32768$. In particular, we used a linear transformation $\mathbf{A} \in \mathbb{C}^{M \times N}$ of the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{B}_1 \mathbf{F} \mathbf{D}_1 \\ \mathbf{B}_2 \mathbf{F} \mathbf{D}_2 \end{bmatrix}, \quad (22)$$

where \mathbf{F} and \mathbf{D}_i were as before¹¹ and \mathbf{B}_i were banded¹² matrices of size $\frac{M}{2} \times N$ with 10 nonzero i.i.d circular-Gaussian entries per column. The use of blurring to enhance phase retrieval was discussed in [51]. As with (21), the \mathbf{A} in (22) has a fast implementation. Running PR-GAMP as before on 100 problem realizations at $\text{SNR} = 30$ dB, we observed $\text{NMSE} < -28$ dB for all 100 realizations and a median runtime of only **7.3 seconds**.

To our knowledge, no existing algorithms are able to perform compressive phase retrieval on images of this size and sparsity with such high speed and accuracy. To put our results in perspective, we recall the image recovery experiment in [28], which shows an example of GESPAR taking 80 seconds to recover a $K = 15$ -sparse image whose support was effectively constrained to $N = 225$ pixels from $M = 38025$ 2D Fourier intensity measurements. In contrast, Fig. 10 shows PR-GAMP taking 1.8 seconds to recover a $K = 6678$ -sparse image whose support was constrained to $N = 65536$ pixels from $M = 65536$ masked 2D Fourier intensity measurements.

¹¹Here, since we used only two masks, we ensured invertibility by constructing the diagonal of \mathbf{D}_1 using exactly $N/2$ unit-valued entries positioned uniformly at random and constructing the diagonal of \mathbf{D}_2 as its complement, so that $\mathbf{D}_1 + \mathbf{D}_2 = \mathbf{I}$.

¹²Since each \mathbf{B}_i was a wide matrix, its nonzero band was wrapped from bottom to top when necessary.

V. CONCLUSIONS

In this paper, we proposed a novel approach to compressive phase retrieval based on the generalized approximate message passing (GAMP) algorithm. Numerical results showed that the proposed PR-GAMP algorithm has excellent phase transition behavior, noise robustness, and runtime. In particular, for successful recovery of synthetic K -sparse signals PR-GAMP requires approximately 4 times the number of measurements as phase-oracle GAMP and achieves NMSE that is only 3 dB worse than phase-oracle GAMP. For recovery of a real-valued 65532-pixel image from 32768 pre-masked and post-blurred Fourier intensities, PR-GAMP returned NMSE < -28 dB for all 100 realizations and a median runtime of only 7.3 seconds. An extensive numerical comparison to the recently proposed CPRL, sparse-Fienup, and GESPAR algorithms suggests that PR-GAMP has superior phase transitions and orders-of-magnitude faster runtimes at large K .

APPENDIX A OUTPUT THRESHOLDING RULES

In this appendix, we derive the expressions (10) and (12) that are used to compute the functions $g_{\text{out},m}$ and $g'_{\text{out},m}$ defined in lines (D2) and (D3) of Table I.

To facilitate the derivations in this appendix,¹³ we first rewrite $p_{Y|Z}(y|z)$ in a form different from (8). In particular, recalling that—under our AWGN assumption—the noisy transform outputs $u = z + w$ are conditionally distributed as $p(u|z) = \mathcal{N}(u; z, \nu^w)$, we first transform $u = ye^{j\theta}$ from rectangular to polar coordinates to obtain

$$p(y, \theta|z) = 1_{y \geq 0} 1_{\theta \in [0, 2\pi)} \mathcal{N}(ye^{j\theta}; z, \nu^w) y \quad (23)$$

where y is the Jacobian of the transformation, and then integrate out the unobserved phase θ to obtain

$$p_{Y|Z}(y|z) = 1_{y \geq 0} y \int_0^{2\pi} \mathcal{N}(ye^{j\theta}; z, \nu^w) d\theta, \quad (24)$$

We begin by deriving the scaling factor

$$\begin{aligned} C(y, \nu^w, \hat{p}, \nu^p) & \triangleq \int_{\mathbb{C}} p_{Y|Z}(y|z) \mathcal{N}(z; \hat{p}, \nu^p) dz \\ & = y 1_{y \geq 0} \int_0^{2\pi} \int_{\mathbb{C}} \mathcal{N}(ye^{j\theta}; z, \nu^w) \mathcal{N}(z; \hat{p}, \nu^p) dz d\theta \quad (25) \\ & = y 1_{y \geq 0} \int_0^{2\pi} \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta, \quad (26) \end{aligned}$$

where we used the Gaussian-pdf multiplication rule¹⁴ in (26). Noting the similarity between (26) and (24), the equivalence between (24) and (8) implies that

$$\begin{aligned} C(y, \nu^w, \hat{p}, \nu^p) & = \frac{2y}{\nu^w + \nu^p} \exp\left(-\frac{y^2 + |\hat{p}|^2}{\nu^w + \nu^p}\right) I_0\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p}\right) 1_{y \geq 0}. \quad (27) \end{aligned}$$

¹³The subscript “ m ” is omitted throughout this appendix for brevity.

¹⁴ $\mathcal{N}(z; a, A)\mathcal{N}(z; b, B) = \mathcal{N}\left(z; \frac{\frac{a}{A} + \frac{b}{B}}{\frac{1}{A} + \frac{1}{B}}, \frac{1}{\frac{1}{A} + \frac{1}{B}}\right)\mathcal{N}(a; b, A + B)$.

In the sequel, we make the practical assumption that $y > 0$, allowing us to drop the indicator “ $1_{y \geq 0}$ ” and invert C .

Next, we derive the conditional mean

$$\begin{aligned} E_{Z|Y,P}\{Z|y, \hat{p}; \nu^p\} & = C(y, \nu^w, \hat{p}, \nu^p)^{-1} \int_{\mathbb{C}} z p_{Y|Z}(y|z; \nu^w) \mathcal{N}(z; \hat{p}, \nu^p) dz. \quad (28) \end{aligned}$$

Plugging (24) into (28) and applying the Gaussian-pdf multiplication rule,

$$\begin{aligned} E_{Z|Y,P}\{Z|y, \hat{p}; \nu^p\} & = C^{-1} y \int_0^{2\pi} \int_{\mathbb{C}} z \mathcal{N}(z; ye^{j\theta}, \nu^w) \mathcal{N}(z; \hat{p}, \nu^p) dz d\theta \quad (29) \\ & = C^{-1} y \int_0^{2\pi} \int_{\mathbb{C}} z \mathcal{N}\left(z; \frac{ye^{j\theta}/\nu^w + \hat{p}/\nu^p}{1/\nu^w + 1/\nu^p}, \frac{1}{1/\nu^w + 1/\nu^p}\right) \\ & \quad \times \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) dz d\theta \quad (30) \end{aligned}$$

$$\begin{aligned} & = C^{-1} y \int_0^{2\pi} \frac{ye^{j\theta}/\nu^w + \hat{p}/\nu^p}{1/\nu^w + 1/\nu^p} \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta \quad (31) \\ & = \frac{y/\nu^w}{1/\nu^w + 1/\nu^p} C^{-1} y \int_0^{2\pi} e^{j\theta} \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta \end{aligned}$$

$$\begin{aligned} & \quad + \frac{\hat{p}/\nu^p}{1/\nu^w + 1/\nu^p} C^{-1} y \int_0^{2\pi} \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta \quad (32) \end{aligned}$$

$$\begin{aligned} & = \frac{y}{\nu^w/\nu^p + 1} C^{-1} y \int_0^{2\pi} e^{j\theta} \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta \\ & \quad + \frac{\hat{p}}{\nu^p/\nu^w + 1}. \quad (33) \end{aligned}$$

Expanding the \mathcal{N} term, the integral in (33) becomes

$$\begin{aligned} & \int_0^{2\pi} e^{j\theta} \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta \\ & = \frac{1}{\pi(\nu^w + \nu^p)} \exp\left(-\frac{y^2 + |\hat{p}|^2}{\nu^w + \nu^p}\right) \\ & \quad \times \int_0^{2\pi} e^{j\theta} \exp\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p} \cos(\theta - \psi)\right) d\theta \quad (34) \end{aligned}$$

$$\begin{aligned} & = \frac{1}{\pi(\nu^w + \nu^p)} \exp\left(-\frac{y^2 + |\hat{p}|^2}{\nu^w + \nu^p}\right) \\ & \quad \times e^{j\psi} \int_0^{2\pi} e^{j\theta'} \exp\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p} \cos(\theta')\right) d\theta' \quad (35) \end{aligned}$$

$$= \frac{2e^{j\psi}}{\nu^w + \nu^p} \exp\left(-\frac{y^2 + |\hat{p}|^2}{\nu^w + \nu^p}\right) I_1\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p}\right) \quad (36)$$

where ψ denotes the phase of \hat{p} , and where the integral in (35) was resolved using the expression in [52, 9.6.19]. Plugging (36) into (33) gives

$$E_{Z|Y,P}\{Z|y, \hat{p}; \nu^p\} = \frac{\hat{p}}{\nu^p/\nu^w + 1} + \frac{ye^{j\psi}}{\nu^w/\nu^p + 1} \frac{I_1\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p}\right)}{I_0\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p}\right)}, \quad (37)$$

which agrees with (10).

Finally, we derive the conditional covariance

$$\begin{aligned} \text{var}_{Z|Y,P}\{Z|y, \hat{p}; \nu^p\} \\ = C(y, \nu^w, \hat{p}, \nu^p)^{-1} \int_{\mathbb{C}} |z|^2 p_{Y|Z}(y|z; \nu^w) \mathcal{N}(z; \hat{p}, \nu^p) dz \\ - |\mathbb{E}_{Z|Y,P}\{Z|y, \hat{p}; \nu^p\}|^2. \end{aligned} \quad (38)$$

Focusing on the first term in (38), if we plug in (24) and apply the Gaussian-pdf multiplication rule, we get

$$\begin{aligned} C(y, \nu^w, \hat{p}, \nu^p)^{-1} \int_{\mathbb{C}} |z|^2 p_{Y|Z}(y|z; \nu^w) \mathcal{N}(z; \hat{p}, \nu^p) dz \\ = C^{-1} y \\ \times \int_0^{2\pi} \int_{\mathbb{C}} |z|^2 \mathcal{N}\left(z; \frac{ye^{j\theta}/\nu^w + \hat{p}/\nu^p}{1/\nu^w + 1/\nu^p}, \frac{1}{1/\nu^w + 1/\nu^p}\right) dz \\ \times \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta \end{aligned} \quad (39)$$

$$\begin{aligned} = C^{-1} y \int_0^{2\pi} \left(\left| \frac{ye^{j\theta}/\nu^w + \hat{p}/\nu^p}{1/\nu^w + 1/\nu^p} \right|^2 + \frac{1}{1/\nu^w + 1/\nu^p} \right) \\ \times \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta \end{aligned} \quad (40)$$

$$\begin{aligned} = C^{-1} y \\ \times \int_0^{2\pi} \frac{|y|^2/(\nu^w)^2 + |\hat{p}|^2/(\nu^p)^2 + 2y|\hat{p}|/(\nu^w \nu^p) \text{Re}\{e^{j(\theta-\psi)}\}}{(1/\nu^w + 1/\nu^p)^2} \\ \times \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta + \frac{1}{1/\nu^w + 1/\nu^p} \end{aligned} \quad (41)$$

$$\begin{aligned} = \frac{|y|^2/(\nu^w)^2 + |\hat{p}|^2/(\nu^p)^2}{(1/\nu^w + 1/\nu^p)^2} + \frac{1}{1/\nu^w + 1/\nu^p} + \frac{2y|\hat{p}|/(\nu^w \nu^p)}{(1/\nu^w + 1/\nu^p)^2} \\ \times C^{-1} y \text{Re} \left\{ e^{-j\psi} \int_0^{2\pi} e^{j\theta} \mathcal{N}(ye^{j\theta}; \hat{p}, \nu^w + \nu^p) d\theta \right\} \end{aligned} \quad (42)$$

$$\begin{aligned} = \frac{|y|^2/(\nu^w)^2 + |\hat{p}|^2/(\nu^p)^2}{(1/\nu^w + 1/\nu^p)^2} + \frac{1}{1/\nu^w + 1/\nu^p} + \frac{2y|\hat{p}|/(\nu^w \nu^p)}{(1/\nu^w + 1/\nu^p)^2} \\ \times C^{-1} y \frac{2}{\nu^w + \nu^p} \exp\left(-\frac{y^2 + |\hat{p}|^2}{\nu^w + \nu^p}\right) I_1\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p}\right) \end{aligned} \quad (43)$$

$$\begin{aligned} = \frac{|y|^2/(\nu^w)^2 + |\hat{p}|^2/(\nu^p)^2}{(1/\nu^w + 1/\nu^p)^2} + \frac{1}{1/\nu^w + 1/\nu^p} + \frac{2y|\hat{p}|/(\nu^w \nu^p)}{(1/\nu^w + 1/\nu^p)^2} \\ \times I_1\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p}\right) / I_0\left(\frac{2y|\hat{p}|}{\nu^w + \nu^p}\right) \end{aligned} \quad (44)$$

where (43) used (36) and (44) used (27). By plugging (44) back into (38), we obtain the expression given in (12).

APPENDIX B POST-INTENSITY NOISE MODELS

In this appendix, we consider the $g_{\text{out},m}$ and $g'_{\text{out},m}$ functions (defined in lines (D2) and (D3) of Table I) for the post-intensity noise model (14) under generic $q(\cdot)$ and p_W .

Following the procedure in Appendix A, we begin by examining the scaling factor

$$\begin{aligned} C(y, \hat{p}, \nu^p) \\ \triangleq \int_{\mathbb{C}} p_{Y|Z}(y|z) \mathcal{N}(z; \hat{p}, \nu^p) dz \\ = \int_0^\infty p_W(y - q(r)) \left(\int_0^{2\pi} \mathcal{N}(re^{j\phi}; \hat{p}, \nu^p) d\phi \right) r dr \\ = \frac{2}{\nu^p} \int_0^\infty r p_W(y - q(r)) \exp\left(-\frac{r^2 + |\hat{p}|^2}{\nu^p}\right) I_0\left(\frac{2r|\hat{p}|}{\nu^p}\right) dr, \end{aligned} \quad (45)$$

where for (45) we used the rectangular-to-polar transformation $z = re^{j\phi}$ with $r \geq 0$ and $\phi \in [0, 2\pi)$, and for (46) we used the Rician result (8).

Next we examine the conditional mean defined in (28). Plugging (15) into (28) and transforming from rectangular to polar coordinates, we get

$$\begin{aligned} \mathbb{E}_{Z|Y,P}\{Z|y, \hat{p}; \nu^p\} \\ = \frac{1}{C} \int_0^\infty r p_W(y - q(r)) \left(\int_0^{2\pi} e^{j\phi} \mathcal{N}(re^{j\phi}; \hat{p}, \nu^p) d\phi \right) r dr \end{aligned} \quad (47)$$

$$= \frac{2e^{j\psi}}{C\nu^p} \int_0^\infty r^2 p_W(y - q(r)) \exp\left(-\frac{r^2 + |\hat{p}|^2}{\nu^p}\right) I_1\left(\frac{2r|\hat{p}|}{\nu^p}\right) dr, \quad (48)$$

where (48) used the result from (36).

Finally we examine the conditional covariance (38), and in particular the first term in (38), which now becomes

$$\begin{aligned} \frac{1}{C} \int_{\mathbb{C}} |z|^2 p_{Y|Z}(y|z) \mathcal{N}(z; \hat{p}, \nu^p) dz \\ = \frac{1}{C} \int_0^\infty r^2 p_W(y - q(r)) \left(\int_0^{2\pi} \mathcal{N}(re^{j\phi}; \hat{p}, \nu^p) d\phi \right) r dr \\ = \frac{2}{C\nu^p} \int_0^\infty r^3 p_W(y - q(r)) \exp\left(-\frac{r^2 + |\hat{p}|^2}{\nu^p}\right) I_0\left(\frac{2r|\hat{p}|}{\nu^p}\right) dr, \end{aligned} \quad (49)$$

where (50) used a computation similar to (46). Further simplification of the above expressions requires specification of $q(\cdot)$ and p_W .

APPENDIX C EM UPDATE FOR NOISE VARIANCE

In this appendix, we derive the EM update (17) of the noise variance. Our approach is based on the use of GAMP's posterior approximation $b_{\mathbf{X}}(\mathbf{x})$ in place of the true posterior distribution $p_{\mathbf{X}|Y}(\mathbf{x}|\mathbf{y})$ in (16). At GAMP iteration t , $b_{\mathbf{X}}(\mathbf{x}) = \prod_{n=1}^N b_{X_n}(x_n)$ for

$$b_{X_n}(x) \triangleq p_{X_n}(x) \mathcal{N}(x; \hat{r}_n(t), \nu_n^r(t)) B_n^{-1} \quad (51)$$

$$B_n \triangleq \int_{\mathbb{C}} p_{X_n}(x) \mathcal{N}(x; \hat{r}_n(t), \nu_n^r(t)) dx, \quad (52)$$

which also appears in line (D4) of Table I.

Under the posterior approximation $b_{\mathbf{X}}$ and large i.i.d \mathbf{A} , the negative log likelihood $-\ln p(\mathbf{y}; \nu^w)$ is well approximated by the variational Bethe free energy¹⁵[53], [54]

$$\begin{aligned} J(\nu^w; \hat{\mathbf{r}}(t), \nu^r(t)) \\ = D(b_{\mathbf{X}} \| p_{\mathbf{X}}) + D(b_{\mathbf{Z}} \| p_{Y|Z}(\mathbf{y}|\cdot; \nu^w)) \Gamma^{-1} \\ + \sum_{m=1}^M \left(\frac{\text{var}\{Z_m | b_{Z_m}\}}{\nu_m^p} + \ln(\pi \nu_m^p) \right). \end{aligned} \quad (53)$$

In (53), the first term measures the KL divergence of the prior $p_{\mathbf{X}}(\mathbf{x}) \triangleq \prod_{n=1}^N p_{X_n}(x_n)$ from the approximated posterior $b_{\mathbf{X}}(\mathbf{x})$. The second term then measures the KL divergence of the

¹⁵Note that the Bethe free energy expressions in this paper are stated for the complex-valued case.

pdf $p_{Y|Z}(\mathbf{y}|z; \nu^w)\Gamma^{-1}$ from $b_Z(z)$, the GAMP-approximated posterior pdf on z . Here, the scaling factor $\Gamma \triangleq \prod_{m=1}^M \Gamma_m$, for

$$\Gamma_m \triangleq \int_{\mathbb{C}} p_{Y|Z}(y_m|z; \nu^w) dz, \quad (54)$$

ensures that $p_{Y|Z}(\mathbf{y}|z; \nu^w)\Gamma^{-1}$ is a valid pdf over $z \in \mathbb{C}^M$, and the approximate posterior takes the form $b_Z(z) = \prod_{m=1}^M b_{Z_m}(z_m)$ for

$$b_{Z_m}(z) \triangleq p_{Y|Z}(y_m|z; \nu^w) \mathcal{N}(z; \bar{p}_m, \nu_m^p) C_m^{-1} \quad (55)$$

$$C_m \triangleq \int_{\mathbb{C}} p_{Y|Z}(y_m|z; \nu^w) \mathcal{N}(z; \bar{p}_m, \nu_m^p) dz, \quad (56)$$

which also appears in line (D1) of Table I. Above, $(\bar{\mathbf{p}}, \nu^p)$ are “fixed point” values that are consistent with $(\hat{\mathbf{r}}(t), \nu^r(t))$ in the sense that

$$\nu_m^p = \sum_{n=1}^N |a_{mn}|^2 \underbrace{\text{var}\{X_n | b_{X_n}(\cdot; \hat{r}_n(t), \nu_n^r(t))\}}_{= \nu_n^x \text{ from (R7)}} \quad (57)$$

$$\begin{aligned} \hat{z}_m &\triangleq \mathbb{E}\{Z_m | b_{Z_m}(\cdot; \bar{p}_m, \nu_m^p)\} \\ &= \sum_{n=1}^N a_{mn} \underbrace{\mathbb{E}\{X_n | b_{X_n}(\cdot; \hat{r}_n(t), \nu_n^r(t))\}}_{= \hat{x}_n \text{ from (R8)}}. \end{aligned} \quad (58)$$

Whereas ν_m^p in (57) can be computed directly from $(\hat{\mathbf{r}}(t), \nu^r(t))$, finding the \bar{p}_m that solves (58) may require numerical search, e.g., via Newton’s method [49].

Plugging (55) into the second term of (53) reveals

$$\begin{aligned} D(b_Z \| p_{Y|Z}(\mathbf{y}|\cdot; \nu^w)\Gamma^{-1}) \\ = \int_{\mathbb{C}^M} b_Z(z) \ln \frac{b_Z(z)}{p_{Y|Z}(\mathbf{y}|z)\Gamma^{-1}} dz \end{aligned} \quad (59)$$

$$= \sum_{m=1}^M \int_{\mathbb{C}} b_{Z_m}(z_m) \ln \frac{b_{Z_m}(z_m)}{p_{Y|Z}(y_m|z_m)\Gamma_m^{-1}} dz_m \quad (60)$$

$$= \sum_{m=1}^M \int_{\mathbb{C}} b_{Z_m}(z) \ln \frac{\mathcal{N}(z; \bar{p}_m(t), \nu_m^p(t)) C_m^{-1}}{\Gamma_m^{-1}} dz \quad (61)$$

$$= \sum_{m=1}^M \left(\ln \frac{\Gamma_m}{C_m} - \ln(\pi \nu_m^p) - \frac{\nu_m^z + |\hat{z}_m - \bar{p}_m|^2}{\nu_m^p} \right), \quad (62)$$

using the shorthand notation $\nu_m^z \triangleq \text{var}\{Z_m | b_{Z_m}\}$. Then plugging (62) into (53) and canceling terms reveals

$$\begin{aligned} J(\nu^w; \hat{\mathbf{r}}(t), \nu^r(t)) \\ = D(b_{\mathbf{X}} \| p_{\mathbf{X}}) - \sum_{m=1}^M \left(\ln \frac{C_m(\nu^w)}{\Gamma_m(\nu^w)} + \frac{|\hat{z}_m - \bar{p}_m(\nu^w)|^2}{\nu_m^p} \right), \end{aligned} \quad (63)$$

where the ν^w dependence of Γ_m , C_m , and \bar{p}_m is made explicit. Note that \hat{z}_m and ν_m^p are completely determined by $(\hat{\mathbf{r}}(t), \nu^r(t))$ via (57)-(58), and thus invariant to ν^w , and $D(b_{\mathbf{X}} \| p_{\mathbf{X}})$ is by definition invariant to ν^w . When $p_{Y|Z}$ is Gaussian, the value of $\bar{p}_m(\nu^w)$ can be computed in closed form [40] after which the resulting expression (63) simplifies.

For non-Gaussian $p_{Y|Z}$, we propose the following EM update procedure. For simplicity, we will assume that one EM update is performed per GAMP iteration, allowing us to write the EM iterations “[i]” as GAMP iterations “[t]”. Recalling (16), we first run GAMP with $\widehat{\nu}^w(t)$ to produce $(\hat{\mathbf{r}}(t), \nu^r(t))$, the approximate posterior $b_{\mathbf{X}}(\mathbf{x})$ in (51), the corresponding $(\bar{\mathbf{p}}(\widehat{\nu}^w(t)), \nu^p)$ from (57)-(58), and finally the approximation of $-\mathbb{E}\{\ln p(\mathbf{y}, \mathbf{x}; \nu^w) | \mathbf{y}; \widehat{\nu}^w(t)\}$ in (63). However, to facilitate the minimization over ν^w , we use

$$\begin{aligned} \tilde{J}(\nu^w; \hat{\mathbf{r}}(t), \nu^r(t), \widehat{\nu}^w(t)) \\ \triangleq D(b_{\mathbf{X}} \| p_{\mathbf{X}}) - \sum_{m=1}^M \left(\ln \frac{C_m(\nu^w)}{\Gamma_m(\nu^w)} + \frac{|\hat{z}_m - \bar{p}_m(\widehat{\nu}^w(t))|^2}{\nu_m^p} \right) \end{aligned} \quad (64)$$

in place of (63), noting that the substitution of $\bar{p}_m(\nu^w)$ by $\bar{p}_m(\widehat{\nu}^w(t))$ preserves the fixed point(s) of the EM procedure. Finally, we assign the value of ν^w that minimizes (64) to $\widehat{\nu}^w(t+1)$. The overall procedure is summarized by (17).

For the $p_{Y|Z}$ in (8) used for PR-GAMP, it can be shown that $\Gamma_m(\nu^w)$ is invariant to ν^w . Thus, $\widehat{\nu}^w(t+1) = \arg \max_{\nu^w} \sum_{m=1}^M \ln C_m(\nu^w)$ for the $C_m(\nu^w)$ given in (27). We numerically compute the maximizing value.

REFERENCES

- [1] P. Schniter and S. Rangan, “Compressive phase retrieval via generalized approximate message passing,” in *Proc. Allerton Conf. Commun. Contr. Comput.*, Monticello, IL, USA, Oct. 2012, pp. 815–822.
- [2] P. Schniter, “Compressive phase retrieval via generalized approximate message passing,” in *Feb. Fourier Talks (FFT) Workshop on Phaseless Reconstruc.*, College Park, MD, USA, Feb. 2013.
- [3] J. R. Fienup, “Phase retrieval algorithms: A comparison,” *Appl. Opt.*, vol. 21, pp. 2758–2769, Aug. 1982.
- [4] R. P. Millane, “Recent advances in phase retrieval,” *Int. Soc. Opt. Eng.*, vol. 6316, 2006.
- [5] O. Bunk, A. Diaz, F. Pfeiffer, C. David, B. Schmitt, D. K. Satapathy, and J. F. Veen, “Diffractive imaging for periodic samples: Retrieving one-dimensional concentration profiles across microfluidic channels,” *Acta Crystallograph. Sect. A: Found. Crystallogr.*, vol. 63, no. 4, pp. 306–314, 2007.
- [6] R. W. Harrison, “Phase problem in crystallography,” *J. Opt. Soc. Amer. A*, vol. 10, no. 5, pp. 1046–1055, 1993.
- [7] R. P. Millane, “Phase retrieval in crystallography and optics,” *J. Opt. Soc. Amer. A*, vol. 7, pp. 394–411, Mar. 1990.
- [8] A. Chai, M. Moscoso, and G. Papanicolaou, “Array imaging using intensity-only measurements,” *Inverse Probl.*, vol. 27, no. 1, pp. 1–16, 2011.
- [9] A. Walther, “The question of phase retrieval in optics,” *Opt. Acta*, vol. 10, no. 1, pp. 41–49, 1963.
- [10] J. C. Dainty and J. R. Fienup, “Phase retrieval and image construction for astronomy,” in *Image Recovery: Theory and Application*, H. Stark, Ed. New York, USA: Academic, 1987, ch. 7, pp. 231–275.
- [11] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, “Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes,” *Ann. Rev. Phys. Chem.*, vol. 59, pp. 387–410, 2008.
- [12] R. Balan, P. G. Casazza, and D. Edidin, “On signal reconstruction without phase,” *Appl. Computat. Harmon. Anal.*, vol. 20, pp. 345–356, May 2006.
- [13] L. Demanet and V. Jugnon, “Convex recovery from interferometric measurements,” 2013, arXiv:1307.6864 [Online]. Available: <http://arxiv.org/abs/1307.6864>
- [14] J. V. Corbett, “The Pauli problem, state reconstruction and quantum-real numbers,” *Rep. Math. Phys.*, vol. 57, no. 1, 2006.
- [15] T. Heinosaari, L. Mazzarella, and M. M. Wolf, “Quantum tomography under prior information,” *Commun. Math. Phys.*, vol. 318, no. 2, pp. 355–374, 2013.
- [16] B. G. Bodmann and N. Hammen, “Stable phase retrieval with low-redundancy frames,” *Adv. Comput. Math.*, pp. 1–15, 2014.

- [17] S. Gazit, A. Szameit, Y. C. Eldar, and M. Segev, "Super-resolution and reconstruction of sparse sub-wavelength images," *Opt. Expr.*, vol. 17, no. 26, pp. 23920–23946, 2009.
- [18] A. Szameit, Y. Shechtman, E. Osherovich, E. Bullklich, P. Sidorenko, H. Dana, S. Steiner, E. B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y. C. Eldar, O. Cohen, and M. Segev, "Sparsity-based single-shot subwavelength coherent diffractive imaging," *Nature Mater.*, vol. 11, pp. 455–459, 2012.
- [19] S. Marchesini, "Ab initio compressive phase retrieval," 2008, arXiv:0809.2006 [Online]. Available: <http://arxiv.org/abs/0809.2006>
- [20] Y. Shechtman, E. Small, Y. Lahini, M. Verbin, Y. Eldar, Y. Silberberg, and M. Segev, "Sparsity-based super-resolution and phase-retrieval in waveguide arrays," *Opt. Express*, vol. 21, no. 20, pp. 24015–24024, 2013.
- [21] X. Li and V. Voroninski, "Sparse signal recovery from quadratic measurements via convex programming," *SIAM J. Math. Anal.*, vol. 45, no. 5, pp. 3019–3033, 2013.
- [22] M. L. Moravec, J. K. Romberg, and R. Baraniuk, "Compressive phase retrieval," in *Proc. SPIE Conf. Ser.*, San Diego, CA, Aug. 2007, vol. 6701.
- [23] S. Mukherjee and C. S. Seelamantula, "Fienup algorithm with sparsity constraints: Application to frequency-domain optical-coherence tomography," *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4659–4672, 2014.
- [24] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, "CPRL – An extension of compressive sensing to the phase retrieval problem," in *Proc. Neural Inf. Process. Syst. Conf.*, 2012.
- [25] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.
- [26] Y. Chen, Y. Chi, and A. J. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," 2013, arXiv:1310.0807 [Online]. Available: <http://arxiv.org/abs/1310.0807>
- [27] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Proc. Neural Inf. Process. Syst. Conf.*, 2013.
- [28] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," *IEEE Trans. Signal Process.*, vol. 62, pp. 928–938, Feb. 2014.
- [29] C. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. New York, USA: Dover, 1998.
- [30] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Aug. 2011, pp. 2168–2172.
- [31] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, pp. 18914–18919, Nov. 2009.
- [32] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," in *Proc. Inf. Theory Workshop*, Cairo, Egypt, Jan. 2010, pp. 1–5.
- [33] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA, USA: Morgan Kaufman, 1988.
- [34] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, pp. 498–519, Feb. 2001.
- [35] B. J. Frey and D. J. C. MacKay, "A revolution: Belief propagation in graphs with cycles," in *Proc. Neural Inf. Process. Syst. Conf.*, Denver, CO, 1997, pp. 479–485.
- [36] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, pp. 764–785, Feb. 2011.
- [37] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and iterative algorithms," in *Proc. IEEE Int. Symp. Inf. Thy.*, Boston, MA, USA, Jun. 2012, pp. 1643–1647.
- [38] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 24, no. 1, pp. 46–156, 1945.
- [39] A. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, vol. 39, pp. 1–17, 1977.
- [40] 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.*, vol. P08009, 2012.
- [41] G. F. Cooper, "The computational complexity of probabilistic inference using Bayesian belief networks," *Artif. Intell.*, vol. 42, pp. 393–405, 1990.
- [42] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, USA, Mar. 2010, pp. 1–6.
- [43] M. Borgerding and P. Schniter, "Generalized approximate message passing for the cosparsity analysis model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2014, accepted for publication.
- [44] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, pp. 4658–4672, Oct. 2013.
- [45] J. P. Vila and P. Schniter, "An empirical-Bayes approach to recovering linearly constrained non-negative sparse signals," *IEEE Trans. Signal Process.*, vol. 62, pp. 4689–4703, Sep. 2014.
- [46] S. Som, L. C. Potter, and P. Schniter, "On approximate message passing for reconstruction of non-uniformly sparse signals," in *Proc. Nat. Aerosp. Electron. Conf.*, Dayton, OH, USA, Jul. 2010.
- [47] T. Heskes, "Stable fixed points of loopy belief propagation are minima of the Bethe free energy," in *Proc. Neural Inf. Process. Syst. Conf.*, Vancouver, BC, Dec. 2002, pp. 343–350.
- [48] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of generalized approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2014, pp. 236–240.
- [49] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2014, accepted for publication.
- [50] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," 2013, arXiv:1310.3240 [Online]. Available: <http://arxiv.org/abs/1310.3240>
- [51] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nature Photon.*, vol. 7, pp. 739–745, 2013.
- [52] *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun, Eds. New York, NY: Dover, 1964.
- [53] S. Rangan, P. Schniter, E. Riegler, A. Fletcher, and V. Cevher, "Fixed points of generalized approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 664–668.
- [54] F. Krzakala, A. Manoel, E. W. Tramel, and L. Zdeborová, "Variational free energies for compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2014, pp. 1499–1503.



Philip Schniter (F'14) received the B.S. and M.S. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively, and the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, in 2000.

From 1993 to 1996 he was with Tektronix Inc., Beaverton, OR as a systems engineer. After receiving the Ph.D. degree, he joined the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, where he is currently a Professor and a member of the Information Processing Systems (IPS) Lab. During 2008–2009, he was a visiting professor at Eurecom, Sophia Antipolis, France, and Supélec, Gif-sur-Yvette, France. His areas of interest currently include statistical signal processing, wireless communications and networks, and machine learning.

Dr. Schniter received the National Science Foundation CAREER Award in 2003.



Sundeep Rangan (SM'14) received the B.A.Sc. degree from the University of Waterloo, Canada, and the M.Sc. and Ph.D. degrees at the University of California, Berkeley, all in electrical engineering.

He has held postdoctoral appointments at the University of Michigan, Ann Arbor, and Bell Labs. In 2000, he cofounded (with four others) Flarion Technologies, a spin off of Bell Labs that developed Flash OFDM, the first cellular OFDM data system and precursor to the current 4G LTE standard. In 2006, Flarion was acquired by Qualcomm Technologies. At Qualcomm, he was a Director of Engineering involved in OFDM infrastructure products. He joined the Electrical and Computer Engineering Department at the New York University Polytechnic School of Engineering in 2010. His research interests are in wireless communications, signal processing, information theory, and control theory.

Dr. Rangan is a Distinguished Lecturer of the IEEE Vehicular Technology Society.