Compressive Phase Retrieval via Generalized Approximate Message Passing

Philip Schniter
Dept. of ECE, The Ohio State University
Columbus, OH 43210
Email: schniter@ece.osu.edu

Sundeep Rangan
Dept. of ECE, Polytechnic Institute of New York University
Brooklyn, NY 11201
Email: srangan@poly.edu

***Abstract*—In this paper, we propose a novel approach to compressive phase retrieval based on loopy belief propagation and, in particular, on the generalized approximate message passing (GAMP) algorithm. Numerical results show that the proposed PR-GAMP algorithm has excellent phase-transition behavior, noise robustness, and runtime. In particular, for successful recovery of synthetic Bernoulli-circular-Gaussian signals, PR-GAMP requires ≈ 4 times the number of measurements as a phase-oracle version of GAMP and, at moderate to large SNR, the NMSE of PR-GAMP is only ≈ 3 dB worse than that of phase-oracle GAMP. A comparison to the recently proposed convex-relation approach known as "CPRL" reveals PR-GAMP's superior phase transition and orders-of-magnitude faster runtimes, especially as the problem dimensions increase. When applied to the recovery of a $65k$-pixel grayscale image from $32k$ randomly masked magnitude measurements, numerical results show a median PR-GAMP runtime of only 13.4 seconds.**

I. Introduction

A. Phase retrieval

Phase retrieval [1], i.e., reconstruction of a signal $\boldsymbol{x}_0 \in \mathbb{C}^n$ from the magnitudes $\boldsymbol{y} = |\boldsymbol{z}|$ of the linear (often Fourier-based) transformation $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}_0 \in \mathbb{C}^m$, is a problem of great importance in many fields, e.g., optics [2], X-ray crystallography [3], astronomy [4], speech [5], and many others. Closely connected to the problem of phase retrieval is the problem of *spectral factorization* [6].

The phase retrieval problem is challenging due to both the nonlinearity of the mapping between $\boldsymbol{x}_0$ and $\boldsymbol{y}$, as well as the *non-unicity* implied by this mapping. For example, if $\boldsymbol{x} = \boldsymbol{x}_0$ yields a given $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|$, then $\boldsymbol{x} = c\boldsymbol{x}_0$ also yields the same magnitudes $\boldsymbol{y}$ for any unit-modulus $c \in \mathbb{C}$. Although such "global phase ambiguity" is tolerable in most applications of phase retrieval, more problematic ambiguities can result under particular instances of $\boldsymbol{A}$. For example, if $\boldsymbol{A}$ is a discrete Fourier transform (DFT) matrix, then all cyclic-shifts of a given $\boldsymbol{x}_0$ will yield identical Fourier magnitudes $\boldsymbol{y}$, as will conjugate flips of $\boldsymbol{x}_0$. But, besides these variants, other $\boldsymbol{x} \neq \boldsymbol{x}_0$ can also yield the same $\boldsymbol{y}$ in the 1D case. In the 2D case, however, it has been shown [7] that real-valued $\boldsymbol{x}_0$ can be uniquely recovered (up to global phase, shift, and flip) from twice-oversampled DFT magnitudes, with probability one. Thus, if the signal support is apriori known, and constructed such that only one shift and flip is feasible, the signal can be recovered up to a global phase.

Even when uniqueness (up to a global phase) is guaranteed, there remains the difficult question of how to recover the signal using a practical algorithm, and whether that algorithm is robust to the presence of measurement noise. Most classical approaches to phase retrieval (e.g., those of Gerchberg and Saxton [8] and Fienup [1]) are iterative and based on alternating projections. For example, it is common to alternate between the signal projection of the signal estimate $\hat{\boldsymbol{x}}$ onto a known support and the projection of its transform magnitudes $\hat{\boldsymbol{y}} = |\boldsymbol{A}\hat{\boldsymbol{x}}|$ onto the observations $\boldsymbol{y}$. Due to the non-convexity of the latter projection, however, such algorithms can easily get trapped in local minima [9].

Recently, convex relaxation techniques have been proposed to solve the phase retrieval problem [10], [11], [12], [13]. There, the first step is to notice that $y_i^2 = |\boldsymbol{a}_i^{\mathsf{H}}\boldsymbol{x}_0|^2 = \mathrm{tr}(\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{H}}\boldsymbol{X}_0)$ for rank-one positive-semidefinite $\boldsymbol{X}_0 \triangleq \boldsymbol{x}_0\boldsymbol{x}_0^{\mathsf{H}}$, where $\boldsymbol{a}_i^{\mathsf{H}}$ denotes the i^{th} row of $\boldsymbol{A}$. Thus, phase retrieval can be expressed as the optimization problem "$\min_{\boldsymbol{X}\succeq 0}\mathrm{rank}(\boldsymbol{X})$ s.t. $\mathrm{tr}(\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{H}}\boldsymbol{X}) = y_i^2$ for $i = 1, ..., m$." The approach known as "PhaseLift" [10], [12] relaxes the non-convex rank term to obtain the convex problem "$\min_{\boldsymbol{X}\succeq 0}\mathrm{tr}(\boldsymbol{X})$ s.t. $\mathrm{tr}(\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{H}}\boldsymbol{X}) = y_i^2$ for $i = 1, ..., m$," which can be recognized as a semidefinite program (SDP), and solved using standard techniques. It was then shown in [14] that, with very high probability, PhaseLift recovers (up to a global phase) an arbitrary $\boldsymbol{x}_0$ from $m \geq c_0 n$ measurements $\{y_i^2\}_{i=1}^m$, where c_0 is a sufficiently large constant, when $\boldsymbol{A}$ is i.i.d random. This result is encouraging because it has been shown that, for $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_0|$ to have a unique solution in $\mathbb{C}^n$ (with probability one and up to a global phase), $m = 3n-2$ measurements are necessary [15] and $m = 4n$–2 are sufficient [16]. Moreover, PhaseLift's recovery was shown [14] to be robust to additive noise on the squared magnitudes y_i^2. The "PhaseCut" approach proposed in [13] instead optimizes over the unknown phases of $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}$, yielding a different SDP, but with similar performance.

B. Compressive phase retrieval

There exist applications (such as Bragg sampling from periodic crystalline structures [17]) where it is inconvenient

Philip Schniter is supported in part by NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090.

or impossible to take $m \geq 3n-2$ magnitude measurements. The case that $m < 3n-2$ can thus be considered as "*compressive phase retrieval*," as coined in [18]. Due to the aforementioned problems of nonlinearity and nonunicity, however, the compressed phase retrieval problem is significantly more difficult than the standard "compressive sensing" problem [19], where the goal is to recover $\boldsymbol{x}_0 \in \mathbb{C}^n$ from $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}_0 \in \mathbb{C}^m$ (or a noisy version thereof) with $m < n$.

Since compressive phase retrieval is a relatively new (and challenging) problem, we are aware of only a few papers on the topic (e.g., [18], [17], [20]). Not surprisingly, they are all based on exploiting the additional structure that exists when the signal $\boldsymbol{x}$ is sparse[1] or compressible, i.e., the same structure exploited in standard compressive sensing. One of the first results was given by Moravec, Romberg, and Baraniuk in [18], where it was shown that, for a k-sparse n-length signal, $m = \mathcal{O}(k^2 \log 4n/k^2)$ random 2D Fourier measurements suffice for exact reconstruction. However, the algorithms proposed in these works leave considerable room for improvement. For example, the algorithm in [18] exploits knowledge of the ℓ_1 norm of the true signal $\boldsymbol{x}_0$, which is rarely (if ever) available in practice. The work by Marchesini [17], on the other hand, is based on heuristics that are not clearly generalizable outside of the applications that they considered. The work by Ohlsson, Yang, Dong, and Sastry [20] is cutting-edge in that it extends the convex PhaseLift algorithm to exploit sparsity, and performs RIP-based and mutual-coherence based analyses to derive sufficient conditions on the sparse-signal recovery performance of the proposed "CPRL" algorithm. CPRL's complexity, however, grows rapidly with the problem dimensions. For completeness, we note that Waldspurger, D'Aspremont, and Mallat mentioned, in [13], a heuristic modification by which sparsity can be leveraged within PhaseCut, but they did not investigate the performance of this approach.

In this paper, we propose a novel approach to compressive phase retrieval based on loopy belief propagation and, in particular, on the *generalized approximate message passing* (GAMP) algorithm [23].

II. Background on GAMP

The approximate message passing (AMP) algorithm was recently proposed by Donoho, Maleki, and Montanari [24], [25] for the task estimating a signal vector $\boldsymbol{x} \in \mathbb{R}^N$ from linearly compressed and additive-Gaussian-noise corrupted measurements[2]

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w} \in \mathbb{C}^m. \tag{1}$$

[1]We note that the works [21], [22] address the problem of recovering sparse signals from Fourier-transform magnitudes, but do not address measurement compression.

[2]Here and elsewhere, we use $\boldsymbol{y}$ when referring to the m measurements that are available for signal reconstruction. In the case of standard compressive sensing, we model these measurements as $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}$ with $\boldsymbol{w}$ being additive white Gaussian noise. In the case of (noisy) compressive phase retrieval, these measurements instead take the form of $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}|$.

The Generalized-AMP (GAMP) algorithm proposed by Rangan [23] then extends the methodology of AMP to the generalized linear measurement model

$$\boldsymbol{y} = q(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}) \in \mathbb{C}^m, \tag{2}$$

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

Both AMP and GAMP can be derived from the perspective of *belief propagation* [26], a Bayesian inference strategy that is based on a factorization of the signal posterior pdf $p(\boldsymbol{x} \,|\, \boldsymbol{y})$ into a product of simpler pdfs that, together, uncover the probabilistic structure in the problem. (Recall that the posterior pdf $p(\boldsymbol{x} \,|\, \boldsymbol{y})$ captures everything that can be learned about $\boldsymbol{x}$ from $\boldsymbol{y}$ under the assumed observation model.) Concretely, if we model the signal coefficients $\boldsymbol{x}$ and noise samples $\boldsymbol{w}$ in (1)-(2) as independent and identically distributed, so that $p(\boldsymbol{x}) = \prod_{j=1}^n p_X(x_j)$ and $p(\boldsymbol{y} \,|\, \boldsymbol{z}) = \prod_{i=1}^m p_{Y|Z}(y_i \,|\, z_i)$ for $\boldsymbol{z} \triangleq \boldsymbol{A}\boldsymbol{x}$, then we can factor the posterior pdf as

$$p(\boldsymbol{x} \,|\, \boldsymbol{y}) \propto p(\boldsymbol{y} \,|\, \boldsymbol{x}) p(\boldsymbol{x}) = \prod_{i=1}^m p_{Y|Z}\big(y_i \,\big|\, [\boldsymbol{A}\boldsymbol{x}]_i\big) \prod_{j=1}^n p_X(x_j), \tag{3}$$

yielding the factor graph in Fig. 1.

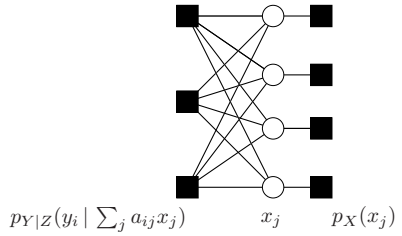

Fig. 1. Factor graph for compressive sensing, with white circles denoting random variables and black squares denoting pdf factors.

In belief propagation [26], 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 into a given variable node are then used to determine the posterior pdf of that variable, or an approximation thereof. The sum-product algorithm [27] 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 [28].

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 $\boldsymbol{A}$ is an appropriately constructed 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

definitions:

$$p_{Z|Y,P}(z|y,\hat{p};\mu^p) = \frac{p_{Y|Z}(y|z)\,\mathcal{CN}(z;\hat{p},\mu^p)}{\int_{z'} p_{Y|Z}(y|z')\,\mathcal{CN}(z';\hat{p},\mu^p)} \quad \text{(D1)}$$

$$g_{\text{out}}(y,\hat{p},\mu^p) = \tfrac{1}{\mu^p}\big(\mathrm{E}_{Z|Y,P}\{Z|y,\hat{p};\mu^p\} - \hat{p}\big) \quad \text{(D2)}$$

$$g'_{\text{out}}(y,\hat{p},\mu^p) = \tfrac{1}{\mu^p}\Big(\tfrac{\mathrm{var}_{Z|Y,P}\{Z|y,\hat{p};\mu^p\}}{\mu^p} - 1\Big) \quad \text{(D3)}$$

$$p_{X|R}(x|\hat{r};\mu^r) = \frac{p_X(x)\,\mathcal{CN}(x;\hat{r},\mu^r)}{\int_{x'} p_X(x')\,\mathcal{CN}(x';\hat{r},\mu^r)} \quad \text{(D4)}$$

$$g_{\text{in}}(\hat{r},\mu^r) = \mathrm{E}_{X|R}\{X|\hat{r};\mu^r\} \quad \text{(D5)}$$

$$g'_{\text{in}}(\hat{r},\mu^r) = \mathrm{var}_{X|R}\{X|\hat{r};\mu^r\} \quad \text{(D6)}$$

initialize:

$$\forall j: \hat{x}_j(1) = \text{user's choice, e.g., } \mathrm{E}\{X\} \quad \text{(I1)}$$

$$\forall j: \mu^x_j(1) = \text{user's choice, e.g., } \mathrm{var}\{X\} \quad \text{(I2)}$$

$$\forall i: \hat{s}_i(0) = 0 \quad \text{(I3)}$$

for $t = 1, 2, 3, \ldots$

$$\forall i: \mu^p_i(t) = \textstyle\sum_{j=1}^n |a_{ij}|^2 \mu^x_j(t) \quad \text{(R1)}$$

$$\forall i: \hat{p}_i(t) = \textstyle\sum_{j=1}^n a_{ij}\hat{x}_j(t) - \mu^p_i(t)\,\hat{s}_i(t-1) \quad \text{(R2)}$$

$$\forall i: \hat{s}_i(t) = g_{\text{out}}(y_i,\hat{p}_i(t),\mu^p_i(t)) \quad \text{(R3)}$$

$$\forall i: \mu^s_i(t) = -g'_{\text{out}}(y_i,\hat{p}_i(t),\mu^p_i(t)) \quad \text{(R4)}$$

$$\forall j: \mu^r_j(t) = \big(\textstyle\sum_{i=1}^m |a_{ij}|^2 \mu^s_i(t)\big)^{-1} \quad \text{(R5)}$$

$$\forall j: \hat{r}_j(t) = \hat{x}_j(t) + \mu^r_j(t) \textstyle\sum_{i=1}^m a^*_{ij}\hat{s}_i(t) \quad \text{(R6)}$$

$$\forall j: \mu^x_j(t+1) = \mu^r_j(t) g'_{\text{in}}(\hat{r}_j(t),\mu^r_j(t)) \quad \text{(R7)}$$

$$\forall j: \hat{x}_j(t+1) = g_{\text{in}}(\hat{r}_j(t),\mu^r_j(t)) \quad \text{(R8)}$$

end

TABLE I
THE GAMP ALGORITHM [23]

LBP-based compressive sensing is not only feasible [24] for *dense* matrices $\boldsymbol{A}$, but provably accurate [29], [30], can be recognized as quite a breakthrough. In particular, they established that, in the large system limit (i.e., as $m, n \to \infty$ with m/n fixed) and under i.i.d sub-Gaussian $\boldsymbol{A}$, the iterations of AMP are governed by a state-evolution whose fixed point—when unique—yields the true posterior means.

From a practical viewpoint, the impact of the original AMP work [24] was not that LBP can solve the compressive sensing problem (1), but that it can solve the problem much faster and more accurately than other methods, whether convex-optimization-based, greedy, or Bayesian. To accomplish this feat, [24] proposed an ingenious set of approximations that become accurate in the limit of large dense $\boldsymbol{A}$, yielding an algorithm that gives accurate results using only $\approx 2mn$ operations per iteration and relatively few iterations (e.g., tens).

Remarkably, the "approximate message passing" (AMP) principles in [24]—including the state evolution—can be extended from the compressive sensing problem (1) to more general compressive inference problems of the form (2), as established in [23]. The GAMP algorithm from [23] is summarized in Table I, where $\mathcal{CN}(z;\hat{z},\mu^z)$ is used to denote the circular-Gaussian pdf in variable z with mean $\hat{z}$ and variance μ^z. In the sequel, we detail how this algorithm allows us to tackle the phase retrieval problem.

III. PHASE RETRIEVAL GAMP

To apply the GAMP algorithm outlined in Table I to compressive phase retrieval, one must specify a measurement likelihood function $p_{Y|Z}(y_i|\cdot)$ that models the lack of phase information in the observations, and a signal prior pdf $p_X(\cdot)$ that facilitates measurement compression, e.g., a sparsity-inducing pdf. In addition, one must make a few modifications to the algorithm, as described in Section III-C.

A. Likelihood

The likelihood $p_{Y|Z}(y_i|\cdot)$ that we assign is based on the assumed measurement model

$$y_i = e^{\mathrm{j}\theta_i}(z_i + w_i) \text{ for } i = 1, \ldots, m, \quad (4)$$

where $\mathrm{j} \triangleq \sqrt{-1}$ is not to be confused with the index j, where $\{\theta_i\}$ are unknown i.i.d phases uniformly distributed over $[0, 2\pi)$, where $z_i \triangleq \boldsymbol{a}_i^{\mathsf{H}}\boldsymbol{x}$, and where $\boldsymbol{a}_i^{\mathsf{H}}$ is the i^{th} row of $\boldsymbol{A}$. Essentially, the i.i.d uniform distribution on $\{\theta_i\}$ makes the phase information in $\{y_i\}$ noninformative, after which it suffices to consider only the magnitudes of $\{y_i\}$ when inferring $\boldsymbol{x}$, which are precisely the observed quantities in phase retrieval.

The additive noise $\{w_i\}$ is modeled as i.i.d circular-Gaussian with mean 0 and variance μ^w. Under this setup, it can be recognized that $p_{Y|Z,\Theta}(y|z,\theta;\mu^w) = \mathcal{CN}(y; ze^{\mathrm{j}\theta}, \mu^w)$, and thus

$$p_{Y|Z}(y|z;\mu^w) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{\pi\mu^w} e^{-\frac{1}{\mu^w}|y - ze^{\mathrm{j}\theta}|^2} d\theta, \quad (5)$$

which, after some calculus, reduces to

$$p_{Y|Z}(y|z;\mu^w) = \frac{1}{\pi\mu^w} e^{-\frac{(|y|-|z|)^2}{\mu^w}} I_0(\rho) e^{-\rho} \text{ for } \rho \triangleq \frac{2|y|\,|z|}{\mu^w} \quad (6)$$

where $I_0(\cdot)$ is the 0^{th}-order modified Bessel function of the first kind. It may be interesting to note that, as ρ grows from 0 to ∞, the term $I_0(\rho)e^{-\rho}$ decreases monotonically from 1 to 0. As anticipated, the measurement-channel pdf depends only on the magnitudes of y and z, and not their phases.

The functions $g_{\text{out}}(\cdot,\cdot,\cdot)$ and $g'_{\text{out}}(\cdot,\cdot,\cdot)$ defined in steps (D2)-(D3) of Table I can then be computed using

$$\begin{aligned}
&\mathrm{E}_{Z|Y,P}\{Z|y,\hat{p};\mu^p\} \\
&= \frac{\int_{\mathbb{C}} z\, p_{Y|Z}(y|z;\mu^w)\mathcal{CN}(z;\hat{p},\mu^p)dz}{\int_{\mathbb{C}} p_{Y|Z}(y|z';\mu^w)\mathcal{CN}(z';\hat{p},\mu^p)dz'} \quad (7)\\
&= \left(\frac{|y|}{1+\mu^w/\mu^p} R_0(\phi) + \frac{|\hat{p}|}{\mu^p/\mu^w + 1}\right)\frac{\hat{p}}{|\hat{p}|} \quad (8)
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{var}_{Z|Y,P}\{Z|y,\hat{p};\mu^p\} \\
&= \frac{\int_{\mathbb{C}} |z|^2\, p_{Y|Z}(y|z;\mu^w)\mathcal{CN}(z;\hat{p},\mu^p)dz}{\int_{\mathbb{C}} p_{Y|Z}(y|z';\mu^w)\mathcal{CN}(z';\hat{p},\mu^p)dz'} \\
&\quad - |\,\mathrm{E}_{Z|Y,P}\{Z|y,\hat{p};\mu^p\}|^2 \\
&= \frac{|y|^2}{(1+\mu^w/\mu^p)^2} + \frac{|\hat{p}|^2}{(\mu^p/\mu^w+1)^2} + \frac{1+\phi R_0(\phi)}{1/\mu^w + 1/\mu^p} \\
&\quad - |\,\mathrm{E}_{Z|Y,P}\{Z|y,\hat{p};\mu^p\}|^2 \quad (9)
\end{aligned}$$

where

$$R_0(\phi) \triangleq \frac{I_1(\phi)}{I_0(\phi)} \text{ and } \phi \triangleq \frac{2|y|\,|\hat{p}|}{\mu^w + \mu^p}. \quad (10)$$

Although the noise variance μ^w may not be known apriori, it can be learned from the observations through an expectation-maximization (EM) [31] procedure, similar to that proposed in [32]. We leave these details to future work.

B. Signal prior

As for the compressibility-enabling (e.g., sparse) signal prior $p_X(\cdot)$, it is not expected to be known apriori in practice. However, it can essentially be learned from the compressed observations $\boldsymbol{y}$ using a scheme akin to EM-GM-AMP [32]. There, $p_X(\cdot)$ is modeled as a Gaussian mixture (GM), and the EM procedure is used to learn the GM parameters (i.e., means, variances, and weights) from the compressed measurements $\boldsymbol{y}$. It turns out that all of the quantities needed for the E-step of the EM algorithm are already computed by GAMP, making the overall approach very computationally efficient. Structured compressibility-enabling priors could also be used, such as the total-variation norm or structured sparsity [33], if appropriate. We leave these details to future work.

C. Stepsize, normalization, initialization, and re-starts

With numerical robustness in mind, we propose a modification of GAMP that normalizes certain terms that grow very large (or small) as the iterations progress. In particular, we define $\alpha(t) \triangleq \{\mu_i^p(t)\}_{i=1}^m$ (which tends to grow exponentially small with t), normalize both $\hat{s}_i(t)$ and $\mu_i^s(t)$—which tend to grow—by $1/\alpha(t)$, and normalize $\mu_j^r(t)$—which tends to shrink—by $\alpha(t)$. The resulting normalized quantities are denoted by underbars in Table II. Under infinite numerical precision, these normalizations would cancel each other out and have no effect. With finite numerical precision, however, they can help to stabilize the algorithm.

Next we propose a "step-size" modification of GAMP, as motivated by the following. In phase retrieval, the circular-Gaussian approximations used within GAMP are mismatched to the fact that the likelihood $p_{Y|Z}(y_i|z_i)$ is not circularly symmetric in z_i. For this reason, the steps taken by the algorithm can be overly aggressive at times. To counteract this behavior, we propose to slow down the algorithm using a positive stepsize $\beta < 1$ that is incorporated as shown in Table II. In our experience, $\beta = 0.25$ works well. One consequence of the stepsize modification is the existence of additional state variables like $\bar{x}_j(t)$. To avoid the need to initialize these variables, we use $\beta = 1$ during the first iteration.

For PR-GAMP, we found that it worked well to initialize each $\hat{x}_j(1)$ randomly using a zero-mean circular-Gaussian distribution of large variance (e.g., 10^4), and then set the corresponding $\mu_j^x(1) = \mathrm{var}\{X\} + |\hat{x}_j(1) - \mathrm{E}\{X\}|^2$.

When operating near PR-GAMP's phase transition (see Section IV-A), it may help to re-start the algorithm from a different random initialization if it seems to have converged to a suboptimal solution, e.g., if the squared residual $\frac{1}{m}\sum_{i=1}^m (y_i - |\boldsymbol{a}_i^{\mathsf{H}}\hat{\boldsymbol{x}}|)^2$ is larger than the assumed noise variance μ^w. However, for problems that are far on the "good" side of the phase transition, we find that there is rarely a need for re-starts, and for problems that are far on the "bad" side of the phase transition, we find that such re-starts are generally a waste of time.

for $t = 1, 2, 3\ldots$

$$\forall i: \mu_i^p(t) = \beta \textstyle\sum_{j=1}^n |a_{ij}|^2 \mu_j^x(t) + (1-\beta)\mu_i^p(t-1) \quad \text{(S1)}$$

$$\alpha(t) = \tfrac{1}{m}\textstyle\sum_{i=1}^m \mu_i^p(t) \quad \text{(S2)}$$

$$\forall i: \hat{p}_i(t) = \textstyle\sum_{j=1}^n a_{ij}\hat{x}_j(t) - \frac{\mu_i^p(t)}{\alpha(t)}\,\underline{\hat{s}}_i(t-1) \quad \text{(S3)}$$

$$\forall i: \underline{\hat{s}}_i(t) = \beta\alpha(t) g_{\mathsf{out}}(y_i, \hat{p}_i(t), \mu_i^p(t)) + (1-\beta)\underline{\hat{s}}_i(t-1) \quad \text{(S4)}$$

$$\forall i: \underline{\mu}_i^s(t) = -\beta\alpha(t) g'_{\mathsf{out}}(y_i, \hat{p}_i(t), \mu_i^p(t)) + (1-\beta)\underline{\mu}_i^s(t-1) \quad \text{(S5)}$$

$$\forall j: \underline{\mu}_j^r(t) = \big(\textstyle\sum_{i=1}^m |a_{ij}|^2 \underline{\mu}_i^s(t)\big)^{-1} \quad \text{(S6)}$$

$$\forall j: \bar{x}_j(t) = \beta\hat{x}_j(t) + (1-\beta)\bar{x}_j(t-1) \quad \text{(S7)}$$

$$\forall j: \hat{r}_j(t) = \bar{x}_j(t) + \underline{\mu}_j^r(t)\textstyle\sum_{i=1}^m a_{ij}^*\underline{\hat{s}}_i(t) \quad \text{(S8)}$$

$$\forall j: \mu_j^x(t+1) = \alpha(t)\underline{\mu}_j^r(t) g'_{\mathsf{in}}\big(\hat{r}_j(t), \alpha(t)\underline{\mu}_j^r(t)\big) \quad \text{(S9)}$$

$$\forall j: \hat{x}_j(t+1) = g_{\mathsf{in}}\big(\hat{r}_j(t), \alpha(t)\underline{\mu}_j^r(t)\big) \quad \text{(S10)}$$

end

TABLE II
GAMP WITH VARIANCE NORMALIZATION $\alpha(t)$ AND STEPSIZE β. DURING THE FIRST ITERATION ($t = 1$), $\beta = 1$ IS USED IN ORDER TO AVOID THE NEED TO INITIALIZE $\mu_i^p(0)$, $\underline{\mu}_i^s(0)$, AND $\bar{x}_j(0)$.

IV. Numerical Results

A. Empirical phase transitions

In this section, we demonstrate the performance of PR-GAMP on a wide range of problem settings. Unless otherwise noted, we generated k-sparse length-n signal vectors $\boldsymbol{x}_0$ with support chosen uniformly at random, where the nonzero coefficients were i.i.d zero-mean circular-Gaussian. We then generated sensing matrices of the form $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{F}$, where $\boldsymbol{F}$ was the unitary $n \times n$ one-dimensional DFT matrix, and the elements of $\boldsymbol{\Phi} \in \mathbb{C}^{m\times n}$ were chosen i.i.d zero-mean circular-Gaussian (as in [20]). PR-GAMP had knowledge of $\boldsymbol{A}$ and the m noisy magnitude measurements $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}|$, where $\boldsymbol{w}$ was i.i.d circular-Gaussian, from which it generated the estimate $\hat{\boldsymbol{x}}$ of the true signal $\boldsymbol{x}_0$. Performance is then assessed using the phase-corrected normalized mean-squared error

$$\mathsf{NMSE} \triangleq \min_\theta \frac{\|\boldsymbol{x}_0 - e^{\mathsf{j}\theta}\hat{\boldsymbol{x}}\|_2^2}{\|\boldsymbol{x}_0\|_2^2}. \quad (11)$$

For these experiments, we allowed PR-GAMP up to 9 re-starts.

We first study the conditions under which PR-GAMP is likely to yield a "successful" recovery, which we define as the event that $\mathsf{NMSE} < 10^{-4}$. Figure 2 plots the empirical success probability (over 500 independent problem realizations) as a function of signal sparsity k and number of measurements m, for fixed signal length $n = 512$ and fixed signal-to-noise ratio $\mathsf{SNR} \triangleq \|\boldsymbol{A}\boldsymbol{x}_0\|_2^2/\|\boldsymbol{w}\|_2^2 = 100$ dB. 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. Given the logarithmic axes of Fig. 3, the near-linearity of the phase transition in the figure suggests a near-polynomial relationship between

the signal sparsity, k, and the minimum required number of measurements, m, at a fixed signal length n. To see how well (versus how often) PR-GAMP is recovering the signal, we plot the median NMSE over the same problem realizations in Fig. 3. There we see that recovery is extremely good throughout the region above the phase transition. Overall, Fig. 2 and Fig. 3 demonstrate that PR-GAMP is indeed capable of *compressive* phase retrieval, i.e., successful $\mathbb{C}^n$-signal recovery from $m < 3n - 2$ magnitude-only measurements.

For reference, Fig. 4 extracts the 50%-success contour from Fig. 2 and compares it to the corresponding contour for phase-oracle (PO)-GAMP (i.e., GAMP operating on measurements of the form $\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}$) calculated from the same problem realizations. There we see that the two phase transitions have the same slope, but that the PO-GAMP phase transition is shifted downward/rightward. Moreover, a closer comparison of the two phase transitions suggests that, for uccessful recovery, GAMP requires about 4 times the number of magnitude-only observations than magnitude-plus-phase observations.

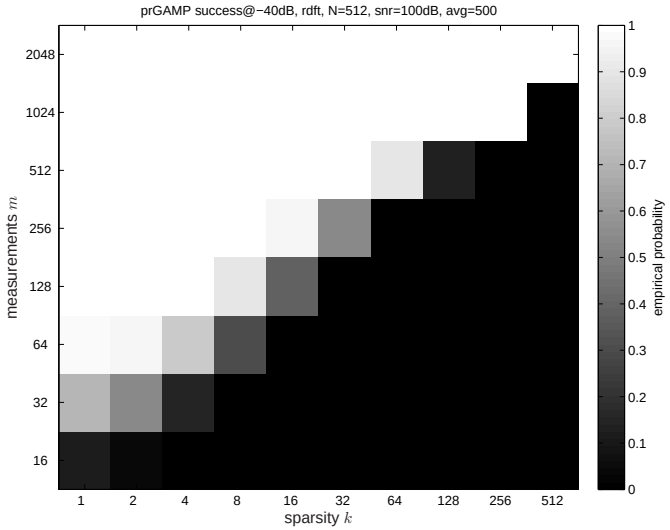

Fig. 2. Empirical probability of successful PR-GAMP recovery of an $n = 512$-length Bernoulli-complex-Gaussian signal, versus signal sparsity k and number of measurements m, from magnitude-only measurements at $\mathsf{SNR} = 100$ dB.

B. Robustness to noise

We now demonstrate the robustness of PR-GAMP to non-trivial levels of additive white circular-Gaussian noise $\boldsymbol{w}$ in the magnitude-only measurements $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}|$. As above, we use a $n = 512$-length Bernoulli-complex-Gaussian signal, but now we focus on the case $(k, m) = (4, 256)$, which is on the "good" side of the phase-transition in Fig. 3. Figure 5 shows median NMSE performance over 2000 independent problem realizations as a function of $\mathsf{SNR} \triangleq \|\boldsymbol{A}\boldsymbol{x}_0\|_2^2/\|\boldsymbol{w}\|_2^2$. At larger values of SNR (i.e., $\mathsf{SNR} \geq 30$ dB), we see that PR-GAMP performs about 3 dB worse than PO-GAMP, which is intuitive given that the latter observes twice as many

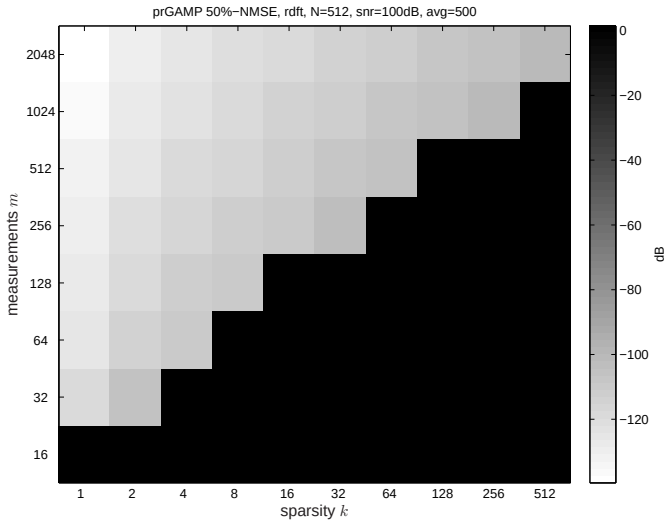

Fig. 3. Median NMSE for PR-GAMP recovery of an $n = 512$-length Bernoulli-complex-Gaussian signal, versus signal sparsity k and number of measurements m, from magnitude-only measurements at $\mathsf{SNR} = 100$ dB.

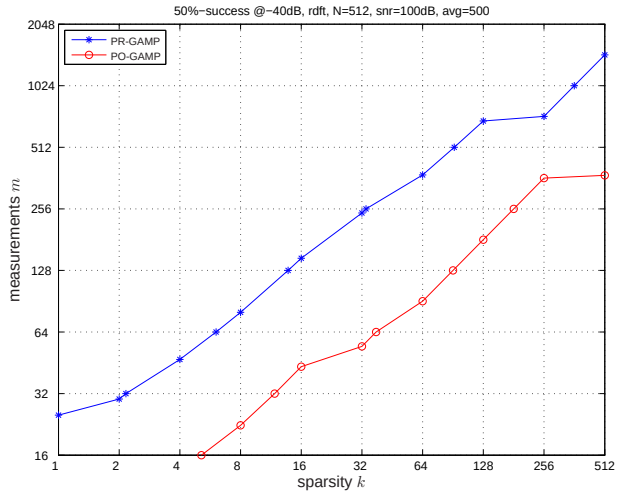

Fig. 4. 50%-success contours for PR-GAMP and phase-oracle GAMP recovery of an $n = 512$-length Bernoulli-complex-Gaussian signal, versus signal sparsity k and number of measurements m, at $\mathsf{SNR} = 100$ dB.

real-valued measurements as PR-GAMP. At smaller values of SNR, we see that PR-GAMP does return reasonable signal estimates, although the gap between its NMSE and that of PO-GAMP is wider.

C. Comparison to CPRL

In this section, we present numerical results that compare PR-GAMP to the state-of-the-art convex-relaxation approach to compressive phase retrieval, CPRL [20]. To implement CPRL, we used the public-domain matlab code posted by the authors,[3] which invokes the optimization package CVX.[4] As before, we examine the probability of success (i.e.,

[3] http://www.control.isy.liu.se/~ohlsson/CPR.zip
[4] http://cvxr.com/cvx/

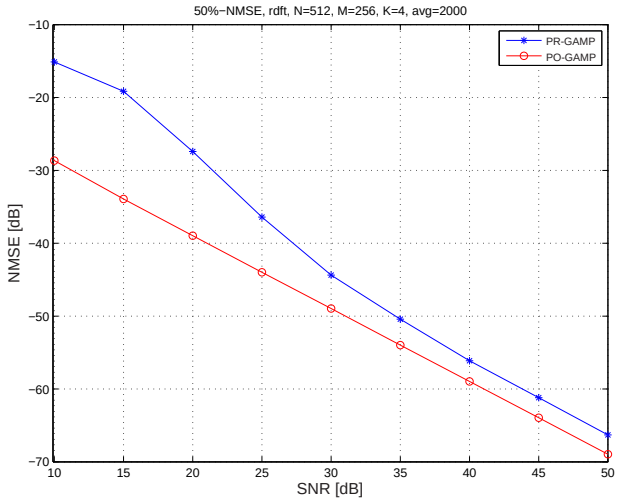

Fig. 5. NMSE for PR-GAMP and phase-oracle GAMP recovery of an $n = 512$-length $k = 4$-sparse Bernoulli-complex-Gaussian signal, versus SNR, from $m = 256$ measurements. The performance shown is the median over 2000 independent problem realizations.

NMSE $< 10^{-4}$) of k-sparse Bernoulli-complex-Gaussian signal recovery from m magnitude only measurements of the form $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}|$, where $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{F}$ for i.i.d circular-Gaussian $\boldsymbol{\Phi}$ and DFT $\boldsymbol{F}$, and where $\boldsymbol{w}$ is i.i.d circular-Gaussian noise yielding 100 dB SNR.

Table III shows empirical success probability 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 both CPRL and PR-GAMP were successful in recovering the signal at all tested combinations of (m, n), although for smaller values of (m, n) PR-GAMP failed on a small percentage of realizations because the central-limit-theorem approximations used within AMP are not well justified at such small problem sizes. On the other hand, Table III shows that the runtime of CPRL increases rapidly with the signal dimension n, whereas that of PR-GAMP remains orders-of-magnitude smaller and independent of (m, n) over the range tested here.[5]

Table IV repeats the experiment carried out in Table III, but at the sparsity $k = 2$. For this more difficult problem, the table shows that CPRL is much less successful at recovering the signal than PR-GAMP. Meanwhile, the runtimes reported in Table IV again show CPRL complexity scaling rapidly with the problem dimension, whereas GAMP complexity stays orders-of-magnitude smaller and constant over the tested problem dimensions. In fact, the comparisons conducted in this section were restricted to very small problem dimensions precisely due to the complexity scaling of CPRL.

[5]Although the complexity of GAMP is known to scale as $\mathcal{O}(mn)$ for this choice of $\boldsymbol{A}$, the values of m and n in Table III are too small for this scaling law to manifest.

	$(m,n) = (20,32)$	$(m,n) = (30,48)$	$(m,n) = (40,64)$
CPRL	0.96 (4.9 sec)	0.97 (51 sec)	0.99 (291 sec)
PR-GAMP	0.83 (0.4 sec)	0.94 (0.3 sec)	0.99 (0.3 sec)

TABLE III
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	0.55 (5.8 sec)	0.55 (58 sec)	0.58 (316 sec)
PR-GAMP	0.72 (0.4 sec)	0.92 (0.3 sec)	1.0 (0.3 sec)

TABLE IV
SUCCESS RATE AND MEDIAN RUNTIME OVER 100 PROBLEM REALIZATIONS FOR SEVERAL COMBINATIONS OF SIGNAL LENGTH n, MEASUREMENT LENGTH m, AND SIGNAL SPARSITY $k = 2$.

D. Practical image recovery

Finally, we demonstrate practical image recovery with PR-GAMP. For this experiment, the signal $\boldsymbol{x}_0$ was the $n = 65536$-pixel grayscale image shown on the left of Fig. 6, which has a sparsity ratio of $k/n \approx 0.1$. Since this image is real and non-negative, we used a Bernoulli-truncated-Gaussian prior for PR-GAMP (as opposed to the Bernoulli-circular-Gaussian prior used in previous experiments).

In addition, we used a linear transformation $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ of the form

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B}_1 & \\ & \boldsymbol{B}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{F} & \\ & \boldsymbol{F} \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_1 \\ \boldsymbol{M}_2 \end{bmatrix}, \tag{12}$$

where $\boldsymbol{F}$ was a 2D DFT matrix of size $n \times n$, $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ were diagonal "masking" matrices of size $n \times n$ with diagonal entries drawn from $\{0, 1\}$, and $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ were banded matrices of size $\frac{m}{2} \times n$ with 10 nonzero i.i.d circular-Gaussian entries per column.[6] The reason for including $\boldsymbol{M}_i$ and $\boldsymbol{B}_i$ was to "randomize" the DFT, i.e., to avoid unicity issues such as shift and flip ambiguities. Moreover, unlike the dense random matrix $\boldsymbol{\Phi}$ used previously for randomization, $\boldsymbol{M}_i$ and $\boldsymbol{B}_i$ are sparse matrices that—together with an FFT implementation of $\boldsymbol{F}$—led to a fast implementation of $\boldsymbol{A}$.

To eliminate the need for the expensive matrix multiplications with the elementwise-squared versions of $\boldsymbol{A}$ and $\boldsymbol{A}^{\mathsf{H}}$, as specified in steps (S1) and (S6) of Table II, GAMP was run in "uniform variance" mode, meaning that $\{\mu_i^p(t)\}_{i=1}^m$ were approximated by $\mu^p(t) \triangleq \frac{1}{m}\sum_{i'=1}^m \mu_{i'}^p(t)$, and similar was done with $\{\mu_i^s(t)\}_{i=1}^m$, $\{\mu_j^r(t)\}_{j=1}^n$, and $\{\mu_j^x(t)\}_{j=1}^n$. The result is that (S1)-(S2) become $\mu^p(t) = \beta\|\boldsymbol{A}\|_F^2\mu^x(t)/m + (1-\beta)\mu^p(t-1) = \alpha(t)$ and (S6) becomes $\underline{\mu}^r(t) = \left(\|\boldsymbol{A}\|_F^2\underline{\mu}^s(t)/n\right)^{-1}$.

As before, the observations took the form $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}|$, but now the noise variance was adjusted to the nontrivial level of SNR $= 30$ dB. To demonstrate *compressive* image

[6]In particular, the diagonal of $\boldsymbol{M}_1$ had $n/2$ ones positioned uniformly at random, and the diagonal of $\boldsymbol{M}_2$ was its complement, so that $\boldsymbol{M}_1 + \boldsymbol{M}_2 = \boldsymbol{I}$. Also, since each $\boldsymbol{B}_i$ was a wide matrix, its nonzero band was wrapped from bottom to top when necessary.

recovery, only $m = n/2 = 32768$ measurements were used. Running PR-GAMP on 100 problem realizations (each with different random $\boldsymbol{A}$ and $\boldsymbol{w}$, and allowing at most 5 re-starts per realization), a **89% success rate** was observed, where for this noisy problem "success" was defined as $\mathsf{NMSE} < -27$ dB (i.e., no more than 3 dB worse than SNR^{-1}). See the right side of Fig. 6 for a typical PR-GAMP recovery. Furthermore, among the successful realizations, the median runtime was only **13.4 seconds**. To put this number in perspective, we recall Tables III and IV, which showed CPRL consuming ~ 300 seconds on toy problems of size $(m, n) = (40, 64)$.

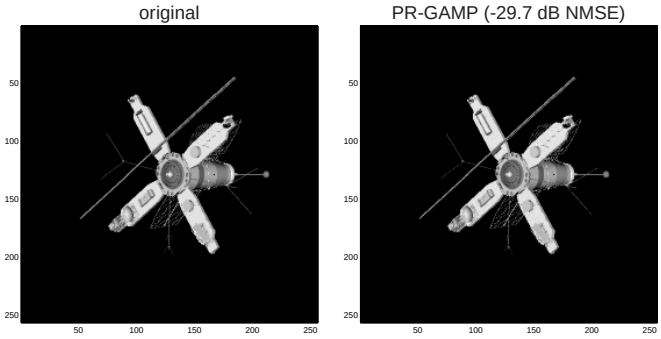

Fig. 6. Original image (left) and a typical PR-GAMP-recovery (right) from $m = n/2$ measurements at $\mathsf{SNR} = 30$ dB, which took 11.1 seconds.

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 Bernoulli-circular-Gaussian signals, PR-GAMP requires ≈ 4 times the number of measurements as phase-oracle GAMP and, at moderate to large SNR, the NMSE of PR-GAMP is only ≈ 3 dB worse than that of phase-oracle GAMP. For recovery of a real-valued 65532-pixel image from 32768 randomly masked magnitude measurements, PR-GAMP was successful 89% of the time, with a median runtime of only 13.4 seconds. Comparison to the state-of-the-art convex relaxation CPRL revealed PR-GAMPs superior phase transition and runtimes that were many orders-of-magnitude faster.

References

[1] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Optics*, vol. 21, pp. 2758–2769, Aug. 1982.

[2] A. Walther, "The question of phase retrieval in optics," *Optica Acta*, vol. 10, no. 1, pp. 41–49, 1963.

[3] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Optical Soc. America A*, vol. 7, pp. 394–411, Mar. 1990.

[4] J. C. Dainty and J. R. Fienup, "Phase retrieval and image construction for astronomy," in *Image Recovery: Theory and Application* (H. Stark, ed.), ch. 7, pp. 231–275, New York: Academic Press, 1987.

[5] L. Rabiner and B. H. Huang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[6] A. Sayed and T. Kailath, "A survey of spectral factorization methods," *Numer. Linear Algebra Appl.*, vol. 8, no. 6-7, pp. 467–496, 2001.

[7] M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier tranform," *IEEE Trans. Acoust. Speech & Signal Process.*, vol. 30, pp. 140–154, 1982.

[8] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optika*, vol. 35, no. 2, pp. 237–246, 1972.

[9] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization," *J. Optical Soc. America A*, vol. 19, pp. 1334–1345, July 2002.

[10] A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," *Inverse Problems*, vol. 27, no. 1, pp. 1–16, 2011.

[11] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *arXiv:1109.0573*, Sept. 2011.

[12] E. J. Candès, T. Strohmer, and V. Voroninski, "Phase lift: Exact and stable signal recovery from magnitude measurements via convex programming," *arXiv:1109.4499*, Sept. 2011.

[13] I. Waldspurger, A. D'Aspremont, and S. Mallat, "Phase recovery, MaxCut and complex semidefinite programming," *arXiv:1206.0102*, June 2012.

[14] E. J. Candès and X. Li, "Solving quadratic equations via PhaseLift when there are about as many equations as unknowns," *arXiv:1208.6247*, Aug. 2012.

[15] J. Finkelstein, "Pure-state informationally complete and "really" complete measurements," *Phys. Rev. A*, vol. 70, pp. 052107–052110, Nov. 2004.

[16] R. Balan, P. G. Casazza, and D. Edidin, "On signal reconstruction without noisy phase," *Appl. Computational Harmonic Anal.*, vol. 20, pp. 345–355, 2006.

[17] S. Marchesini, "Ab initio compressive phase retrieval," *arXiv:0809.2006*, Sept. 2008.

[18] M. L. Moravec, J. K. Romberg, and R. Baraniuk, "Compressive phase retrieval," in *SPIE Conf. Series*, vol. 6701, (San Diego, CA), Aug. 2007.

[19] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications* (Y. C. Eldar and G. Kutyniok, eds.), Cambridge Univ. Press, 2012.

[20] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, "Compressive phase retrieval from squared output measurements via semidefinite programming," *arXiv:1111.6323*, Nov. 2011.

[21] Y. M. Lu and M. Vetterli, "Sparse spectral factorization: Unicity and reconstruction algorithms," in *Proc. IEEE Int. Conf. Acoust. Speech & Signal Process.*, (Prague, Czech Republic), pp. 5976–5979, Mar. 2011.

[22] K. Jaganathan, S. Oymak, and B. Hassibi, "Recovery of sparse 1-D signals from the magnitudes of their Fourier transform," *arXiv:1206.0102*, June 2012.

[23] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Thy.*, (Saint Petersburg, Russia), Aug. 2011. (See also *arXiv:1010.5141*).

[24] 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.

[25] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," in *Proc. Inform. Theory Workshop*, (Cairo, Egypt), pp. 1–5, Jan. 2010.

[26] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufman, 1988.

[27] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[28] B. J. Frey and D. J. C. MacKay, "A revolution: Belief propagation in graphs with cycles," in *Proc. Neural Inform. Process. Syst. Conf.*, (Denver, CO), pp. 479–485, 1997.

[29] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, pp. 764–785, Feb. 2011.

[30] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and iterative algorithms," in *Proc. IEEE Int. Symp. Inform. Thy.*, (Boston, Ma), pp. 1–5, June 2012.

[31] 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.

[32] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *arXiv:1207.3107*, July 2012.

[33] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. Conf. Inform. Science & Syst.*, (Princeton, NJ), pp. 1–6, Mar. 2010.