# GENERALIZED APPROXIMATE MESSAGE PASSING FOR COSPARSE ANALYSIS COMPRESSIVE SENSING

*Mark Borgerding, Philip Schniter, Jeremy Vila*

Dept. ECE, The Ohio State University
Columbus, OH 43210

*Sundeep Rangan*

Dept. of ECE, NYU Polytechnic Institute
Brooklyn, NY 11201

## ABSTRACT

In cosparse analysis compressive sensing (CS), one seeks to estimate a non-sparse signal vector from noisy sub-Nyquist linear measurements by exploiting the knowledge that a given linear transform of the signal is cosparse, i.e., has sufficiently many zeros. We propose a novel approach to cosparse analysis CS based on the generalized approximate message passing (GAMP) algorithm. Unlike other AMP-based approaches to this problem, ours works with a wide range of analysis operators and regularizers. In addition, we propose a novel $\ell_0$-like soft-thresholder based on MMSE denoising for a spike-and-slab distribution with an infinite-variance slab. Numerical demonstrations on synthetic and practical datasets demonstrate advantages over existing AMP-based, greedy, and reweighted-$\ell_1$ approaches.

***Index Terms—*** Approximate message passing, belief propagation, compressed sensing.

## 1. INTRODUCTION

We consider the problem of recovering a signal $\boldsymbol{x} \in \mathbb{R}^N$ (e.g., an $N$-pixel image) from the possibly noisy linear measurements

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{w} \in \mathbb{R}^M, \tag{1}$$

where $\boldsymbol{\Phi}$ represents a known linear measurement operator $\boldsymbol{w}$ represents noise, and $M \ll N$. We focus on the *analysis compressive sensing* (CS) problem [1, 2] where, for a given analysis operator $\boldsymbol{\Omega}$,

$$\boldsymbol{u} \triangleq \boldsymbol{\Omega}\boldsymbol{x} \in \mathbb{R}^D \tag{2}$$

is assumed to be *cosparse* (i.e., contain sufficiently many zero-valued coefficients). This differs from the *synthesis CS* problem, where $\boldsymbol{x}$ itself is assumed to be *sparse* (i.e., contain sufficiently few non-zero coefficients). Although the two problems become interchangeable when $\boldsymbol{\Omega}$ is invertible, we are mainly interested in non-invertible $\boldsymbol{\Omega}$, e.g., the "overcomplete" case where $D > N$. For notational simplicity, we assume real-valued quantities in the sequel. However, our methods are easily generalized to complex-valued quantities, as we demonstrate in the numerical experiments.

The analysis CS problem is typically formulated as a regularized loss minimization problem of the form

$$\widehat{\boldsymbol{x}}_{\mathsf{rlm}} = \arg\min_{\boldsymbol{x}} \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + h(\boldsymbol{\Omega}\boldsymbol{x}), \tag{3}$$

with separable regularizer $h(\boldsymbol{u}) = \sum_{d=1}^{D} h_d(u_d)$. One of the most famous instances of $h(\boldsymbol{u})$ is that of total-variation (TV) regularization [3], where $h(\boldsymbol{u}) = \lambda\|\boldsymbol{u}\|_1$ and $\boldsymbol{\Omega}$ computes variation across

neighboring pixels. In the anisotropic case, this variation is measured by finite difference operators, e.g., $\boldsymbol{\Omega} = [\boldsymbol{D}_h^{\mathsf{H}}, \boldsymbol{D}_v^{\mathsf{H}}]^{\mathsf{H}}$, where $\boldsymbol{D}_h$ computes horizontal differences and $\boldsymbol{D}_v$ computes vertical differences. Of course, $\ell_1$ regularization can be used with generic $\boldsymbol{\Omega}$, with the desirable property that it always renders (3) convex. The resulting problem, sometimes referred to as the *generalized LASSO* (GrLASSO) [4], is amenable to a wide range of efficient optimization techniques like Douglas-Rachford splitting [5] and NESTA [6].

Despite the elegance of the $\ell_1$ norm, several studies have shown improvements from the use of $\ell_0$-like norms for $h(\boldsymbol{u})$, especially for highly overcomplete $\boldsymbol{\Omega}$ (i.e., $D \gg N$). For example, the use of iteratively reweighted $\ell_1$ [7] has demonstrated significant improvements over $\ell_1$ regularization in the context of analysis CS [8, 9]. Likewise, greedy approaches to locate the zero-valued elements in $\boldsymbol{u}$ have also demonstrated significant improvements over $\ell_1$. Examples include greedy analysis pursuit (GAP), analysis iterative hard thresholding (AIHT), analysis hard thresholding pursuit (AHTP), analysis CoSaMP (ACoSaMP), and analysis subspace pursuit (ASP) [2, 10].

In this paper, we propose a Bayesian approach to analysis CS that leverages recent advances in approximate message passing (AMP) algorithms [11, 12], and in particular the generalized AMP (GAMP) algorithm from [13].

While other AMP-based approaches have been recently proposed for the special case where $\boldsymbol{\Omega}$ is a 1D finite difference operator, i.e., the TV-AMP from [14] and the ssAMP from [15], our approach works with a *generic* analysis operator $\boldsymbol{\Omega}$ and a much broader range of signal priors and likelihoods. Furthermore, our approach facilitates both MAP and (approximate) MMSE estimation of $\boldsymbol{x}$ in a computationally efficient manner. We also note that a different Bayesian approach to cosparse analysis CS, based on multivariate Gauss-mixture priors, was recently presented in [16]. The MAP and MMSE estimation methods proposed in [16], which employ greedy pursuit and Gibbs sampling, respectively, have computational complexities that scale as $O(D^2N)$ and $O(D^2N^3)$ (assuming $M \leq D$). In contrast, ours scales like $O(DN)$ for generic $\boldsymbol{\Omega}$, or $O(N \log N)$ when $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$ have fast implementations, which is often the case in imaging applications.

## 2. GENERALIZED AMP FOR ANALYSIS CS

### 2.1. The proposed Bayesian model

Our approach is Bayesian in that it treats the true signal $\boldsymbol{x}$ as a realization of a random vector $\boldsymbol{\mathsf{x}} \in \mathbb{R}^N$ with prior pdf $p_{\boldsymbol{\mathsf{x}}}(\boldsymbol{x})$ and likelihood function $p_{\boldsymbol{\mathsf{y}}|\boldsymbol{\mathsf{q}}}(\boldsymbol{y}|\boldsymbol{\Phi}\boldsymbol{x})$, where $\boldsymbol{y}$ are the observed noisy measurements and $\boldsymbol{\mathsf{q}} \triangleq \boldsymbol{\Phi}\boldsymbol{\mathsf{x}}$ are akin to hidden noiseless measurements. (For clarity, we write random quantities using san-serif fonts and deterministic ones using serif fonts.) Furthermore, we assume that the

prior and likelihood have the forms

$$p_{\boldsymbol{y}|\boldsymbol{q}}(\boldsymbol{y}|\boldsymbol{\Phi}\boldsymbol{x}) \propto \prod_{m=1}^{M} \exp(-l_m([\boldsymbol{\Phi}\boldsymbol{x}]_m)) \tag{4}$$

$$p_{\boldsymbol{x}}(\boldsymbol{x}) \propto \prod_{d=1}^{D} \exp(-h_d([\boldsymbol{\Omega}\boldsymbol{x}]_d)) \prod_{n=1}^{N} \exp(-g_n(x_n)) \tag{5}$$

with scalar functions $l_m(\cdot)$, $h_d(\cdot)$, and $g_n(\cdot)$. Note that the function $l_m(\cdot)$ is dependent on the measured value $y_m$. We discuss the design of these functions in the sequel.

Given the form of (4) and (5), the MAP estimate $\widehat{\boldsymbol{x}}_{\mathsf{MAP}} \triangleq \arg\max_{\boldsymbol{x}} p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ can be written (using Bayes rule) as

$$\widehat{\boldsymbol{x}}_{\mathsf{MAP}} = \arg\min_{\boldsymbol{x}} \left\{ l(\boldsymbol{\Phi}\boldsymbol{x}) + h(\boldsymbol{\Omega}\boldsymbol{x}) + g(\boldsymbol{x}) \right\} \tag{6}$$

with separable loss function $l(\boldsymbol{q}) = \sum_{m=1}^{M} l_m(q_m)$ and separable regularizers $g(\boldsymbol{x}) = \sum_{n=1}^{N} g_n(x_n)$ and $h(\boldsymbol{u}) = \sum_{d=1}^{D} h_d(u_d)$. Note that, with trivial $g(\boldsymbol{x}) = 0$ and quadratic loss $l(\boldsymbol{q}) = \frac{1}{2}\|\boldsymbol{q} - \boldsymbol{y}\|_2^2$, the MAP estimation problem (6) reduces to the regularized loss minimization problem (3). But clearly (6) is more general.

As for the MMSE estimate $\widehat{\boldsymbol{x}}_{\mathsf{MMSE}} \triangleq \int \boldsymbol{x}\, p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x}$, exact evaluation requires the computation of a high dimensional integral that is intractable for most problem sizes of interest. In the sequel, we present a computationally efficient approach to approximate MMSE estimation that is based on loopy belief propagation and, in particular, the GAMP algorithm from [13].

## 2.2. Background on GAMP

The GAMP algorithm [13] aims to estimate the signal $\boldsymbol{x}$, which is assumed to be a realization of random vector $\boldsymbol{X} \in \mathbb{R}^N$ with known prior $p_{\boldsymbol{x}}(\boldsymbol{x})$, from the observations $\boldsymbol{y}$ generated in accordance with the likelihood function $p_{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{x})$. Here, the prior and likelihood are assumed to be separable in the sense that

$$p_{\boldsymbol{x}}(\boldsymbol{x}) \propto \prod_{n=1}^{N} \exp(-g_n(x_n)), \quad p_{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{y}|\boldsymbol{z}) \propto \prod_{i=1}^{I} \exp(-f_i(z_i)), \tag{7}$$

where $\boldsymbol{z} \triangleq \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^I$ can be interpreted as hidden transform outputs. The MAP version of GAMP aims to compute $\widehat{\boldsymbol{x}}_{\mathsf{MAP}} = \arg\max_{\boldsymbol{x}} p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$, i.e., solve the optimization problem

$$\widehat{\boldsymbol{x}}_{\mathsf{MAP}} = \arg\min_{\boldsymbol{x}} \sum_{i=1}^{I} f_i([\boldsymbol{A}\boldsymbol{x}]_i) + \sum_{n=1}^{N} g_n(x_n), \tag{8}$$

while the MMSE version of GAMP aims to compute the MMSE estimate $\widehat{\boldsymbol{x}}_{\mathsf{MMSE}} \triangleq \int \boldsymbol{x}\, p_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x}$, in both cases by iterating simple, scalar optimizations. MAP-GAMP can be considered as the extension of the AMP algorithm [11] from the quadratic loss $f(\boldsymbol{z}) = \|\boldsymbol{y} - \boldsymbol{z}\|_2^2$ to generic separable losses of the form $f(\boldsymbol{z}) = \sum_{i=1}^{I} f_i(z_i)$. Likewise, MMSE-GAMP can be considered as a similar extension of the Bayesian-AMP algorithm [12] from additive white Gaussian noise (AWGN) $p_{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{y}|\boldsymbol{z})$ to generic $p_{\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{y}|\boldsymbol{z})$ of the form in (7).

In the large-system limit (i.e., $I, N \to \infty$ with $I/N$ converging to a finite positive constant) under i.i.d sub-Gaussian $\boldsymbol{A}$, GAMP is characterized by a state evolution whose fixed points, when unique, are Bayes optimal [17, 18]. For generic $\boldsymbol{A}$, it has been shown [19] that MAP-GAMP's fixed points coincide with the critical points of the cost function (8) and that MMSE-GAMPs fixed points coincide with those of a certain variational cost that was connected to the

definitions for MMSE-GAMP:

$$F_i(\widehat{p}, \nu^p) \triangleq \frac{\int z \exp(-f_i(z)) \mathcal{N}(z; \widehat{p}, \nu^p) dz}{\int \exp(-f_i(z)) \mathcal{N}(z; \widehat{p}, \nu^p) dz} \tag{D1}$$

$$G_n(\widehat{r}, \nu^r) \triangleq \frac{\int x \exp(-g_n(x)) \mathcal{N}(x; \widehat{r}, \nu^r) dx}{\int \exp(-g_n(x)) \mathcal{N}(x; \widehat{r}, \nu^r) dx} \tag{D2}$$

definitions for MAP-GAMP:

$$F_i(\widehat{p}, \nu^p) \triangleq \arg\min_z f_i(z) + \frac{1}{2\nu^p}|z - \widehat{p}|^2 \tag{D3}$$

$$G_n(\widehat{r}, \nu^r) \triangleq \arg\min_x g_n(x) + \frac{1}{2\nu^r}|x - \widehat{r}|^2 \tag{D4}$$

inputs:
$$\forall i, n: F_i, G_n, \widehat{x}_n(1), \nu_n^x(1), a_{in}, T_{\max} \geq 1, \epsilon \geq 0, \beta_0 \in (0, 1]$$

initialize:
$$\forall i: \widehat{s}_i(0) = 0, \ t = 1$$

for $t = 1, \ldots, T_{\max}$,

$$\text{if } t = 1, \text{ then } \beta = 1, \text{ else } \beta = \beta_0 \tag{R1}$$

$$\forall i: \nu_i^p(t) = \beta \sum_{n=1}^{N} |a_{in}|^2 \nu_n^x(t) + (1-\beta)\nu_i^p(t-1) \tag{R2}$$

$$\forall i: \widehat{p}_i(t) = \sum_{n=1}^{N} a_{in}\widehat{x}_n(t) - \nu_i^p(t)\widehat{s}_i(t-1) \tag{R3}$$

$$\forall i: \nu_i^z(t) = \nu_i^p(t) F_i'(\widehat{p}_i(t), \nu_i^p(t)) \tag{R4}$$

$$\forall i: \widehat{z}_i(t) = F_i(\widehat{p}_i(t), \nu_i^p(t)) \tag{R5}$$

$$\forall i: \nu_i^s(t) = \beta \left(1 - \frac{\nu_i^z(t)}{\nu_i^p(t)}\right) \frac{1}{\nu_i^p(t)} + (1-\beta)\nu_i^s(t-1) \tag{R6}$$

$$\forall i: \widehat{s}_i(t) = \beta \frac{\widehat{z}_i(t) - \widehat{p}_i(t)}{\nu_i^p(t)} + (1-\beta)\widehat{s}_i(t-1) \tag{R7}$$

$$\forall n: \widetilde{x}_n(t) = \beta\widehat{x}_n(t) + (1-\beta)\widetilde{x}_n(t-1) \tag{R8}$$

$$\forall n: \nu_n^r(t) = \beta \left(\frac{1}{\sum_{i=1}^{I} |a_{in}|^2 \nu_i^s(t)}\right) + (1-\beta)\nu_n^r(t-1) \tag{R9}$$

$$\forall n: \widehat{r}_n(t) = \widetilde{x}_n(t) + \nu_n^r(t) \sum_{i=1}^{I} a_{in}^* \widehat{s}_i(t) \tag{R10}$$

$$\forall n: \nu_n^x(t+1) = \nu_n^r(t) G_n'(\widehat{r}_n(t), \nu_n^r(t)) \tag{R11}$$

$$\forall n: \widehat{x}_n(t+1) = G_n(\widehat{r}_n(t), \nu_n^r(t)) \tag{R12}$$

$$\text{if } \|\widehat{\boldsymbol{x}}(t) - \widehat{\boldsymbol{x}}(t+1)\|/\|\widehat{\boldsymbol{x}}(t+1)\| < \epsilon, \text{then stop} \tag{R13}$$

end

outputs: $\forall n: \widehat{x}_n(t+1)$

**Table 1**. The damped GAMP algorithm. In (R4) and (R11), $F_i'$ and $G_n'$ denote the derivatives of $F_i$ and $G_n$ w.r.t their first arguments.

Bethe free entropy in [20]. However, with general $\boldsymbol{A}$ (e.g., non-zero-mean $\boldsymbol{A}$ [21] or ill-conditioned $\boldsymbol{A}$ [22]) GAMP may not converge to its fixed points, i.e., it may diverge. In an attempt to prevent divergence with generic $\boldsymbol{A}$, damped [22], adaptively damped [23], sequential [20], and double-loop [24] versions of GAMP have been proposed.

A damped version of the GAMP algorithm is summarized in Table 1. There, smaller values of the damping parameter $\beta_0$ make GAMP more robust to difficult $\boldsymbol{A}$ at the expense of convergence speed, and $\beta_0 = 1$ recovers the original GAMP algorithm from [13]. Note that the only difference between MAP-GAMP and MMSE-GAMP is the definition of the scalar denoisers in (D1)-(D4). Denoisers of the type in (D3)-(D4) are often referred to "proximal operators" in the optimization literature. In fact, as noted in [19] and [22], max-sum GAMP is closely related to primal-dual algorithms from convex optimization, such as the classical Arrow-Hurwicz and recent Chambolle-Pock and primal-dual hybrid gradient algorithms [25–27]. The primary difference between MAP-GAMP and those algorithms is that the primal and dual stepsizes (i.e., $\nu_n^r(t)$ and $1/\nu_i^p(t)$ in Table 1) are adapted, rather than fixed or scheduled.

### 2.3. GAMP Enables Analysis CS

If we configure GAMP's transform $\boldsymbol{A}$ and loss function $f(\cdot)$ as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\Phi} \\ \boldsymbol{\Omega} \end{bmatrix}, \quad f_i(\cdot) = \begin{cases} l_i(\cdot) & i \in \{1, \ldots, M\} \\ h_{i-M}(\cdot) & i \in \{M+1, \ldots, M+D\} \end{cases} \tag{9}$$

where $\mathbf{\Phi}$ and $\mathbf{\Omega}$ are the measurement and analysis operators from Sec. 2.1, and $l_i(\cdot)$ and $h_d(\cdot)$ are the loss and regularization functions from Sec. 2.1, then MAP-GAMP's optimization problem (8) coincides with the MAP optimization (6), which (as discussed earlier) is a generalization of the analysis-CS problem (3). Likewise, MMSE-GAMP will return an approximation of the MMSE estimate $\widehat{\boldsymbol{x}}_{\mathsf{MMSE}}$ under the statistical model (4)-(5).

In the sequel, we refer to GAMP under (9) (with suitable choices of $f_q$, $g_n$, and $h_d$) as "Generalized AMP for Analysis CS," or GrAMPA. Despite the simplicity of this idea and its importance to, e.g., image recovery, it has (to our knowledge) not been proposed before, outside of our preprint [28].

## 2.4. Choice of loss and regularization

One of the strengths of GrAMPA is the freedom to choose the loss function $l_i(\cdot)$ and the regularizations $g_n(\cdot)$ and $h_d(\cdot)$.

The quadratic loss $l_m(q) = |y_m - q|^2$, as used in (3), is appropriate for many applications. GrAMPA, however, also supports non-quadratic losses, as needed for 1-bit compressed sensing [29], phase retrieval [30], and Poisson-based photon-limited imaging [31].

The pixel regularization $g_n(\cdot)$ could be used to enforce known positivity in $x_n$ (via $g_n(x) = -\ln \mathbb{1}_{x \geq 0}$), real-valuedness in $x_n$ despite complex-valued measurements (via $g_n(x) = -\ln \mathbb{1}_{x \in \mathbb{R}} \; \forall n$), or zero-valuedness in $x_n$ (via $g_n(x) = -\ln \mathbb{1}_{x=0} \; \forall n$). Here, we use $\mathbb{1}_A \in \{0, 1\}$ to denote the indicator of the event $A$.

As for the analysis regularization $h_d(\cdot)$, the use of $h_d(u) = \lambda|u|$ with MAP-GrAMPA would allow it to tackle the GrLASSO and anisotropic TV problems defined in Sec. 1. With MMSE-GrAMPA, a first instinct might be to use the Bernoulli-Gaussian (BG) prior commonly used for synthesis CS, i.e., $h_d(u) = -\ln\big((1-\beta)\delta(u) + \beta\mathcal{N}(u; 0, \sigma^2)\big)$, where $\delta(\cdot)$ is the Dirac delta pdf and the parameters $\beta$ and $\sigma^2$ control sparsity and variance, respectively. But the need to tune two parameters is inconvenient, and bias effects from the use of finite $\sigma^2$ can degrade performance, especially when $D \gg N$.

Thus, for the MMSE case, we propose a *sparse non-informative parameter estimator* (SNIPE) that can be understood as the MMSE denoiser for a "spike-and-slab" prior with an infinite-variance slab. In particular, SNIPE computes the MMSE estimate of random variable $\mathsf{u}_d$ from a $\mathcal{N}(0, \nu_d^q)$-corrupted observation $\widehat{q}_d$ under the prior

$$p_{\mathsf{u}_d}(u) = \beta_d \, p_0(u/\sigma)/\sigma + (1-\beta_d)\delta(u), \qquad (10)$$

in the limiting case that $\sigma \to \infty$. Here, $\beta_d \in (0, 1]$ is the prior probability that $\mathsf{u}_d \neq 0$ and the "slab" pdf $p_0(u)$ is continuous, finite, and non-zero at $u = 0$, but otherwise *arbitrary*. Note that, for fixed $\sigma$ and $\beta_d$, the MMSE estimator can be stated as

$$F_d(\widehat{q}_d; \nu_d^q) \triangleq \mathrm{E}\{\mathsf{u}_d | \widehat{q}_d; \nu_d^q\} = \frac{\int u \, p_{\mathsf{u}_d}(u)\mathcal{N}(u; \widehat{q}_d, \nu_d^q)du}{\int p_{\mathsf{u}_d}(u)\mathcal{N}(u; \widehat{q}_d, \nu_d^q)du}$$

$$= \frac{\int u p_0(u/\sigma)\mathcal{N}(u; \widehat{q}_d, \nu_d^q)du}{\int p_0(u/\sigma)\mathcal{N}(u; \widehat{q}_d, \nu_d^q)du + \sigma\frac{1-\beta_d}{\beta_d}\mathcal{N}(0; \widehat{q}_d, \nu_d^q)}. \qquad (11)$$

Since, with any fixed sparsity $\beta_d < 1$, the estimator (11) trivializes to $F_d(\widehat{q}_d; \nu_d^q) = 0 \; \forall \widehat{q}_d$ as $\sigma \to \infty$, we scale the sparsity with $\sigma$ as $\beta_d = \sigma/\big(\sigma + p_0(0)\sqrt{2\pi\nu_d^q}\exp(\omega)\big)$ for a tunable parameter $\omega \in \mathbb{R}$, in which case it can be shown that

$$F_d(\widehat{q}_d; \nu_d^q, \omega) \stackrel{\sigma \to \infty}{=} \frac{\widehat{q}_d}{1 + \exp(\omega - \frac{1}{2}|\widehat{q}_d|^2/\nu_d^q)}. \qquad (12)$$
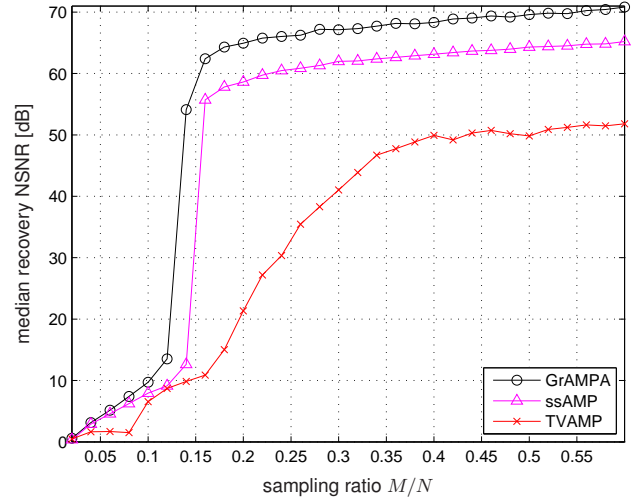


**Fig. 1**. Recovery of 0.05-sparse Bernoulli-Gaussian finite-difference signals from AWGN-corrupted measurements at $\mathsf{SNR} = 60$ dB.

## 3. NUMERICAL RESULTS

We now provide numerical results that compare GrAMPA with SNIPE denoising to several existing algorithms for cosparse analysis CS. In all cases, recovery performance was quantified using $\mathsf{NSNR} \triangleq \|\boldsymbol{x}\|^2/\|\widehat{\boldsymbol{x}} - \boldsymbol{x}\|^2$. Each algorithm was given perfect knowledge of relevant statistical parameters (e.g., noise variance) or in cases were an algorithmic parameter needed to be tuned (e.g., GrLASSO $\lambda$ or SNIPE $\omega$), the $\mathsf{NSNR}$-maximizing value was used.

### 3.1. Comparison to ssAMP and TV-AMP

We first replicate an experiment from the ssAMP paper [15]. Using the demonstration code for [15], we generated signal realizations $\boldsymbol{x} \in \mathbb{R}^N$ that yield BG 1D-finite-difference sequences $\mathbf{\Omega}\boldsymbol{x}$ with sparsity rate 0.05. Then we attempted to recover those signals from AWGN-corrupted observations $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x} + \boldsymbol{w} \in \mathbb{R}^M$, at an $\mathsf{SNR} \triangleq \|\mathbf{\Phi}\boldsymbol{x}\|_2^2/\|\boldsymbol{w}\|_2^2$ of 60 dB, generated with i.i.d Gaussian measurement matrices $\mathbf{\Phi}$.

Figure 1 shows median $\mathsf{NMSE}$ versus sampling ratio $M/N$ for ssAMP, TV-AMP, and GrAMPA, over 100 problem realizations. There we see GrAMPA uniformly outperforming ssAMP, which uniformly outperforms TV-AMP. We attribute the performance differences to choice of regularization: GrAMPA's SNIPE regularization is closer to $\ell_0$ than ssAMP's BG-based regularization, which is closer to $\ell_0$ than TV-AMP's $\ell_1$ regularization. We note the performance of GrAMPA in Fig. 1 is much better than that reported in [15] due to the misconfiguration of GrAMPA in [15].

### 3.2. Comparison to GAP: Synthetic cosparse recovery

We now compare GrAMPA with SNIPE denoising to Greedy Analysis Pursuit (GAP) [2] using an experiment from [2] that constructed $\mathbf{\Omega}^{\mathsf{T}} \in \mathbb{R}^{N \times D}$ as a random, almost-uniform, almost-tight frame and $\boldsymbol{x}$ as an exactly $L$-cosparse vector. The objective was then to recover $\boldsymbol{x}$ from noiseless measurements $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}$ using analysis operator $\mathbf{\Omega}$ and i.i.d Gaussian $\mathbf{\Phi}$. For this experiment, we used $N = 200$.

Figure 2 shows the empirical phase-transition curves (PTCs) for GAP and GrAMPA versus sampling ratio $\delta = M/N$ and uncer-
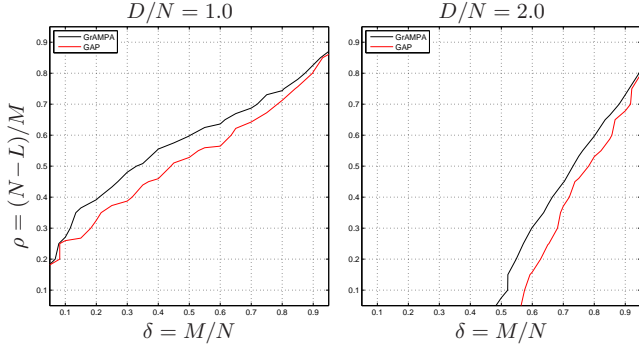
**Fig. 2**. Phase transition curves for recovery of $L$-cosparse $N$-length signals from $M$ noiseless measurements under i.i.d Gaussian $\boldsymbol{\Phi}$ and an $N \times D$ random, almost-uniform, almost-tight frame $\boldsymbol{\Omega}^{\mathsf{T}}$.
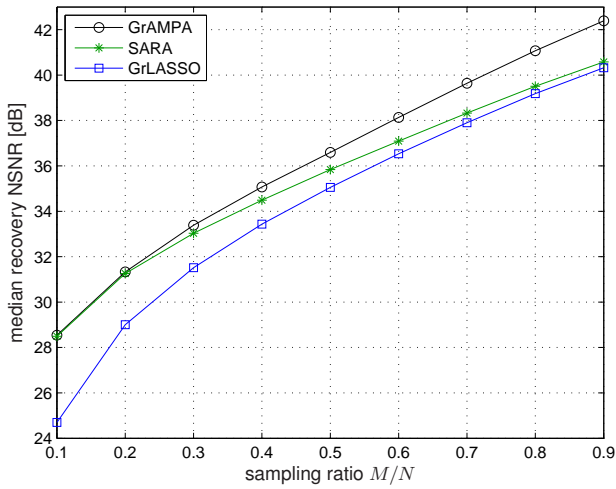


**Fig. 3**. Recovery of the $512 \times 512$ Lena image under a measurement SNR of 40 dB, spread-spectrum $\boldsymbol{\Phi}$, and Db1-8 concatenated $\boldsymbol{\Omega}$.

tainty ratio $\rho = (N - L)/M$. For points below the PTC, recovery was successful with high probability, while for points above the PTC, recovery was unsuccessful with high probability. Here, we defined "success" as NSNR $\geq 10^6$. Figure 2 shows that the PTC of GrAMPA is uniformly better than that of GAP. It also shows that, for both algorithms, the PTC approaches the feasibility boundary (i.e., $\rho = 1$) as $M/N \to 1$ but that, as the analysis operator becomes more overcomplete (i.e., $D/N$ increases), the PTC progressively weakens.

### 3.3. Compressive image recovery via sparsity averaging

Next, we repeat an experiment from [8], where the $N = 512 \times 512$ Lena image $\boldsymbol{x}$ was recovered from $M$ noisy complex-valued measurements $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{w}$ at SNR = 40 dB. The measurements were of the "spread spectrum" form: $\boldsymbol{\Phi} = \boldsymbol{MFC}$, where $\boldsymbol{C}$ was diagonal with random $\pm 1$ entries, $\boldsymbol{F}$ was an $N$-FFT, and $\boldsymbol{M} \in \{0, 1\}^{M \times N}$ contained rows of $\boldsymbol{I}_N$ selected uniformly at random. An overcomplete dictionary $\boldsymbol{\Psi} \in \mathbb{R}^{N \times 8N}$ was constructed from a horizontal concatenation of the first 8 Daubechies orthogonal DWT matrices, yielding the analysis operator $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{\mathsf{T}}$. The use of highly overcomplete concatenated dictionaries is dubbed "sparsity averaging" in [8].

Figure 3 shows median NSNR (over 30 Monte-Carlo trials) versus sampling ratio $M/N$ for GrAMPA with SNIPE denoising; for
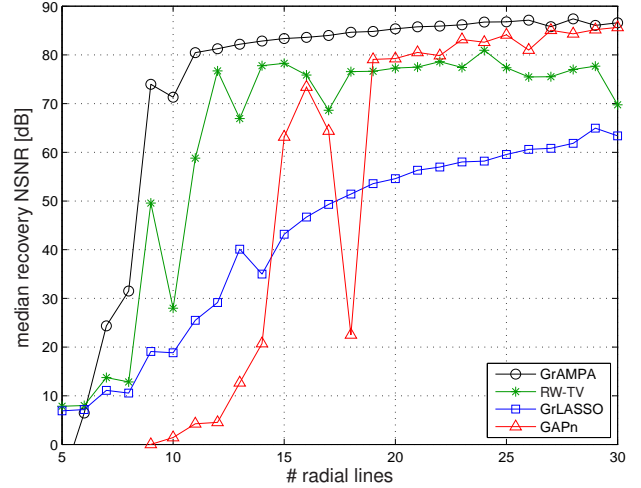


**Fig. 4**. Recovery of the $64 \times 64$ Shepp-Logan phantom from 2D FFT $\boldsymbol{\Phi}$ and 2D horizontal, vertical, and diagonal finite-difference $\boldsymbol{\Omega}$.

"SARA" from [8], which employs iteratively-reweighted-$\ell_1$ [7]; and for GrLASSO implemented via the "SOPT" Matlab code that accompanies [8], which employs Douglas-Rachford splitting [5]. All algorithms enforced non-negativity in the estimate. Figure 3 shows GrAMPA outperforming the other algorithms in NSNR at all sampling ratios $M/N$. Averaging over trials where all algorithms gave recovery NSNR $\geq 30$ dB, the runtimes of GrAMPA, GrLASSO, and SARA were 220, 255, and 2687 seconds, respectively.

### 3.4. Shepp-Logan phantom recovery via 2D finite-differences

Finally, we investigated the recovery of the $N = 64 \times 64$ Shepp-Logan Phantom image from 2D Fourier radial-line measurements $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{w}$ at SNR = 80 dB, using an analysis operator $\boldsymbol{\Omega}$ composed of horizontal, vertical, diagonal, and anti-diagonal 2D finite differences, as described in the noise-tolerant GAP paper [32].

Figure 4 plots median recovery NSNR (over 11 Monte-Carlo trials) versus number of radial lines for GrAMPA with SNIPE denoising, GAPn [32], the "RW-TV" approach from [8], which employs iteratively-weighted-$\ell_1$ [7], and GrLASSO, implemented using the Douglas-Rachford based "SOPT" Matlab code from [8]. The figure shows that GrAMPA achieved the best phase transition and also the best NSNR (for all numbers of radial lines above 6). Averaging over trials where all algorithms gave recovery NSNR $\geq 30$ dB, the runtimes of GrAMPA, GrLASSO, RW-TV, and GAP were 0.28, 1.8, 9.7, and 30.1 seconds, respectively.

### 4. CONCLUSIONS

In this work, we proposed the "Generalized AMP for Analysis CS" (GrAMPA) algorithm, a new AMP-based approach to analysis CS that can be used with a wide range of loss functions, regularization terms, and analysis operators. In addition, we proposed the "Sparse Non-informative Parameter Estimator" (SNIPE), an $\ell_0$-like soft thresholder that corresponds to the MMSE denoiser for a spike-and-slab distribution with an infinite-variance slab. Numerical experiments comparing GrAMPA with SNIPE to several other recently proposed analysis-CS algorithms show improved recovery performance and excellent runtime. Online tuning of the SNIPE parameter $\omega$ will be considered in future work.

# 5. REFERENCES

[1] Michael Elad, Peyman Milanfar, and Ron Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, pp. 947–968, 2007.

[2] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Computational Harmonic Anal.*, vol. 34, no. 1, pp. 30–56, Jan. 2013.

[3] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[4] Ryan J. Tibshirani, "Solution path of the generalized lasso," *Ann. Statist.*, vol. 39, no. 3, pp. 1335–1371, 2011.

[5] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 6564–574, Dec. 2007.

[6] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 1–39, 2011.

[7] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted $\ell_1$ minimization," *J. Fourier Anal. App.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[8] Rafael E. Carrillo, Jason D. McEwen, Dimitri Van De Ville, Jean-Philippe Thiran, and Yves Wiaux, "Sparsity averaging for compressive imaging," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 591–594, 2013.

[9] R. Chartrand, E. Y. Sidky, and X. Pan, "Nonconvex compressive sensing for X-ray CT: An algorithm comparison," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 2013, pp. 665–669.

[10] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies, "Greedy-like algorithms for the cosparse analysis model," *arXiv:1207.2456*, July 2012.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

[12] 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, Jan. 2010, pp. 1–5.

[13] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Thy.*, Aug. 2011, pp. 2168–2172, (full version at *arXiv:1010.5141*).

[14] D. L. Donoho, I. M. Johnstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," *IEEE Trans. Inform. Theory*, vol. 59, no. 6, June 2013.

[15] Jaewook Kang, Hyoyoung Jung, Heung-No Lee, and Kiseon Kim, "Spike-and-slab approximate message-passing for high-dimensional piecewise-constant recovery," *arXiv:1408.3930v1*, Aug. 2014, (Matlab codes at https://sites.google.com/site/jwkang10/ssAMP_algorithm_ver1_Aug2014.zip).

[16] J. S. Turek, I. Yavneh, and M. Elad, "On MAP and MMSE estimators for the co-sparse analysis model," *Digital Signal Process.*, vol. 28, pp. 57–74, 2014.

[17] Adel Javanmard and Andrea Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *Inform. Inference*, vol. 2, no. 2, pp. 115–144, 2013.

[18] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and iterative algorithms," in *Proc. IEEE Int. Symp. Inform. Thy.*, Boston, MA, June 2012, pp. 1643–1647, (full paper at *arXiv:1207.7321*).

[19] 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. Inform. Thy.*, July 2013, pp. 664–668, (full version at *arXiv:1301.6295*).

[20] F. Krzakala, A. Manoel, E. W. Tramel, and L. Zdeborová, "Variational free energies for compressed sensing," in *Proc. IEEE Int. Symp. Inform. Thy.*, July 2014, pp. 1499–1503.

[21] F. Caltagirone, F. Krzakala, and L. Zdeborová, "On convergence of approximate message passing," in *Proc. IEEE Int. Symp. Inform. Thy.*, July 2014, pp. 1812–1816.

[22] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of generalized approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inform. Thy.*, July 2014, pp. 236–240, (full version at *arXiv:1402.3210*).

[23] Jeremy Vila, Philip Schniter, Sundeep Rangan, Florent Krzakala, and Lenka Zdeborová, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *IEEE ICASSP*, 2015, to appear.

[24] S. Rangan, A. K. Fletcher, P. Schniter, and U. Kamilov, "Inference for generalized linear models via alternating directions and Bethe free energy minimization," *arXiv:1501:01797*, Jan. 2015.

[25] E. Esser, X. Zhang, and T. F. Chan, "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sci.*, vol. 3, no. 4, pp. 1015–1046, 2010.

[26] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, pp. 120–145, 2011.

[27] B. He and X. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective," *SIAM J. Imaging Sci.*, vol. 5, pp. 119–149, 2012.

[28] M. Borgerding and P. Schniter, "Generalized approximate message passing for the cosparse analysis model," *arXiv:1312.3968*, 2013, (Matlab codes at http://www2.ece.ohio-state.edu/~schniter/GrAMPA).

[29] U. S. Kamilov, V. K. Goyal, and S. Rangan, "Message-passing de-quantization with applications to compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6270–6281, Dec. 2012.

[30] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," in *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Oct. 2012, pp. 815–822.

[31] D. L. Snyder, A. M. Hammond, and R. L. White, "Image recovery from data acquired with a charge-coupled-device camera," *J. Am. Statist. Assoc.*, vol. 10, pp. 1014–1023, 1993.

[32] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "Recovery of cosparse signals with greedy analysis pursuit in the presence of noise," in *Proc. IEEE Workshop Comp. Adv. Multi-Sensor Adaptive Process.*, Puerto Rico, Dec. 2011.