

An Empirical-Bayes Approach to Recovering Linearly Constrained Non-Negative Sparse Signals

Jeremy Vila and Philip Schniter

Dept. of ECE, The Ohio State University, Columbus, OH 43210. (Email: vilaj@ece.osu.edu, schniter@ece.osu.edu)

Abstract—We consider the recovery of an (approximately) sparse signal from noisy linear measurements, in the case that the signal is a priori known to be non-negative and obeys certain linear equality constraints. For this, we propose a novel empirical-Bayes approach that combines the Generalized Approximate Message Passing (GAMP) algorithm with the expectation maximization (EM) algorithm. To enforce both sparsity and non-negativity, we employ an i.i.d Bernoulli non-negative Gaussian mixture (NNGM) prior and perform approximate minimum mean-squared error (MMSE) recovery of the signal using sum-product GAMP. To learn the NNGM parameters, we use the EM algorithm with a suitable initialization. Meanwhile, the linear equality constraints are enforced by augmenting GAMP’s linear observation model with noiseless pseudo-measurements. Numerical experiments demonstrate the state-of-the-art mean-squared-error and runtime of our approach.¹

I. INTRODUCTION

We consider the recovery of an (approximately) sparse signal $\mathbf{x} \in \mathbb{R}^N$ from the noisy linear measurements

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

where \mathbf{A} is a known sensing matrix, \mathbf{w} is additive white Gaussian noise (AWGN), and M may be $\ll N$. In this paper, we focus on non-negative signals (i.e., $x_n \geq 0 \forall n$) that obey linear equality constraints $\mathbf{B}\mathbf{x} = \mathbf{c} \in \mathbb{R}^P$. A notable example is the *simplex* constraint, i.e., $\mathbf{x} \in \Delta^+ \triangleq \{\mathbf{x} \in \mathbb{R}^N : \mathbf{x} \geq 0, \mathbf{1}^\top \mathbf{x} = 1\}$, occurring in hyperspectral unmixing [1], portfolio optimization [2], density learning [3], and many other applications.

For this task, we propose a novel empirical-Bayes approach that combines truncated Gaussian-mixture models, the expectation maximization (EM) algorithm [4], and the Generalized Approximate Message Passing (GAMP) algorithm [5], summarized in Table I and contextualized in the sequel.

GAMP is a computationally efficient approach to (approximate) maximum a posteriori (MAP) or minimum mean-squared error (MMSE) inference of a random vector $\mathbf{x} \in \mathbb{R}^N$ with a known i.i.d prior pdf $p_X(\cdot)$ from a corrupted observation $\mathbf{y} \in \mathbb{R}^M$ of the linear transform outputs $\mathbf{z} \triangleq \mathbf{A}\mathbf{x} \in \mathbb{R}^M$, where $\{y_m\}$ are conditionally independent (given \mathbf{z}) with known likelihood $p_{Y|Z}(y|\cdot)$. GAMP, recently proposed by Rangan, generalizes Donoho, Maleki, and Montanari’s Approximate Message Passing (AMP) algorithm [6] from AWGN observations to arbitrary likelihoods $p_{Y|Z}(y|\cdot)$. AMP and GAMP are both derived from (Taylor-series and central-limit-theorem based) approximations of loopy belief propagation, and yield

inputs:	$p_X(\cdot), p_{Y Z}(y_m \cdot), \{A_{mn}\}, T_{\max}, \tau_{\text{gamp}}$
definitions:	
	$p_{Z P}(z_m \hat{p}_m; \mu_m^p) \triangleq \frac{p_{Y Z}(y_m z_m) \mathcal{N}(z_m; \hat{p}_m, \mu_m^p)}{\int_z p_{Y Z}(y_m z) \mathcal{N}(z; \hat{p}_m, \mu_m^p)}$ (D1)
	$p_{X R}(x_n \hat{r}_n; \mu_n^r) \triangleq \frac{p_X(x_n) \mathcal{N}(x_n; \hat{r}_n, \mu_n^r)}{\int_x p_X(x) \mathcal{N}(x; \hat{r}_n, \mu_n^r)}$ (D2)
initialize:	
	$\forall n : \hat{x}_n(1) = \int_x x p_X(x)$ (I1)
	$\forall n : \mu_n^x(1) = \int_x x - \hat{x}_n(1) ^2 p_X(x)$ (I2)
	$\forall m : \hat{s}_m(0) = 0$ (I3)
for $t = 1 : T_{\max}$,	
	$\forall m : \mu_m^p(t) = \sum_{n=1}^N A_{mn} ^2 \mu_n^x(t)$ (R1)
	$\forall m : \hat{p}_m(t) = \sum_{n=1}^N A_{mn} \hat{x}_n(t) - \mu_m^p(t) \hat{s}_m(t-1)$ (R2)
	$\forall m : \mu_m^z(t) = \text{var}\{Z P = \hat{p}_m(t); \mu_m^p(t)\}$ (R3)
	$\forall m : \hat{z}_m(t) = \text{E}\{Z P = \hat{p}_m(t); \mu_m^p(t)\}$ (R4)
	$\forall m : \mu_m^s(t) = (1 - \mu_m^z(t)/\mu_m^p(t))/\mu_m^p(t)$ (R5)
	$\forall m : \hat{s}_m(t) = (\hat{z}_m(t) - \hat{p}_m(t))/\mu_m^s(t)$ (R6)
	$\forall n : \mu_n^r(t) = (\sum_{m=1}^M A_{mn} ^2 \mu_m^s(t))^{-1}$ (R7)
	$\forall n : \hat{r}_n(t) = \hat{x}_n(t) + \mu_n^r(t) \sum_{m=1}^M A_{mn}^* \hat{s}_m(t)$ (R8)
	$\forall n : \mu_n^x(t+1) = \text{var}\{X R = \hat{r}_n(t); \mu_n^r(t)\}$ (R9)
	$\forall n : \hat{x}_n(t+1) = \text{E}\{X R = \hat{r}_n(t); \mu_n^r(t)\}$ (R10)
	if $\sum_{n=1}^N \hat{x}_n(t+1) - \hat{x}_n(t) ^2 < \tau_{\text{gamp}} \sum_{n=1}^N \hat{x}_n(t) ^2$, break (R11)
end	
outputs:	$\{\hat{z}_m(t), \mu_m^z(t)\}, \{\hat{r}_n(t), \mu_n^r(t)\}, \{\hat{x}_n(t+1), \mu_n^x(t+1)\}$

TABLE I. THE GAMP ALGORITHM FROM [5]

computationally simple “first-order” algorithms that admit rigorous analysis in the large-system limit (i.e., $M, N \rightarrow \infty$ for fixed ratio M/N) for i.i.d zero-mean sub-Gaussian \mathbf{A} [7] and that have been demonstrated to work well for many deterministic matrices [8]. As we shall see, the generalization that GAMP affords over AMP is essential for our approach to enforcing the linear equality constraints $\mathbf{B}\mathbf{x} = \mathbf{c}$.

The GAMP algorithm requires specification of both the signal prior p_X and the likelihood $p_{Y|Z}$, which are typically unknown in practice. Recently, it was proposed to model p_X as Bernoulli-Gaussian-Mixture (BGM) with deterministic unknown parameters, and $p_{Y|Z}$ as Gaussian with deterministic unknown variance, and then compute (approximate) maximum-likelihood estimates of these parameters using the EM algorithm [9]. This combination of Bayesian and frequentist techniques is usually referred to as “empirical Bayes” [10]. Extensive empirical evidence showed that the resulting EM-GM-AMP algorithm offers state-of-the-art mean-squared error (MSE) and runtime for large random \mathbf{A} .

In this work, we extend the EM-GM-AMP approach from [9] to support non-negativity and linear equality constraints on \mathbf{x} . In the sequel, we provide details on our new approach and then present empirical evidence of its performance using phase-transition curves for the recovery of simplex signals and a sparse non-negative image-recovery experiment.

¹This work has been supported in part by NSF grants IIP-0968910, CCF-1018368, CCF-1218754, and by DARPA/ONR grant N66001-10-1-4090.

II. AUGMENTED MEASUREMENT MODEL

To enforce the linear equality constraint $\mathbf{B}\mathbf{x} = \mathbf{c}$ using GAMP, we extend the observation model (1) into

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{c} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{w} \\ 0 \end{bmatrix} \quad (2)$$

and exploit the fact that GAMP supports a likelihood function that varies with the measurement index m . Assuming the elements of \mathbf{w} are i.i.d Gaussian with variance ψ , and defining $\bar{\mathbf{y}} \triangleq [\bar{\mathbf{y}}_c]$, $\bar{\mathbf{A}} \triangleq [\bar{\mathbf{A}}]$, and $\bar{\mathbf{z}} \triangleq \bar{\mathbf{A}}\mathbf{x}$, the likelihood function corresponding to the augmented measurement model (2) is

$$p_{\bar{\mathbf{Y}}_m | \bar{\mathbf{Z}}_m}(\bar{\mathbf{y}}_m | \bar{\mathbf{z}}_m) = \begin{cases} \mathcal{N}(\bar{\mathbf{y}}_m; \bar{\mathbf{z}}_m, \psi) & m = 1, \dots, M \\ \delta(\bar{\mathbf{y}}_m - \bar{\mathbf{z}}_m) & m = M+1, \dots, M+P. \end{cases} \quad (3)$$

GAMP approximates the marginal posterior $p(\bar{\mathbf{z}}_m | \mathbf{y})$ by

$$p_{\bar{\mathbf{Z}}_m | P_m}(\bar{\mathbf{z}}_m | \hat{p}_m(t); \mu_m^p(t)) \propto p_{\bar{\mathbf{Y}}_m | \bar{\mathbf{Z}}_m}(\bar{\mathbf{y}}_m | \bar{\mathbf{z}}_m) \mathcal{N}(\bar{\mathbf{z}}_m; \hat{p}_m(t), \mu_m^p(t)) \quad (4)$$

at iteration t , where $\hat{p}_m(t)$ and $\mu_m^p(t)$ are given in Table I. The variance and mean of this posterior, used in lines (R3) and (R4) of Table I, are (dropping the t notation for brevity)

$$\hat{z}_m = \begin{cases} \hat{p}_m + \frac{\mu_m^p}{\mu_m^p + \psi}(\bar{\mathbf{y}}_m - \hat{p}_m) & m = 1, \dots, M \\ \bar{\mathbf{y}}_m & m = M+1, \dots, M+P \end{cases} \quad (5)$$

$$\mu_m^z = \begin{cases} \frac{\mu_m^p \psi}{\mu_m^p + \psi} & m = 1, \dots, M \\ 0 & m = M+1, \dots, M+P. \end{cases} \quad (6)$$

III. NON-NEGATIVE GAUSSIAN MIXTURE GAMP

We now describe how GAMP can be used with an i.i.d Bernoulli non-negative Gaussian mixture (NNGM) prior pdf of the form

$$p_X(x) = (1 - \lambda)\delta(x) + \lambda \sum_{\ell=1}^L \omega_\ell \mathcal{N}_+(x; \theta_\ell, \phi_\ell), \quad (7)$$

where $\mathcal{N}_+(\cdot)$ denotes the non-negative Gaussian pdf, i.e.,

$$\mathcal{N}_+(x; \theta, \phi) = \begin{cases} \frac{\mathcal{N}(x; \theta, \phi)}{\Phi_c(-\theta/\sqrt{\phi})} & x \geq 0 \\ 0 & x < 0 \end{cases}, \quad (8)$$

$\Phi_c(\cdot)$ is the complimentary cdf of the standard normal distribution, $\lambda \in (0, 1]$ is the sparsity rate, and ω_ℓ, θ_ℓ , and ϕ_ℓ are the weight, location, and scale, respectively, of the ℓ^{th} mixture component. For now, we treat the parameters $\mathbf{q} \triangleq [\lambda, \boldsymbol{\omega}, \boldsymbol{\theta}, \boldsymbol{\phi}, \psi]$ and the model order L as fixed and known.

GAMP approximates the marginal posterior $p(x_n | \mathbf{y})$ by

$$p_{X|R}(x_n | \hat{r}_n; \mu_n^r) \propto p_X(x_n) \mathcal{N}(x_n; \hat{r}_n, \mu_n^r) \quad (9)$$

where the quantities \hat{r}_n and μ_n^r (see Table I) vary with the GAMP iteration t . Given this posterior, the sum-product GAMP updates (R9) and (R10) are [11]

$$\hat{x}_n = \frac{\lambda}{\zeta_n} \sum_{\ell=1}^L \beta_{n,\ell} (\gamma_{n,\ell} + \sqrt{\nu_{n,\ell}} f(\alpha_{n,\ell})), \quad (10)$$

$$\mu_n^x = \frac{\lambda}{\zeta_n} \sum_{\ell=1}^L \beta_{n,\ell} (\nu_{n,\ell} g(\alpha_{n,\ell}) + (\gamma_{n,\ell} + \sqrt{\nu_{n,\ell}} f(\alpha_{n,\ell}))^2) - \hat{x}_n^2, \quad (11)$$

in terms of the normalization factor

$$\zeta_n \triangleq (1 - \lambda) \mathcal{N}(0; \hat{r}_n, \mu_n^r) + \lambda \sum_{\ell=1}^L \beta_{n,\ell}, \quad (12)$$

the $(\hat{r}_n, \mu_n^r, \mathbf{q})$ -dependent quantities

$$\alpha_{n,\ell} \triangleq \frac{-\gamma_{n,\ell}}{\sqrt{\nu_{n,\ell}}} \quad (13)$$

$$\gamma_{n,\ell} \triangleq \frac{\hat{r}_n / \mu_n^r + \theta_\ell / \phi_\ell}{1 / \mu_n^r + 1 / \phi_\ell}, \quad (14)$$

$$\nu_{n,\ell} \triangleq \frac{1}{1 / \mu_n^r + 1 / \phi_\ell} \quad (15)$$

$$\beta_{n,\ell} \triangleq \frac{\omega_\ell \mathcal{N}(\hat{r}_n; \theta_\ell, \mu_n^r + \phi_\ell) \Phi_c(\alpha_{n,\ell})}{\Phi_c(-\theta_\ell / \sqrt{\phi_\ell})}, \quad (16)$$

and the functions

$$f(a) \triangleq \frac{\varphi(a)}{\Phi_c(a)} \quad (17)$$

$$g(a) \triangleq 1 - f(a)(f(a) - a), \quad (18)$$

where $\varphi(\cdot)$ is the pdf of the standard normal distribution. From (9), it is straightforward to show that the corresponding posterior support probabilities $\pi_n \triangleq \Pr\{x_n \neq 0 | \mathbf{y}; \mathbf{q}\}$ are

$$\pi_n = \frac{1}{1 + \left(\frac{\lambda}{1-\lambda} \sum_{\ell=1}^L \beta_{n,\ell} \right)^{-1}}. \quad (19)$$

IV. EM PARAMETER LEARNING

Since the parameters $\mathbf{q} \triangleq [\lambda, \boldsymbol{\omega}, \boldsymbol{\theta}, \boldsymbol{\phi}, \psi]$ that best “fit” the true signal and noise distributions are typically unknown, we propose to learn them using an EM procedure [4]. The EM algorithm is an iterative technique that is guaranteed to converge to a local maximum (or a saddle point) of the likelihood $p(\mathbf{y}; \mathbf{q})$. We choose the “hidden” data to be \mathbf{x} , resulting in the iteration- i EM update

$$\mathbf{q}^{i+1} = \arg \max_{\mathbf{q}} \mathbb{E}\{\ln p(\mathbf{x}, \mathbf{y}; \mathbf{q}) | \mathbf{y}; \mathbf{q}^i\}. \quad (20)$$

For reasons of tractability, we evaluate the “arg max” in (20) one component at a time, while holding the others fixed, as in the “incremental” variant of EM from [12], and we use GAMP’s approximate posteriors to evaluate the expectation in (20). With this approximation, the updates become² [11]

$$\theta_k^{i+1} = \frac{\sum_{n=1}^N \pi_n \bar{\beta}_{n,k} (\gamma_{n,k} + \sqrt{\nu_{n,k}} f(\alpha_{n,k}))}{\sum_{n=1}^N \pi_n \bar{\beta}_{n,k}} \quad (21)$$

$$\phi_k^{i+1} = \frac{\sum_{n=1}^N \pi_n \bar{\beta}_{n,k} ((\gamma_{n,k} + \sqrt{\nu_{n,k}} f(\alpha_{n,k})) - \theta_k)^2 + \nu_{n,k} g(\alpha_{n,k})}{\sum_{n=1}^N \pi_n \bar{\beta}_{n,k}} \quad (22)$$

$$\omega_k^{i+1} = \frac{\sum_{n=1}^N \pi_n \bar{\beta}_{n,k}}{\sum_{n=1}^N \pi_n}, \quad (23)$$

where $\bar{\beta}_{n,k} \triangleq \beta_{n,k} / \sum_{\ell} \beta_{n,\ell}$. Because all quantities needed for the EM updates are by-products of sum-product GAMP, the EM approach does not substantially increase the complexity of our approach.

²The EM updates for the sparsity rate λ and noise variance ψ are identical to those given in [9].

A good EM initialization is essential to avoiding bad local minima. We propose to set the initial sparsity rate at

$$\lambda^0 = \frac{M}{N} \rho_{SE}\left(\frac{M}{N}\right), \quad (24)$$

where $\rho_{SE}(\cdot)$ is the theoretical noiseless phase-transition-curve (PTC) for ℓ_1 recovery of sparse non-negative (SNN) signals, shown to have the closed-form expression

$$\rho_{SE}(\delta) = \max_{c \geq 0} \frac{1 - (1/\delta)[(1+c^2)\Phi(-c) - c\varphi(c)]}{1 + c^2 - [(1+c^2)\Phi(-c) - c\varphi(c)]} \quad (25)$$

in [13], where $\Phi(\cdot)$ and $\varphi(\cdot)$ denote the cdf and pdf of the standard normal distribution. We then propose to set the initial values of the NNGM weights $\{\omega_\ell\}$, locations $\{\theta_\ell\}$, and scales $\{\phi_\ell\}$ at the values that best fit the uniform pdf on $[0, 1]$ (which can be computed offline using the standard EM-based approach described in [14, p. 435]). We propose to set the initial noise variance at $\psi^0 = \|\mathbf{y}\|_2^2 / ((\text{SNR} + 1)M)$, where, without knowledge of the true $\text{SNR} \triangleq \|\mathbf{A}\mathbf{x}\|_2^2 / \|\mathbf{w}\|_2^2$, we suggest using the value $\text{SNR} = 100$.

V. NUMERICAL RESULTS

A. Phase Transition Curves

First, we present empirically generated phase-transition curves (PTCs) for the recovery of K -sparse N -length simplex signals from M noiseless measurements.

To evaluate each PTC, we fixed $N = 500$ and constructed a 20×20 uniformly spaced grid on the $\frac{M}{N}$ -versus- $\frac{K}{M}$ plane over the ranges $\frac{M}{N} \in [0.05, 1]$ and $\frac{K}{M} \in [0.05, 1]$. At each grid point, we drew $R = 100$ independent realizations of the pair (\mathbf{A}, \mathbf{x}) , where $\mathbf{A} \in \mathbb{R}^{M \times N}$ was constructed with i.i.d $\mathcal{N}(0, M^{-1})$ entries and $\mathbf{x} \in \mathbb{R}^N$ was constructed with K nonzero elements $\{\mathbf{x}_k\}_{k=1}^K$ (placed uniformly at random), drawn i.i.d from a symmetric Dirichlet distribution with concentration parameter $a > 0$, whose pdf can be written as

$$p(\mathbf{x}_1, \dots, \mathbf{x}_{K-1}) = \begin{cases} \frac{\Gamma(aK)}{\Gamma(a)^K} \prod_{k=1}^K \mathbf{x}_k^{a-1}, & \mathbf{x}_k \in [0, 1] \\ 0 & \text{else} \end{cases} \quad (26a)$$

$$p(\mathbf{x}_K | \mathbf{x}_1, \dots, \mathbf{x}_{K-1}) = \delta(1 - \mathbf{x}_1 - \dots - \mathbf{x}_K) \quad (26b)$$

where $\Gamma(\cdot)$ is the gamma function. Note that (26) enforces the simplex constraint. For the r^{th} realization of (\mathbf{A}, \mathbf{x}) , we then recovered \mathbf{x} from the noiseless observation $\mathbf{y} = \mathbf{A}\mathbf{x}$ and defined the recovery $\hat{\mathbf{x}}$ to be “successful” if $\text{NMSE} \triangleq \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 / \|\mathbf{x}\|_2^2 < 10^{-6}$. Using $S_r = 1$ to mark a success and $S_r = 0$ otherwise, the average success rate was then computed via $\bar{S} \triangleq \frac{1}{R} \sum_{r=1}^R S_r$, and the empirical PTC was then plotted as the $\bar{S} = 0.5$ level-curve using Matlab’s `contour` command.

Figures 1–2 show the empirical PTCs for symmetric Dirichlet distributions with parameter $a = 1$ (i.e., uniformly distributed over the simplex) and $a = 100$ (i.e., $\mathbf{x}_k \approx \frac{1}{K} \forall k$), respectively, using the proposed “simplex EM-NNGM-AMP” approach, greedy selection and simplex projection (GSSP) [3], and Matlab’s `lsqnonneg` (using the augmented model (2)). In addition, Figures 1–2 show the theoretical SNN ℓ_1 -recovery PTC $\rho_{SE}(\frac{M}{N})$ from [13], given by (25). For the GSSP algorithm, we initialized $\hat{\mathbf{x}}^0$ as the SPGL1 [15] solution (using the augmented model from (2)) and set the step size as $100/\|\mathbf{A}\|_F^2$, as it yielded the best overall PTCs.

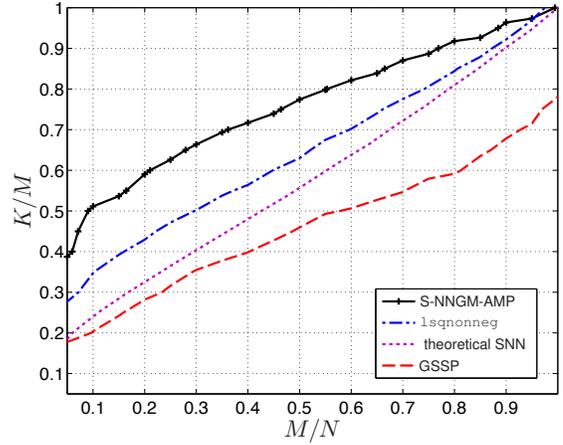


Fig. 1. Empirical PTCs and ℓ_1 -SNN theoretical PTC for noiseless recovery of length- $N = 500$, K -sparse, symmetric Dirichlet signals with concentration $a = 1$ from M measurements.

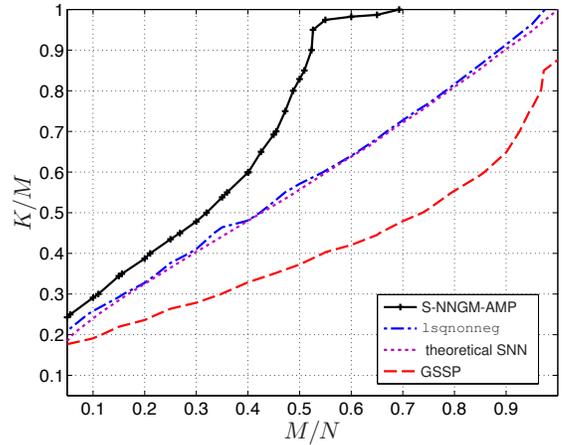


Fig. 2. Empirical PTCs and ℓ_1 -SNN theoretical PTC for noiseless recovery of length- $N = 500$, K -sparse, symmetric Dirichlet signals with concentration $a = 100$ from M measurements.

Figures 1–2 show that the proposed simplex-EM-NNGM-AMP approach yields empirical PTCs that are significantly better than those of the competing approaches. Its excellent performance is the result of three factors: i) the generality of the NNGM prior, ii) the ability of the EM approach to accurately learn the prior parameters, and iii) the ability of GAMP to exploit the learned prior. In fact, Fig. 2 shows simplex-NNGM-AMP accurately reconstructing K -sparse signals from only $M = K$ measurements even while $N > M$, thereby achieving a sort of “holy grail” in sparse reconstruction. The latter is possible because the signal \mathbf{x} is highly structured when $a = 100$, and because the proposed approach can learn and exploit that structure.

B. Non-negative Image Recovery

As a practical example, we experimented with the recovery of a sparse non-negative image. For this, we used the $N = 256 \times 256$ satellite image shown in Fig. 3, which contained $K = 6678$ nonzero pixels and $N - K = 58858$ zero-valued pixels, and thus was approximately 10% sparse. Linearly compressed measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ were collected under i.i.d Gaussian noise \mathbf{w} whose variance was selected to achieve $\text{SNR} = 60$

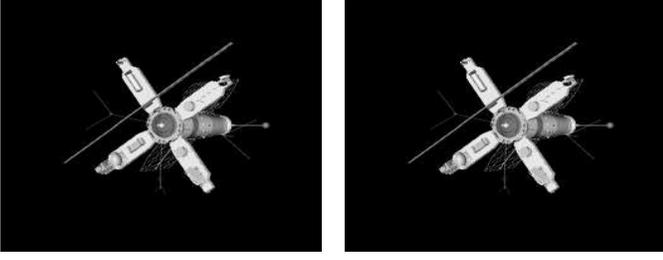


Fig. 3. Sparse non-negative image of a satellite: original image on left and EM-NNGM-AMP recovery at $\frac{M}{N} = \frac{1}{4}$ on right.

dB. Here, \mathbf{x} represents the (rasterized) image and \mathbf{A} a linear measurement operator configured as $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{S}$, where $\mathbf{\Phi} \in \{0, 1\}^{M \times N}$ was constructed from rows of the $N \times N$ identity matrix selected uniformly at random, $\mathbf{\Psi} \in \{-1, 1\}^{N \times N}$ was a Hadamard transform, and $\mathbf{S} \in \mathbb{R}^{N \times N}$ was a diagonal matrix with ± 1 diagonal entries chosen uniformly at random. Note that multiplication by \mathbf{A} can be executed using a fast binary algorithm, making it attractive for hardware implementation. For this experiment, no linear equality constraints exist and so the observation model was not augmented.

As a function of the sampling ratio $\frac{M}{N}$, Fig. 4 shows the NMSE and runtime averaged over $R = 100$ realizations of \mathbf{A} and \mathbf{w} for the proposed EM-NNGM-AMP in comparison to EM-GM-AMP from [9], genie-tuned non-negative LASSO via TFOCS [16],³ EM-tuned non-negative LASSO GAMP (EM-NNL-AMP) from [11],⁴ and genie-tuned standard LASSO implemented via SPGL1⁵ [15]. Results for `lsqnonneg` are not shown because its per-realization runtime exceeded three hours (since `lsqnonneg` treats \mathbf{A} as an explicit matrix).

Figure 4 (top) shows that the proposed EM-NNGM-AMP algorithm provided the most accurate signal recoveries for all undersampling ratios. Remarkably, its phase-transition occurred at $\frac{M}{N} \approx 0.25$, whereas that of the other algorithms occurred at $\frac{M}{N} \approx 0.35$. The gain of EM-NNGM-AMP over EM-GM-AMP can be attributed to the former's exploitation of signal non-negativity, whereas the gain of EM-NNGM-AMP over non-negative LASSO (either via EM-NNL-AMP or genie-tuned TFOCS) can be attributed to former's learning/exploitation of the true signal distribution. Finally, the gain of non-negative LASSO over standard LASSO can be attributed to the former's exploitation of signal non-negativity. In terms of runtime, EM-NNGM-AMP was about $3\times$ as fast as EM-GM-AMP, between $3\times$ to $15\times$ as fast as standard LASSO (via SPGL1), and $10\times$ to $20\times$ as fast as non-negative LASSO (via TFOCS).

REFERENCES

- [1] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical,

³Using EM-NNL-AMP's $\hat{\mathbf{x}}$, we ran TFOCS over an 11-point grid of hypothesized non-negative ℓ_1 penalty $\lambda \in \{0.5\|\mathbf{A}^T(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})\|_\infty, \dots, 2\|\mathbf{A}^T(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})\|_\infty\}$ and then reported the total runtime and best NMSE.

⁴EM-NNL-AMP runs max-product GAMP [5] under an AWGN likelihood and exponential signal prior whose noise variance and rate parameter, respectively, were tuned using an EM-based approach. (See [11] for details.)

⁵We ran SPGL1 in "BPDN mode," i.e., solving $\min_{\mathbf{x}} \|\mathbf{x}\|_1$ s.t. $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 < \sigma$ for hypothesized tolerances $\sigma^2 \in \{0.3, 0.6, \dots, 1.5\} \times M\psi$ and then reported the total runtime and best NMSE.

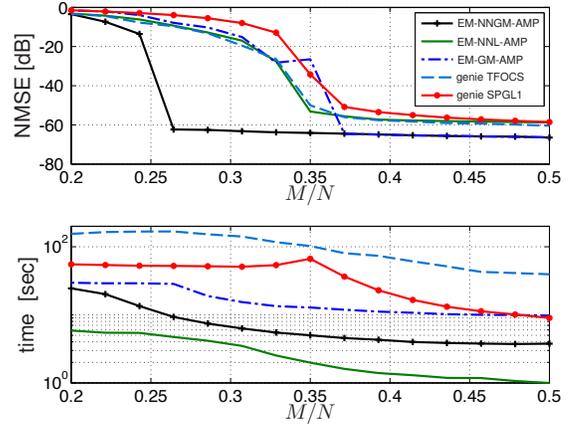


Fig. 4. Recovery NMSE (top) and runtime (bottom) versus $\frac{M}{N}$ for the sparse NN satellite image for the proposed EM-NNGM-AMP compared to EM-GM-AMP, non-negative LASSO via TFOCS and EM-NNL-AMP, and standard LASSO via SPGL1.

- statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, 2012.
- [2] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. New York: Wiley, 1991.
- [3] A. Kyrillidis, S. Becker, V. Cevher, and C. Koch, "Sparse projections onto the simplex," *arXiv:1206.1529*, Apr. 2013.
- [4] 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.
- [5] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Thy.*, (Saint Petersburg, Russia), pp. 2168–2172, Aug. 2011. (Full version at *arXiv:1010.5141*).
- [6] 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.
- [7] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *arXiv:1211.5164*, Nov. 2012.
- [8] H. Monajemi, S. Jafarpour, M. Gavish, Stat 330/CME 362 Collaboration, and D. L. Donoho, "Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices," *Proc. Nat. Acad. Sci.*, vol. 110, pp. 1181–1186, Jan. 2013.
- [9] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, pp. 4658–4672, Oct. 2013.
- [10] B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. New York: Cambridge University Press, 2010.
- [11] J. P. Vila and P. Schniter, "An empirical-Bayes approach to recovering linearly constrained non-negative sparse signals," *arXiv:1310.2806*, 2013.
- [12] R. Neal and G. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models* (M. I. Jordan, ed.), pp. 355–368, MIT Press, 1999.
- [13] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Nat. Acad. Sci.*, vol. 102, no. 27, pp. 9446–9451, 2005.
- [14] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2007.
- [15] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Scientific Comput.*, vol. 31, no. 2, pp. 890–912, 2008.
- [16] S. Becker, E. Candès, and M. M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Program. Comput.*, vol. 3, no. 3, pp. 165–218, 2011.