Hyperspectral Image Unmixing via Bilinear Generalized Approximate Message Passing

Jeremy Vila, Philip Schniter,[a] and Joseph Meola[b]

[a]Dept. of Electrical & Computer Engineering, The Ohio State University, Columbus, OH, US;
[b]Air Force Research Laboratory, Wright Patterson Air Force Base, OH, US

ABSTRACT

In hyperspectral unmixing, the objective is to decompose an electromagnetic spectral dataset measured over M spectral bands and T pixels, into N constituent material spectra (or "endmembers") with corresponding spatial abundances. In this paper, we propose a novel approach to hyperspectral unmixing (i.e., joint estimation of endmembers and abundances) based on loopy belief propagation. In particular, we employ the bilinear generalized approximate message passing algorithm (BiG-AMP), a recently proposed belief-propagation-based approach to matrix factorization, in a "turbo" framework that enables the exploitation of spectral coherence in the endmembers, as well as spatial coherence in the abundances. In conjunction, we propose an expectation-maximization (EM) technique that can be used to automatically tune the prior statistics assumed by turbo BiG-AMP. Numerical experiments on synthetic and real-world data confirm the state-of-the-art performance of our approach.

Keywords: Hyperspectral image unmixing, loopy belief propagation, approximate message passing

1. INTRODUCTION

In hyperspectral imaging (HSI), sensors capture M spectral measurements across each of $T = T_1 \times T_2$ pixels with the goal of identifying which materials are present throughout the scene. In the hyperspectral image unmixing problem, the macroscopic signal model[1] is traditionally assumed, where spectral measurements are a (noisy) linear combination of N material spectra, called *endmembers*, with per-pixel fractional *abundances*.[2] The goal is to then "unmix" the data into these endmembers and their corresponding abundance maps.

Mathematically, we write the received radiance data* $\boldsymbol{Y}$ as the bilinear model

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{A} + \boldsymbol{W} \in \mathbb{R}^{M \times T}, \tag{1}$$

where, with N materials, the columns of $\boldsymbol{S} \in \mathbb{R}_+^{M \times N}$ are the non-negative (NN) endmembers, the rows of $\boldsymbol{A} \in \mathbb{R}_+^{N \times T}$ are abundance maps, and $\boldsymbol{W}$ is noise. By definition, the abundances must satisfy the simplex constraint $a_{nt} \geq 0 \ \forall n, t$, and $\sum_{n=1}^{N} a_{nt} = 1 \ \forall t$. Fortunately, it is straightforward to enforce the sum-to-one constraint on the per-pixel abundances by augmenting[3,4] the system model (1) as

$$\underbrace{\begin{bmatrix} \boldsymbol{Y} \\ \mathbf{1}^{\mathsf{T}} \end{bmatrix}}_{\overline{\boldsymbol{Y}}} = \underbrace{\begin{bmatrix} \boldsymbol{S} \\ \mathbf{1}^{\mathsf{T}} \end{bmatrix}}_{\overline{\boldsymbol{S}}} \boldsymbol{A} + \begin{bmatrix} \boldsymbol{W} \\ \mathbf{0}^{\mathsf{T}} \end{bmatrix}, \tag{2}$$

where $\mathbf{1}^{\mathsf{T}}$ and $\mathbf{0}^{\mathsf{T}}$ are rows of ones and zeros, respectively.

In most real-world scenarios, however, there exists additional structure beyond endmember non-negativity and abundance simplex constraints. For instance, the amplitudes of each endmember are usually correlated, a phenomenon that we refer to as *spectral coherence*. Additionally, the abundances are typically *sparse*, in that each material usually inhabits only a fraction of the T pixels in the scene. Moreover, the abundance supports

Send correspondence to Jeremy Vila, E-mail: vilaj@ece.osu.edu.

*We assume throughout this paper that we are working directly on radiance data that the HSI sensors directly observe.

typically exhibit spatial structure in that, when a material inhabits a given pixel, it is more likely to inhabit neighboring pixels, a phenomenon that we refer to as *spatial coherence*. By incorporating spectral coherence and spatial coherence into our model, we can improve the recovery of both endmembers and abundances.

Traditional approaches to hyperspectral image unmixing apply an extraction step to estimate the endmembers, followed by inversion step to estimate the abundances. Here, the endmember extraction step leverages the so-called *pure pixel* assumption, where, for each of the materials, it is assumed that there is at least one pixel of the observed data $\boldsymbol{Y}$ containing only that material. Examples of such endmember extraction algorithms include N-FINDR,[5] vertex component analysis (VCA),[6] and fast separable non-negative matrix factorization (FSNMF).[7] The inversion step then traditionally employs fully constrained least squares (FCLS)[3,8] in order to enforce the simplex constraint on the per-pixel abundances. Unconstrained least squares (UCLS) would be computationally simpler, but the lack of simplex enforcement implies that the outputs would not truly be abundances. Notice that neither the extraction step nor the inversion step leverages the spatial or spectral coherence often exhibited in practice.

More recently, Bayesian approaches to *joint* estimation of the endmembers and abundances have also been proposed. For example, Bayesian Linear Unmixing (BLU)[9] employs priors that enforce both NN constraints on the endmembers and simplex constraints on the per-pixel abundances, and then approximates the posterior distributions $p_{\boldsymbol{S}|\boldsymbol{Y}}(\cdot)$ and $p_{\boldsymbol{A}|\boldsymbol{Y}}(\cdot)$ using Gibbs sampling, from which it returns (approximately) minimum mean-square error (MMSE) or maximum a posteriori (MAP) estimates. The spatially constrained unmixing (SCU)[10] algorithm expands upon BLU by employing a hierarchical Dirichlet process prior to exploit spatial coherence. Both BLU and SCU have been shown to outperform N-FINDR and VCA (plus FCLS) in certain conditions,[10] but at the cost of runtimes that are many orders-of-magnitude larger.

In this paper, we propose a novel Bayesian approach to HSI with complexity on par with traditional approaches (e.g., FCLS plus either N-FINDER, VCA, or FSNMF) yet with performance that far exceeds them. To do this, we model each material's spectral amplitudes as a Markov chain, and its abundances as structured-sparse with support governed by a Markov random field (MRF), and we attack the problem of joint endmember and abundance estimation through the framework of loopy belief propagation[11] (BP) on the factor graph shown in Fig. 1.

Our approach, which we refer to as HSI-AMP, is based on a partitioning of the factor graph into three sub-graphs: one that models spectral coherence via N Gauss-Markov chains, one that models spatial coherence using N binary MRFs, and one that models the bilinear structure of (2). While inference on the first two sub-graphs can be easily accomplished using standard methods,[12] the third poses a more difficult problem that we tackle using the recently proposed Bilinear Generalized Approximate Message Passing (BiG-AMP) algorithm.[13] BiG-AMP uses central limit theorem (CLT) and Taylor series approximations (which become accurate in the large system limit, i.e., $M, N, T \to \infty$ for fixed $\frac{M}{T}$ and $\frac{N}{T}$) to drastically reduce the complexity of exact loopy BP, and can be understood as an extension of the approximate message passing (AMP) technique[14–16] originally proposed for the linear observation models that arise in compressive sensing, to bilinear models like (2). The merging of AMP-based inference methods with more general BP inference methods to exploit, e.g., spatial and temporal coherence, goes by the name "turbo AMP" and has been exploited for compressive image retrieval,[17,18] compressive video retrieval,[19] and joint channel estimation and decoding,[20] all with state-of-the-art results. Furthermore, since in practice the parameters of the prior distributions are unknown, we automatically *tune* them using an expectation-maximization (EM) approach similar to that used in the state-of-the-art sparse reconstruction algorithm EM-GM-AMP.[21]

2. THE SIGNAL MODEL

As described in Section 1, we expect that each of the N abundance maps $\{a_{nt}\}_{t=1}^{T}$ exhibits spatial coherence. To model this structure, we introduce the binary support variables $d_{nt} \in \{0, 1\}$, where, $d_{nt} = 1$ indicates that the abundance coefficient a_{nt} is non-zero (with probability 1), i.e., "active." Assuming that the active signal coefficients are i.i.d conditional on $\{d_{nt}\}$, we write the (conditional, sparse) prior on a_{nt} as

$$p_{A|D}(a_{nt}|d_{nt}) = (1 - d_{nt})\delta(a_{nt}) + d_{nt}h_A(a_{nt}), \tag{3}$$

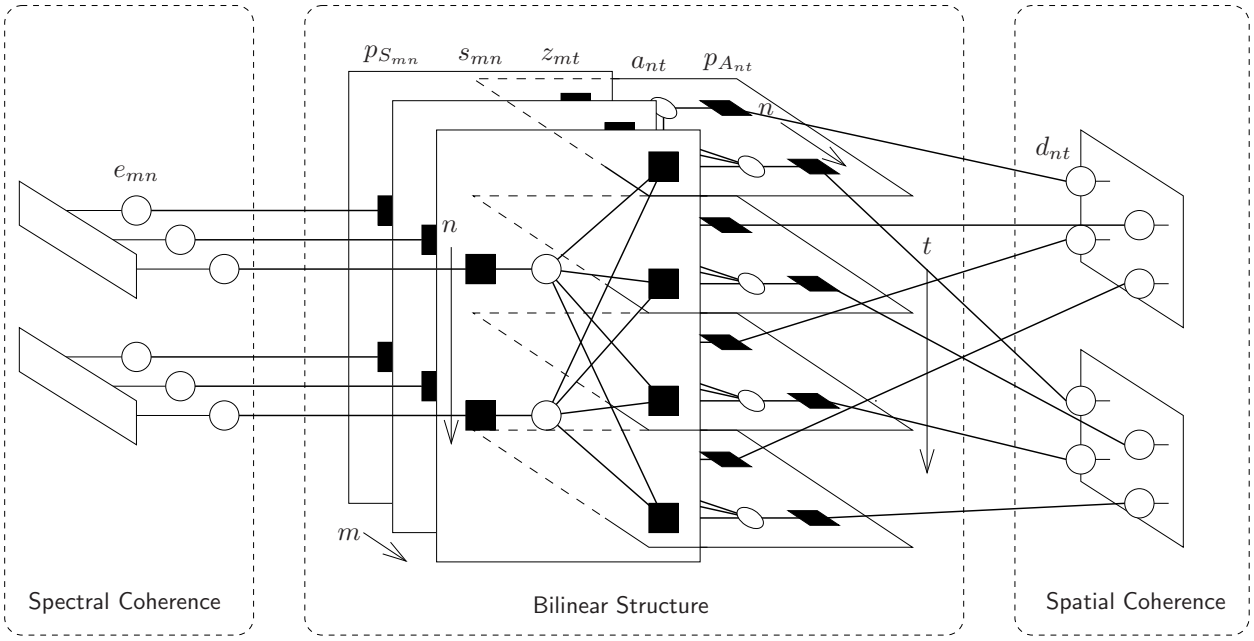

Figure 1. The factor graph for HSI-AMP for the toy-problem dimensions $M = 3$, $N = 2$, and $T = 4$. Circles represent random variables and dark squares represent pdf factors. Each elongated bar in the left sub-graph conglomerates the factors associated with an M-variable Markov chain, while each square in the right sub-graph conglomerates the factors associated with a $T_1 \times T_2$-pixel Markov random field (which is further detailed in Fig. 2).

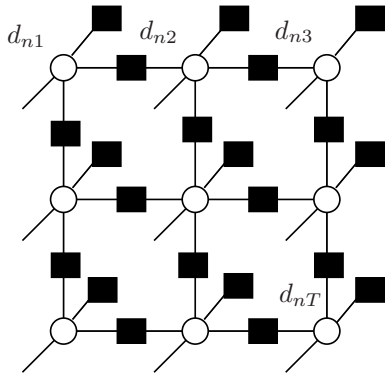

Figure 2. The MRF factor graph of size $T = 3 \times 3$ modeling the support of the n^{th} abundance map.

where $\delta(\cdot)$ is the Dirac delta and $h_A(\cdot)$ is the pdf on a_{nt} when active. Referring to $\{d_{nt}\}_{t=1}^T$ as the *sparsity pattern* of the n^{th} material, we exploit spatial coherence in the abundance maps by modeling $\{d_{nt}\}_{t=1}^T$ as a Markov Random Field (MRF), whose factor graph is illustrated in Fig. 2. Given this MRF structure, inference of $\{d_{nt}\}_{t=1}^T$ can be efficiently implemented using standard loopy BP techniques.[12]

Similarly, we model spectral coherence as correlation in the spectral amplitudes $\{s_{mn}\}_{m=1}^M$. To do this, we introduce the auxiliary amplitude sequence $\{e_{mn}\}_{m=1}^M$ for each material n, which we model as a stationary Gauss-Markov chain, i.e.,

$$p(e_{mn}|e_{(m-1)n}) = \mathcal{N}\big(e_{mn}; (1-\eta_n)e_{(m-1)n} + \eta_n\kappa_n, \eta_n^2\sigma_n^2\tfrac{2-\eta_n}{\eta_n}\big), \tag{4}$$

where $\kappa_n \in \mathbb{R}$ controls the mean of the process, σ_n^2 controls the variance, and $\eta_n \in [0,1]$ controls the correlation. Inference of each $\{e_{mn}\}_{m=1}^M$ can be accomplished using the classical forward-backward algorithm. The actual spectra $\{s_{mn}\}$ are then modeled as positively truncated versions of these auxiliary amplitudes, as described in Section 3.2.

3. THE HSI-AMP ALGORITHM

3.1 Turbo BiG-AMP

Our primary goal is to estimate the structured sparse abundances $\boldsymbol{A}$ and the amplitude correlated endmembers $\boldsymbol{S}$ given the observations $\boldsymbol{Y}$. Unfortunately, performing exact inference of the posterior $p(\boldsymbol{S}, \boldsymbol{A}|\boldsymbol{Y})$ is NP-hard because the factor graph contains many loops.[22] So, we settle for (approximate) MMSE estimates of $\boldsymbol{A}$ and $\boldsymbol{S}$.

As previously described, we use the "turbo AMP" approach[23] to separate the factor graph into an AMP-appropriate sub-graph and, in this case, two additional sub-graphs, one modeling spectral coherence and the other spatial coherence (recall Fig. 1). We then use AMP methods (in this case, BiG-AMP) to perform inference on the first subgraph (which, without approximation, is a daunting task), standard methods to perform inference on the other two subgraphs, and we exchange beliefs between subgraphs using sum-product-type updates until all beliefs "agree." As described in Section 2, our spectral coherence sub-graph consists of N independent M-element Gauss-Markov chains, and our spatial coherence sub-graph consists of N independent T-element binary MRFs, for which standard inference methods apply. For the bilinear structured sub-graph, we perform approximate loopy-BP using the recently-proposed BiG-AMP[13] algorithm. The BiG-AMP algorithm uses a series of CLT and Taylor series approximations to drastically reduce the complexity of (approximate) loopy BP. In the end, the complexity of BiG-AMP is dominated by four matrix multiplies (e.g., $\boldsymbol{SA}$) per iteration, with no more than a few hundred iterations.

3.2 BiG-AMP priors

When viewed in isolation, BiG-AMP assumes that the random variables $\{s_{mn}\}$ and $\{a_{mn}\}$ are statistically independent with specified priors, which we will refer to as "local" priors. In our turbo framework, these local priors will depend on the most recent messages coming from the two other sub-graphs and, as a consequence of the spatial and spectral dependencies in these other sub-graphs, we emphasize that the effective "global" priors on $\{s_{mn}\}$ and $\{a_{mn}\}$ are *not* independent. Still, to employ BiG-AMP on the inner sub-graph, we must specify the local priors on $\{s_{mn}\}$ and $\{a_{mn}\}$, as well as how these local priors are computed from the messages arriving from the outer sub-graphs. In choosing these local priors, we desire that 1) our priors to match to the true marginal distributions (e.g., enforcing non-negativity in $\{s_{mn}\}$ and $\{a_{mn}\}$), and that 2) they yield tractable BiG-AMP update steps.

With these considerations in mind, we assign the local prior on s_{mn} as

$$p_{S_{mn}}(s) = \mathcal{N}_+(s; \theta^s_{mn}, \phi^s_{mn}), \tag{5}$$

for which $\mathcal{N}_+$ refers to the non-negative Gaussian pdf

$$\mathcal{N}_+(x; \theta, \phi) \triangleq \begin{cases} 0 & x < 0 \\ \frac{\mathcal{N}(x;\theta,\phi)}{\Phi_c\left(\theta/\sqrt{\phi}\right)} & x \geq 0 \end{cases}, \tag{6}$$

where θ can be interpreted as a location parameter, ϕ is a scale parameter, and $\Phi_c(\cdot)$ refers to the complimentary cdf of the $\mathcal{N}(0,1)$ distribution. The parameters θ^s_{mn} and ϕ^s_{mn} are then set equal to the mean and variance, respectively, of the most recent message arriving from the e_{mn} node of the spectral coherence sub-graph.

We then choose the local prior on a_{nt} as a Bernoulli non-negative Gaussian mixture (BNNGM) pdf

$$p_{A_{nt}}(a) = (1 - \lambda^a_{nt})\delta(a) + \lambda^a_{nt} \sum_{\ell=1}^{L} \omega^a_{n\ell} \mathcal{N}_+(a; \theta^a_{n\ell}, \phi^a_{n\ell}), \tag{7}$$

where the sparsity rate λ^a_{nt} is set from the most recent messages from support node d_{nt} in the spatial coherence sub-graph, and the n^{th} material's NNGM parameters $\{\omega^a_{n\ell}, \theta^a_{n\ell}, \phi^a_{n\ell}\}_{\ell=1}^L$ are learned via the EM procedure described in Section 3.3. Notice that the assumed local prior enforces both sparsity and non-negativity in the abundances. We emphasize that unique NNGM parameters are learned for each of the N materials, since the marginal abundance distributions are likely to differ across materials.

Finally, we assume that the noise samples used to construct $\boldsymbol{W}$ in (1) are i.i.d Gaussian, i.e.,

$$p_W(w) = \mathcal{N}(w; 0, \psi), \tag{8}$$

with variance ψ learned through the EM procedure described in Section 3.3. Recall, however, that BiG-AMP operates on the augmented system model (2), which contains the unit-valued "pseudo" observations $\{\overline{y}_{M+1,t}\}_{t=1}^{T}$ used to enforce the sum-to-one constraint. In the end, BiG-AMP requires us to specify the scalar likelihood functions $p_{\overline{Y}_m|\overline{Z}_m}(\overline{y}_{mt}|\overline{z}_{mt})$ relating the noisy observations $\overline{y}_{mt} \triangleq [\overline{\boldsymbol{Y}}]_{mt}$ to the noiseless counterparts $\overline{z}_{mt} \triangleq [\overline{\boldsymbol{S}}\boldsymbol{A}]_{mt}$, for which we have

$$p_{\overline{Y}_m|\overline{Z}_m}(\overline{y}_{mt}|\overline{z}_{mt}) = \begin{cases} \mathcal{N}(\overline{y}_{mt}; \overline{z}_{mt}, \psi) = \mathcal{N}(y_{mt}; \overline{z}_{mt}, \psi) & m = 1, \ldots, M \text{ and } t = 1, \ldots, T \\ \delta(\overline{y}_{mt} - \overline{z}_{mt}) = \delta(1 - \overline{z}_{mt}) & m = M{+}1 \text{ and } t = 1, \ldots, T. \end{cases} \tag{9}$$

Importantly, the priors described in (5), (7), and (9) yield closed-form BiG-AMP updates.

3.3 EM learning of the prior parameters

In practice, the parameters $\boldsymbol{\Omega} = \{\psi, \{\omega_{n\ell}^a, \theta_{n\ell}^a, \phi_{n\ell}^a\}_{\forall n\ell}\}$ that govern BiG-AMP's local priors are unknown, and so we propose to learn them on-the-fly using an expectation-maximization (EM)[24] procedure, similar to what was done for the AMP-based sparse-reconstruction algorithm EM-GM-AMP.[4] The EM algorithm iterates

$$\boldsymbol{\Omega}^{i+1} = \arg\max_{\boldsymbol{\Omega}} \mathrm{E}\{\ln p(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{Y}; \boldsymbol{\Omega}) \,|\, \boldsymbol{Y}; \boldsymbol{\Omega}^i\} \tag{10}$$

with the goal of increasing a lower bound on the likelihood $p(\boldsymbol{Y}; \boldsymbol{\Omega})$ at each iteration. In our application of EM, we update $\boldsymbol{\Omega}$ one component at a time (while holding the others fixed), which is the well known "incremental" variant of EM.[25] Also, we use BiG-AMP's approximate marginal posteriors to evaluate the expectation in (10), and thus our approach is technically an approximate-EM algorithm. As a result of our re-use of BiG-AMP's posteriors, however, the complexity of the EM tuning procedure is a small fraction of BiG-AMP itself.

We then use standard methods to learn the parameters $\{\kappa_n, \eta_n, \sigma_n^2\}_{n=1}^{N}$ governing the Gauss-Markov chains[19] and the parameters governing the MRFs.[18]

3.4 Initializations

Since the EM algorithm may converge to a local, rather than global, maximum of the likelihood, care must be taken when initializing HSI-AMP.

First, we initialize the endmembers (i.e., $(\hat{\boldsymbol{S}})^0$) at the solutions provided by VCA.[6] Using these endmembers, we apply UCLS to compute the initial (approximate) abundance maps (i.e., $(\hat{\boldsymbol{A}})^0$). As for the endmembers' non-negative-Gaussian distributions, we initialize the location parameters at the solution provided by VCA, i.e., $(\boldsymbol{\theta}^s)^0 = (\hat{\boldsymbol{S}})^0$, and the scale parameters $(\boldsymbol{\phi}^s)^0$ to 1. For the abundances' BNNGM parameters, the initial sparsity rate is set at $(\lambda^a)^0 = \frac{1}{2}$, and the NNGM weights $\{\omega_{n\ell}^a\}$, locations $\{\theta_{n\ell}^a\}$, and scales $\{\phi_{n\ell}^a\}$ were initialized to best fit the uniform pdf on $[0, 1]$. Here, it is our experience that $L = 3$ suffices for the number of NNGM components. Finally, we initialize the noise variance as

$$\psi^0 = \frac{\|\boldsymbol{Y}\|_F^2}{(\mathsf{SNR}^0 + 1)MT} \tag{11}$$

where, without any user guidance on the true signal-to-noise ratio $\mathsf{SNR} \triangleq E\{|z_{mt}|^2\}/\psi$, we suggest $\mathsf{SNR}^0 = 100$.

4. RESULTS

4.1 Pure-Pixel Synthetic Data

To gain insights on HSI-AMP's performance, we first compared it to several other algorithms on a synthetic pure-pixel dataset. For this, a scene of $T = 50 \times 50$ pixels was partitioned into $N = 5$ equally sized vertical strips, each consisting of only a single material. The endmembers, representing dry grass, generic leaf, terra

Figure 3. Color image of the noiseless measurements $\boldsymbol{Z}$. Since the pixels are pure, each strip represents the color image of the $N = 5$ base endmembers. They are, in order from left to right: dry grass, generic leaf, terra cota, silver paint, and light gray shingle, as in Fig. 4.

	$\boldsymbol{S}$ Runtime	$\boldsymbol{A}$ Runtime	Total Runtime	$\mathsf{NMSE}_{\boldsymbol{S}}$	$\mathsf{NMSE}_{\boldsymbol{A}}$
HSI-AMP	-	-	5.35 sec	-57.1 dB	-37.3 dB
SCU	-	-	2808 sec	-30.6 dB	-20.5 dB
VCA + FCLS	0.05 sec	4.08	4.13 sec	-39.6 dB	-30.5 dB
VCA + UCLS	0.05 sec	0.0007 sec	0.05 sec	-39.6 dB	-12.0 dB
FSNMF + FCLS	0.002 sec	3.97 sec	3.97 sec	-25.3 dB	-12.5 dB
FSNMF + UCLS	0.002 sec	0.0008 sec	0.002 sec	-23.4 dB	-6.8 dB

Table 1. Runtime and endmember/abundance map MMSE recovery for various algorithms on the pure synthetic dataset with $\mathsf{SNR} = 30$ dB.

cota, silver paint, and light gray shingle, were taken from a reflectance library with $M = 121$ spectral bands from $400 - 1000$ nm. Finally, additive white Gaussian noise was added to yield a desired SNR. Figure 3 shows the resulting color image constructed from the red, green, and blue wavelengths of the noiseless measurements $\boldsymbol{Z} = \boldsymbol{S}\boldsymbol{A}$.

Averaging over $R = 10$ noise realizations, Table 1 shows the normalized mean-squared error (i.e., $\mathsf{NMSE}_{\boldsymbol{S}} \triangleq \|\boldsymbol{S} - \hat{\boldsymbol{S}}\|_F^2 / \|\boldsymbol{S}\|_F^2$ and $\mathsf{NMSE}_{\boldsymbol{A}} \triangleq \|\boldsymbol{A} - \hat{\boldsymbol{A}}\|_F^2 / \|\boldsymbol{A}\|_F^2$) of the estimated endmembers and abundances, as well as the corresponding algorithmic runtimes on a standard PC, at $\mathsf{SNR} = 30$ dB. (For the HSI-AMP and SCU algorithms, we reported only the total runtime, since estimation of $\boldsymbol{S}$ and $\boldsymbol{A}$ are done jointly.) For qualitative comparison, Fig. 4 shows the endmembers recovered by HSI-AMP, VCA, SCU, and FSNMF for a typical realization.

Table 1 shows HSI-AMP yielding significant gains in NMSE over all competing techniques in both endmember and abundance estimation. In particular, HSI-AMP outperformed the best competing technique (VCA+FCLS) by more than 16 dB in $\mathsf{NMSE}_{\boldsymbol{S}}$ and about 7 dB in $\mathsf{NMSE}_{\boldsymbol{A}}$. Moreover, HSI-AMP's runtime is comparable to VCA+FCLS (and FSNMF+FCLS), and 2-3 orders of magnitude faster than SCU, on this problem. That said, HSI-AMP's runtime is still 2-3 orders-of-magnitude slower than the combination of UCLS and VCA or FSNMF on this problem, although this speed comes at the cost of 25-30 dB worse $\mathsf{NMSE}_{\boldsymbol{A}}$. Not surprisingly, Fig. 4 shows that, in a qualitative sense, the endmembers recovered by HSI-AMP most accurately match the true endmembers in this experiment. Those recovered by VCA look good for the first three materials but not for the last two, those recovered by SCU look good only for the first material, and those recovered by FSNMF look very noisy throughout.

We attribute HSI-AMP's excellent performance on this experiment to its ability to *jointly* estimate endmembers and abundances, to exploit endmember *spectral coherence* and abundance *spatial coherence*, and to exploit the per-pixel abundance *sparsity*. VCA and FSNMF, while extremely quick for endmember extraction, suffer in endmember estimation quality, even in the presence of pure pixels, because they estimate endmembers separately from abundances and because they exploit neither spatial nor spectral coherence. As for SCU, we are unsure as to why it performed so poorly, but we conjecture that its Gibbs sampling procedure may be finicky and/or that its priors may not be well suited to exclusively pure pixels.

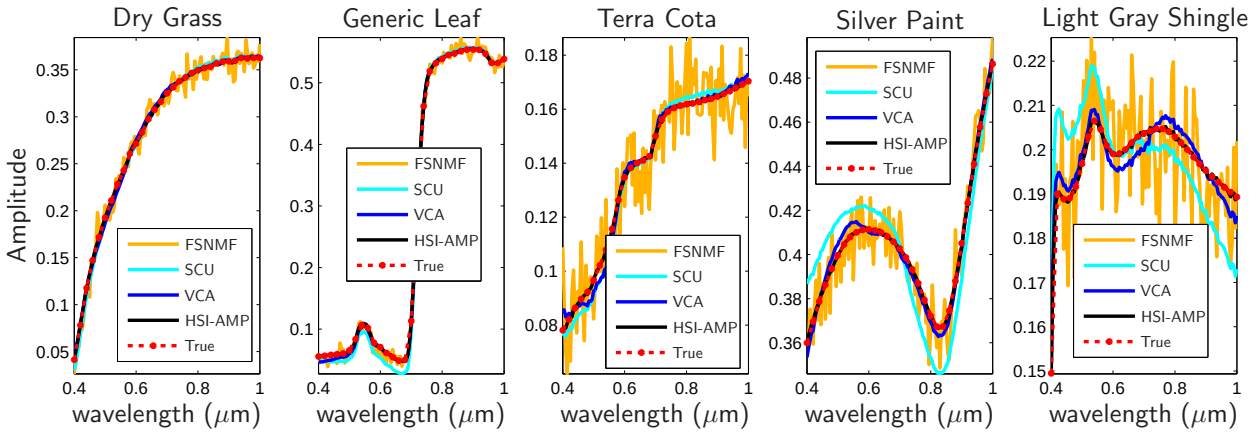

Figure 4. True endmember spectra (red) vs their estimations via HSI-AMP (black), VCA (blue), SCU (cyan), and FSNMF (orange), for the synthetic experiment under 30 dB.

4.2 SHARE 2012 dataset

We now test the same HSI unmixing algorithms on the SHARE 2012 dataset.[26] This data consisted of $M = 360$ spectral bands, ranging from 400-2450 nm, taken over a larger rural scene, from which we extracted a smaller scene of $T = 150 \times 100$ pixels. The smaller scene consisted of grass, trees, pavement, rooftops, and various targets, and is shown as a color image in Fig. 5(a) (and Fig. 5(b)). At the time of this paper's submission, ground truth data was unavailable.

Assuming that $N = 7$ materials were present,[†] Fig. 5 shows the recovered abundance maps for HSI-AMP, VCA+FCLS, FSNMF+FCLS, and SCU, respectively. In Fig. 5, we ordered the $N = 7$ abundance maps recovered by HSI-AMP arbitrarily. Then, for each competing algorithm, we proceeded from left to right over HSI-AMP's abundances, and selected the abundance that was most correlated to the given HSI-AMP abundance.

Without ground truth, we are limited to qualitative comparisons, such as checking whether each algorithm recovers what are surmised to be pure pixel materials. For example, looking at the pixels in the color image that seem to pertain exclusively to grass, HSI-AMP seems to best separate out these pixels (see the leftmost panel in Fig. 5(b)). Similarly, HSI-AMP also seems to best separate out the pixels corresponding to leafy trees (see the rightmost panel in Fig. 5(b)). Finally, the 4^{th} and 5^{th} panels in Fig. 5(b) show that HSI-AMP separated out two distinct targets in a parking lot, whereas VCA+FCLS (and the other algorithms) mixed those two materials together.

5. CONCLUSIONS

HSI unmixing is a powerful tool for estimating collection of constituent material endmembers and their fractional abundances. Traditional endmember extraction algorithms like VCA or FSNMF are based on the exploitation of pure pixels, which may not exist in real data, and do not exploit the spatial and spectral coherence that usually do exist. Moreover, although VCA and FSNMF run very quickly, they do not estimate abundances (only endmembers). If accurate abundance estimation is required, then the standard approach to doing so, FCLS, becomes the computational bottleneck, at which point the advantage of very fast endmember extraction could be questioned.

In this paper, we proposed HSI-AMP, a novel approach to HSI unmixing based on loopy belief propagation, that showed state-of-the-art joint endmember and abundance recovery in two preliminary experiments: a "synthetic" one that used abundances based on uniform strips of pure pixels and endmembers from a reflectance library, and the other based on the SHARE 2012 dataset. For the first, HSI-AMP showed NMSE improvements of 16 dB in endmember estimation and 7 dB in abundance estimation relative to the closest competitor (VCA+FCLS) while approximately preserving the runtime. For the second, a lack of ground truth prevents us

[†]Automatic selection of the model order N is an important topic for future research.

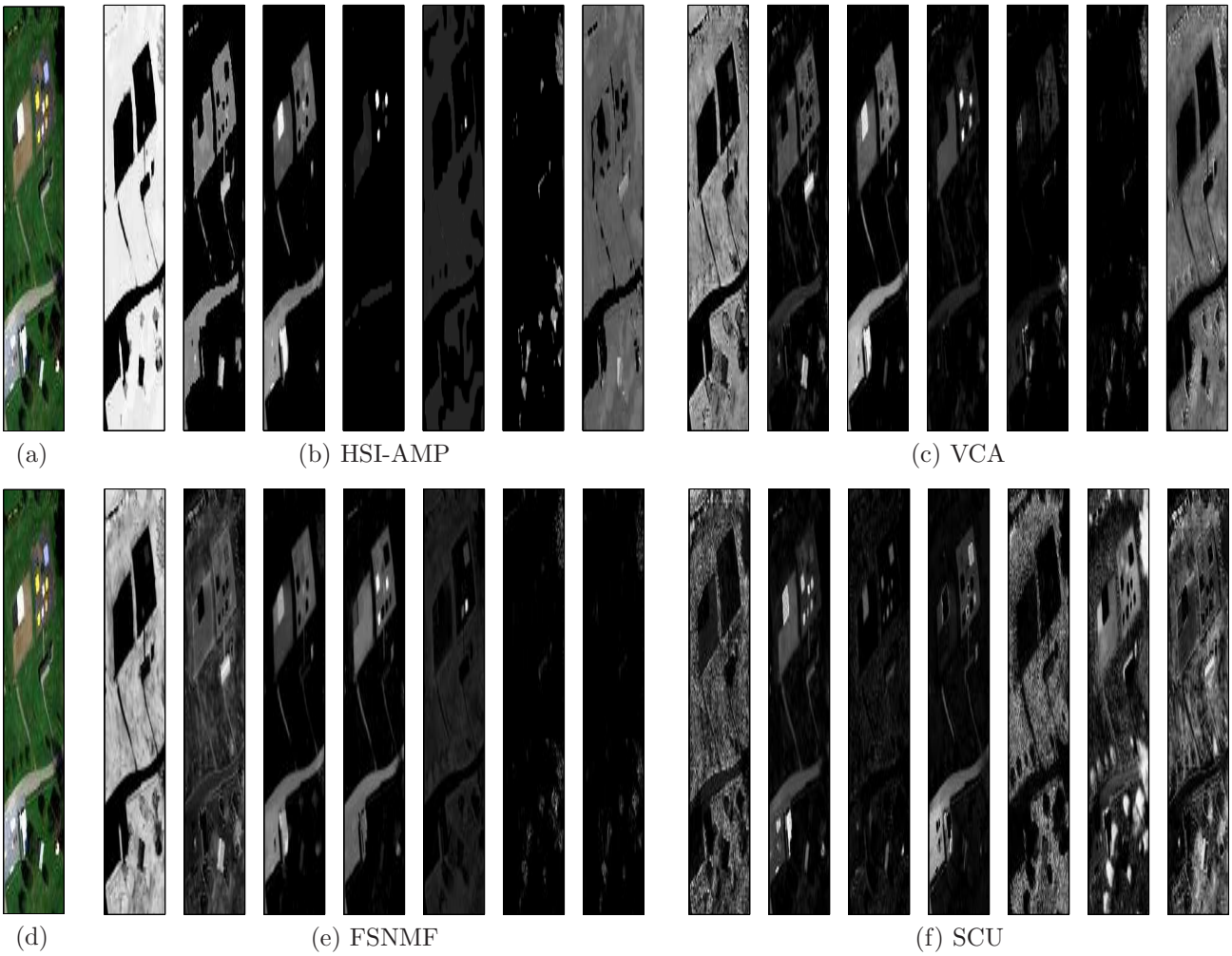

Figure 5. The $N = 7$ recovered abundance maps using the (b) HSI-AMP, (c) VCA, (e) FSNMF, and (f) SCU algorithms on the SHARE 2012 dataset. For visual comparison, (a) and (d) show the color image associated with the scene.

from making quantitative claims, but qualitatively HSI-AMP seemed to better extract various distinct materials present in the scene. We attribute HSI-AMP's success to its ability to jointly estimate $\boldsymbol{S}$ and $\boldsymbol{A}$ while leveraging spectral coherence in $\boldsymbol{S}$, and both sparsity and spatial coherence in $\boldsymbol{A}$, and automatically tuning its distributional parameters via the EM algorithm.

ACKNOWLEDGMENTS

This work has been supported in part by NSF-I/UCRC grant IIP-0968910 and NSF grant CCF-1218754. We'd also like to thank Nina Raqueno et al. at the Rochester Institute of Technology for providing the SHARE 2012 dataset, and Jason Parker for his assistance with BiG-AMP.

REFERENCES

[1] Johnson, P. E., Smith, M. O., Taylor-George, S., and Adams, J. B., "A semiempirical method for analysis of the reflectance spectra of binary mineral mixtures," *Journal of Geophysical Research* **88**, 3557–3561 (1983).

[2] Bioucas-Dias, J., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., and Chanussot, J., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **5**(2), 354–379 (2012).

[3] Heinz, D. and Chang, C.-I., "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on* **39**(3), 529–545 (2001).

[4] Vila, J. and Schniter, P., "Generalized approximate message passing with simplex constraints," in [*in preparation*], (2013).

[5] Winter, M. E., "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE* , 266–275 (1999).

[6] Nascimento, J. and Bioucas Dias, J., "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on* **43**(4), 898–910 (2005).

[7] Gillis, N. and Vavasis, S. A., "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *arXiv:1208.1237v2* (2012).

[8] Heylen, R., Burazerovic, D., and Scheunders, P., "Fully constrained least squares spectral unmixing by simplex projection," *Geoscience and Remote Sensing, IEEE Transactions on* **49**(11), 4112–4122 (2011).

[9] Dobigeon, N., Moussaoui, S., Coulon, M., Tourneret, J. Y., and Hero, A., "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *Signal Processing, IEEE Transactions on* **57**(11), 4355–4368 (2009).

[10] Mittelman, R., Dobigeon, N., and Hero, A., "Hyperspectral image unmixing using a multiresolution sticky hdp," *Signal Processing, IEEE Transactions on* **60**(4), 1656–1671 (2012).

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

[12] Li, S. Z., [*Markov Random Field Modeling in Image Analysis*], Springer, London, 3rd ed. (2009).

[13] Parker, J., Schniter, P., and Cevher, V., "Bilinear generalized approximate message passing," in [*in preparation*], (2013).

[14] Donoho, D. L., Maleki, A., and Montanari, A., "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.* **106**, 18914–18919 (Nov. 2009).

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

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

[17] Som, S. and Schniter, P., "Compressive imaging using approximate message passing and a Markov-tree prior," *IEEE Trans. Signal Process.* (to appear 2012). (see also arXiv:1108.2632).

[18] Som, S. and Schniter, P., "Approximate message passing for recovery of sparse signals with Markov-random-field support structure," presented at the *Internat. Conf. Mach. Learning*, (Bellevue, WA) (July 2011).

[19] Ziniel, J. and Schniter, P., "Dynamic compressive sensing of time-varying signals via approximate message passing," *arXiv:1205.4080* (May 2012).

[20] Schniter, P., "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.* **5**, 1462–1474 (Dec. 2011).

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

[22] Cooper, G. F., "The computational complexity of probabilistic inference using Bayesian belief networks," *Artificial Intelligence* **42**, 393–405 (1990).

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

[24] Dempster, A., Laird, N. M., and Rubin, D. B., "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.* **39**, 1–17 (1977).

[25] Neal, R. and Hinton, G., "A view of the EM algorithm that justifies incremental, sparse, and other variants," in [*Learning in Graphical Models*], Jordan, M. I., ed., 355–368, MIT Press (1999).

[26] Giannandrea, A., Raqueno, N., Messinger, D. W., Faulring, J., Kerekes, J. P., van Aardt, J., Canham, K., Hagstrom, S., Ontiveros, E., Gerace, A., Kaufman, J., Vongsy, K. M., Griffith, H., and Bartlett, B. D., "The SHARE 2012 data collection campaign," (April 2013).