Hyperspectral Unmixing via Turbo Bilinear Approximate Message Passing

Jeremy Vila, Philip Schniter* and J. Meola

Abstract—The goal of hyperspectral unmixing 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 based on loopy belief propagation (BP) that enables the exploitation of spectral coherence in the endmembers and spatial coherence in the abundances. In particular, we partition the factor graph into spectral bands, spatial coherence, and bilinear subgraphs, and pass messages between them using a “turbo” approach. To perform message passing within the bilinear subgraph, we employ the bilinear generalized approximate message passing algorithm (BiG-AMP), a recently proposed belief-propagation-based approach to matrix factorization. Furthermore, we propose an expectation-maximization (EM) strategy to tune the prior parameters of a collection strategy to select the number of materials $N$. Numerical experiments conducted with both synthetic and real-world data show favorable unmixing performance relative to existing methods.

Index Terms—Hyperspectral imaging, approximate message passing, belief propagation, expectation-maximization algorithms

I. INTRODUCTION

In hyperspectral unmixing (HU), the objective is to jointly estimate the radiance spectra and per-pixel abundances of the $N$ materials present in a scene, given measurements across $M$ spectral bands at each of $T = T_1 \times T_2$ pixels. Often, linear mixing [2], [3] is assumed, in which case the measurements $Y \in \mathbb{R}^{M \times T}$ are modeled as

$$Y = SA + W,$$

where the $n$th column of $S \in \mathbb{R}^{M \times N}$ represents the spectrum (or “endmember”) of the $n$th material, the $n$th row of $A \in \mathbb{R}^{N \times T}$ represents the spatial abundance of the $n$th material, and $W$ represents noise. Both $S$ and $A$ must contain only non-negative (NN) elements, and each column of $A$ must obey the simplex constraint (i.e., NN and sum-to-one).

Traditionally, hyperspectral unmixing is a two-step procedure, consisting of endmember extraction (EE) to recover the endmembers followed by inversion to recover the abundances.

Many EE algorithms leverage the “pure pixel” assumption: for each material, there exists at least one observed pixel containing only that material (i.e., all columns of the $N \times N$ identity matrix can be found among the columns of $A$). Well-known examples of pure-pixel-based EE algorithms include N-FINDR [4] and VCA [5]. The existence of pure pixels in HU is equivalent to “separability” in the problem of non-negative matrix factorization (NMF), where the goal is to find $S \in \mathbb{R}^{M \times N}$ and $A \in \mathbb{R}^{N \times T}$ matching a given $Z = SA$. There, separability has been shown to be sufficient for the existence of unique factorizations [6] and polynomial-time solvers [7], with a recent example being the FSNMF algorithm from [8]. In HU, however, the limited spatial-resolution of hyperspectral cameras implies that the pure-pixel assumption does not always hold in practice. With “mixed pixel” scenarios in mind, algorithms such as Minimum Volume Simplex Analysis (MVSA) [9] and Minimum Volume Enclosing Simplex (MVES) [10] attempt to find the minimum-volume simplex that contains the data $Y$.

In the inversion step, the extracted endmembers in $\hat{S}$ are used to recover the simplex-constrained abundances in $A$. Often this is done by solving [11], [12]

$$\hat{A} = \arg \min_{A \geq 0} \|Y - \hat{SA}\|_F^2 \text{ s.t. } 1_N^T A = 1_T,$$

where $1_N$ denotes the $N \times 1$ vector of ones, which is usually referred to as fully constrained least squares (FCLS).

Real-world hyperspectral datasets can contain significant structure beyond non-negativity on $s_{mn}$ and simplex constraints on $\{a_{nt}\}_{n=1}^N$. For example, the abundances $\{a_{nt}\}_{n=1}^N$ will be sparse if most pixels contain significant contributions from only a small subset of the $N$ materials. Also, the abundances $\{a_{nt}\}_{t=1}^T$ will be spectrally coherent if the presence of a material in a given pixel makes it more likely for that same material to exist in neighboring pixels. Likewise, the endmembers $\{s_{mn}\}_{m=1}^M$ will be spectrally coherent if the radiance values are correlated across frequency.

Various unmixing algorithms have been proposed to leverage these additional structures. For example, given an end-member estimate $S$, the SUnSAL algorithm [13] estimates sparse abundances $A$ using $\ell_1$-regularized least-squares (LS), and the SUnSAL-TV algorithm [14] adds total-variation (TV) regularization [15] to also penalize changes in abundance across neighboring pixels (i.e., to exploit spatial coherence). SUnSAL and SUnSAL-TV can be categorized as unmixing algorithms, rather than inversion algorithms, since their $\ell_1$-regularization supports the use of large (i.e., $N > M$) and scene-independent endmember libraries for $S$. However, there
are limitations on the size of the library $\hat{S}$, and it can be difficult to determine suitable choices for the $\ell_1$ and TV regularization weights.

Bayesian approaches to hyperspectral unmixing have also been proposed. For example, the Bayesian Linear Unmixing (BLU) algorithm [16] employs priors that enforce NN constraints on the endmembers and simplex constraints on the per-pixel abundances, and returns either (approximately) minimum mean-square error (MMSE) or maximum a posteriori (MAP) estimates using Gibbs sampling. The Spatially Constrained Unmixing (SCU) [17] algorithm, an extension of BLU, furthermore exploits spatial coherence using a hierarchical Dirichlet-process prior. Both BLU and SCU have been shown to outperform N-FINDR and VCA-plus-FCLS under certain conditions [17], but at the cost of several orders-of-magnitude increase in runtime.

In this paper, we propose a novel Bayesian approach to HU that is based on loopy belief propagation. Our approach, referred to as HU turbo-AMP (HUT-AMP), partitions the factor graph (see Fig. 1) into three subgraphs: one that models spectral coherence (using $N$ Gauss-Markov chains), one that models spatial coherence (using $N$ binary Markov Random Fields (MRFs)), and one that models the NN bilinear structure of $[1]$. While the first two subgraphs yield inference problems that are handled efficiently by standard methods [18], [19], the third does not. Thus, to perform efficient inference on the latter subgraph, we apply the recently proposed Bilinear Generalized Approximate Message Passing (BiG-AMP) algorithm [20]. BiG-AMP can be interpreted as an extension of approximate message passing (AMP) techniques [21]–[23], originally proposed for the linear observation models that arise in compressive sensing, to bilinear models like [1]. To merge BiG-AMP-based inference with Markov-chain and MRF-based inference, we leverage the “turbo AMP” approach first proposed in [24] and subsequently applied to joint channel-estimation and decoding [25], [26], compressive image retrieval [27], [28], and compressive video retrieval [29], all with state-of-the-art results. Furthermore, since the parameters of the various prior distributions are unknown in practice, we use the expectation-maximization (EM) algorithm to automatically tune them, building on the NN sparse reconstruction work in [30]. Lastly, when the number of materials $N$ is unknown, we show how it can be accurately estimated using a classical model-order selection (MOS) strategy [31].

We evaluate the performance of our proposed technique, in comparison to several recently proposed methods, through a detailed numerical study that includes both synthetic and real-world datasets. The results, presented in Sec. IV suggest that HUT-AMP yields an excellent combination of unmixing performance and computational complexity.

**Notation:** For matrices, we use boldface capital letters like $A$, and we use $A^T$, $\text{tr}(A)$, and $\|A\|_F$ to denote the transpose, trace, and Frobenius norm, respectively. For vectors, we use boldface small letters like $x$, and we use $\|x\|_p = \left(\sum_n |x_n|^p\right)^{1/p}$ to denote the $\ell_p$ norm, with $x_n = |x_n|$ representing the $n^{th}$ element of $x$. We use $1_N$ to denote the $N \times 1$ vector of ones. Deterministic quantities are denoted using serif typeface (e.g., $x$, $x$, $X$), while random quantities are denoted using san-serif typeface (e.g., $X$, $x$, $X$). For random variable $x$, we write the pdf as $p_x(x)$, the expectation as $E_x\{x\}$, and the variance as $\text{var}\{x\}$. For a Gaussian random variable $x$ with mean $\mu$ and variance $\nu$, we write the pdf as $N(x; \mu, \nu)$, and, for the special case of $N(x; 0, 1)$, we abbreviate the pdf as $\varphi(x)$ and write the complimentary cdf as $\Phi_c(x)$. Finally, we use $\delta(x)$ (where $x \in \mathbb{R}$) to denote the Dirac delta distribution and $\delta_n$ (where $n \in \mathbb{Z}$) to denote the Kronecker delta sequence.

**II. SIGNAL AND OBSERVATION MODELS**

**A. Background on BiG-AMP**

As described in the introduction, a distinguishing feature of our approach is the use of BiG-AMP [20] for bilinear inference. We begin by overviewing BiG-AMP, since its operating assumptions affect the construction of our statistical model.

Consider the problem of estimating the matrices $S \in \mathbb{R}^{M \times N}$ and $\mathbf{A} \in \mathbb{R}^{N \times T}$ from a noisy observation $\mathbf{Y} \in \mathbb{R}^{M \times T}$ of the hidden bilinear form $\mathbf{Z} \triangleq S \mathbf{A} \in \mathbb{R}^{M \times T}$. (Our use of overbar notation will become clear in the sequel.) BiG-AMP treats the elements in both $S$ and $\mathbf{A}$ as independent random variables $S_{mn}$ and $A_{nt}$ with known prior pdfs $p_{smn}(\cdot)$ and $p_{amt}(\cdot)$, respectively, with $S_{mn}$ being zero-mean. Furthermore, it assumes that the likelihood function of $\mathbf{Z}$ is known and separable, i.e., of the form

$$p_{\mathbf{Z}|\mathbf{Y}}(\mathbf{Z}|\mathbf{Y}) = \prod_{m=1}^{M} \prod_{t=1}^{T} p_{Z_{mt}|Z_{mt}}(S_{mn}|A_{nt}).$$

BiG-AMP then converts the computationally intractable matrix-inference problem to a sequence of simple scalar-inference problems. In particular, it iteratively computes approximations to the marginal posterior pdfs $p_{S_{mn}}(\cdot)$ and $p_{A_{nt}}(\cdot)$, respectively, using an approximation of the sum-product algorithm (SPA) [12].
One important property of this approximation is that the messages from the \( S_{mn} \) nodes to the \( f_{mn} \) nodes in Fig. 1 are Gaussian, as are those from the \( a_{nt} \) nodes to the \( g_{nt} \) nodes. We refer the interested reader to \citep{20} for derivation details and a full statement of the algorithm.

BiG-AMP’s complexity is dominated by ten matrix multiplies (of the form \( \mathbf{SA} \)) per iteration, although simplifications can be made in the case of Gaussian \( p_{\mathbf{Y} | \mathbf{Z}, \mathbf{m}, \mathbf{mt}, \mathbf{mt}} \) that reduce the complexity to three matrix multiplies per iteration \citep{20}. Furthermore, when BiG-AMP’s likelihood function and priors include unknown parameters \( \Omega \), expectation-maximization (EM) methods can be used to learn them, as described in \citep{20}. BiG-AMP was shown \citep{33} to yield excellent performance on matrix completion, robust PCA, and dictionary learning problems, and here we show that it performs very well on the NMF and HU problems as well.

\subsection*{B. Augmented Observation Model}

We model the elements of the additive noise matrix \( \mathbf{W} \) in (1) as i.i.d zero-mean Gaussian with variance \( \psi > 0 \). Thus, the BiG-AMP marginal likelihoods take the form

\[
p_{\mathbf{Y} | \mathbf{Z}, \mathbf{m}, \mathbf{mt}, \mathbf{mt}} = \mathcal{N}(\mathbf{Y} | \mathbf{Z}, \mathbf{mt}, \mathbf{mt}; \psi).
\]

For now we treat \( \psi \) as known, but later (in Sec. III-C) we describe how it and other model parameters can be learned from \( \mathbf{Y} \).

Leveraging the zero-mean property of the noise, we first perform mean-removal on the observations \( \mathbf{Y} \). In particular, we subtract the empirical mean

\[
\mu \triangleq \frac{1}{MT} \sum_{t=1}^{T} \sum_{m=1}^{M} y_{mt} = \frac{1}{MT} \mathbf{1}_M^T \mathbf{Y} \mathbf{1}_T
\]

from \( \mathbf{Y} \) to obtain

\[
\mathbf{Y} \triangleq \mathbf{Y} - \mu \mathbf{1}_M \mathbf{1}_T^T
\]

\[
= (\mathbf{S} - \mu \mathbf{1}_M \mathbf{1}_N^T) \mathbf{A} + \mathbf{W},
\]

\[
\triangleq \mathbf{S} \mathbf{A} + \mathbf{W}
\]

where (6) employed (1) and \( \mathbf{1}_N \mathbf{A} = \mathbf{1}_T^T \), the latter of which results from the simplex constraint on the columns of \( \mathbf{A} \). It can then be shown (see Appendix A) that the elements of \( \mathbf{S} \) in (6) are approximately zero-mean.

To enforce the linear equality constraint \( \mathbf{1}_N \mathbf{A} = \mathbf{1}_T^T \), we augment the observation model (6) into the form

\[
\begin{bmatrix} \mathbf{Y}^T \\ \mathbf{1}_N^T \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{S} \\ \mathbf{1}_N \end{bmatrix} \mathbf{A} \triangleq \mathbf{S} \mathbf{A} + \mathbf{W}
\]

\[
\begin{bmatrix} \mathbf{1}_N^T \end{bmatrix} \triangleq \mathbf{W}
\]

For the augmented model (7), the likelihood function of \( \mathbf{S} \triangleq \mathbf{SA} \) takes the form in (3) with

\[
p_{\mathbf{Y} | \mathbf{Z}, \mathbf{m}, \mathbf{mt}, \mathbf{mt}} = \begin{cases} 
\mathcal{N}(\mathbf{Y} | \mathbf{Z}, \mathbf{mt}, \mathbf{mt}; \psi) & m = 1, \ldots, M \\
\delta(\mathbf{Y} - \mathbf{Z}) & m = M + 1.
\end{cases}
\]

We note that, ignoring spectral and spatial coherence, the model (7) is appropriate for the application of BiG-AMP, since the likelihood function \( p_{\mathbf{Y} | \mathbf{Z}}(\mathbf{Y} | \mathbf{Z}) \) is known (up to \( \psi \) and

separable, and since the elements in \( \mathbf{S} \) and \( \mathbf{A} \) can be treated as independent random variables with priors known up to a set of parameters, with those in \( \mathbf{S} \) being approximately zero-mean. In the sequel, we describe how the model (7) can be extended to capture spectral and spatial coherence.

\subsection*{C. Endmember Model}

We desire a model that promotes spectral coherence in the endmembers, i.e., correlation among the (mean removed) spectral amplitudes \( \{s_{mn}\}_{m=1}^M \) of each material \( n \). However, since BiG-AMP needs \( \mathbf{S}_{mn} \) to be independent, we cannot impose correlation on these variables directly. Instead, we introduce an auxiliary sequence of correlated amplitudes \( \{e_{mn}\}_{m=1}^M \) such that \( \mathbf{S}_{mn} \) is independent conditional on \( \mathbf{e}_{mn} \). In particular,

\[
p_{\mathbf{S}_{mn} | \mathbf{e}_{mn}} = \prod_{m=1}^M \prod_{n=1}^N p_{\mathbf{S}_{mn} | \mathbf{e}_{mn}}(s_{mn} | e_{mn})
\]

\[
p_{\mathbf{e}_{mn}} = p_{\mathbf{e}_{mn}}(e_{mn})
\]

implying that \( \mathbf{e}_{mn} \) is merely a copy of \( \mathbf{S}_{mn} \). To impart correlation within the auxiliary sequences \( \{\mathbf{e}_{mn}\}_{m=1}^M \), we model them as independent Gaussian-Markov models

\[
p_{\mathbf{E}} = \prod_{n=1}^N \prod_{m=2}^M p(e_{mn} | e_{m-1,n}),
\]

where \( \mathbf{e}_n = [e_{1n}, \ldots, e_{Mn}]^T \), \( e_n = [e_{1n}, \ldots, e_{Mn}]^T \), and

\[
p(e_{mn} | e_{m-1,n}) = \mathcal{N}(e_{mn}; \kappa_n, \sigma_n^2)
\]

\[
\kappa_n \in \mathbb{R}
\]

In (12)-(13), \( \kappa_n, \sigma_n \in \mathbb{R} \) controls the mean of the \( n \)th process, \( \sigma_n^2 \) controls the variance, and \( \eta_n \in [0, 1] \) controls the correlation. The resulting factor graph is illustrated in Fig. 2.

We note that the model (12)-(13) does not explicitly enforce non-negativity in \( s_{mn} \) because, for simplicity, we have omitted the constraint \( s_{mn} \geq -\mu \). Enforcement of \( s_{mn} \geq -\mu \) could be accomplished by replacing the pdfs in (12)-(13) with truncated Gaussian versions, but the computations required for inference would become much more tedious. In our experience, this tedium is not warranted: with practical HU datasets\textsuperscript{1}, it suffices to enforce non-negativity in \( \mathbf{A} \) and keep \( \mathbf{Y} \approx \mathbf{SA} \).

\textsuperscript{1}Throughout our numerical experiments, the proposed inference method never produced a negative estimate of \( s_{mn} \).
particular, we assume that probability valued, and we impose structure on the D. Abundance Model

where support of the Roughly speaking, larger and larger Gaussian mixture (NNGM) distribution for model \[18\], as illustrated by the factor graph in Fig. 3.

We desire a model that promotes both sparsity and spatial coherence for each material \(n\). For this purpose, we introduce the support variables \(d_{nt} \in \{-1,1\}\), where \(d_{nt} = -1\) indicates that \(a_{nt}\) is zero-valued, and \(d_{nt} = 1\) indicates that \(a_{nt}\) is non-zero with probability 1, which we will refer to as “active.” By modeling the abundances \(a_{nt}\) as independent conditional on \(d_{nt}\), we comply with the independence assumptions of BiG-AMP. In particular, we assume that

\[
p_{\text{AD}}(A | D) = \prod_{n=1}^{N} \prod_{t=1}^{T} p_{a_{nt}}(a_{nt} | d_{nt})
\]

where \(\zeta_n(\cdot)\) denotes the pdf of \(a_{nt}\) when active. Essentially, we employ a Bernoulli-\(\zeta_n(\cdot)\) distribution for the \(n\)th material.

We then place a Markov random field (MRF) prior on the support of the material \(d_{nt}\), \(d_{nt} \triangleq [d_{n1}, \ldots, d_{nT}]^T\):

\[
p_{D}(D) = \prod_{n=1}^{N} p_d(a_{n})
\]

\[
p_{d_{n}}(a_{n}) \propto \exp \left( \sum_{t=1}^{T} \left( \frac{1}{2} \sum_{i \in D_t} \beta_t d_{n_i} - \alpha_n \right) d_{nt} \right),
\]

where \(D_t \subseteq \{1, \ldots, T\} \setminus t\) denotes the neighbors of pixel \(t\). Roughly speaking, larger \(\beta_t\) yields higher spatial coherence and larger \(\alpha_n\) yields higher sparsity. For simplicity, we adopt a neighborhood structure corresponding to the classical Ising model \([18]\), as illustrated by the factor graph in Fig. 3.

As for the active abundances, we adopt a non-negative Gaussian mixture (NNGM) distribution for \(\zeta_n(\cdot)\):

\[
\zeta_n(a) = \sum_{\ell=1}^{L} \omega^{n}_{\ell} \mathcal{N}_{+}(a; \theta^{n}_{\ell}, \phi^{n}_{\ell}),
\]

where \(\omega^{n}_{\ell} \geq 0\) and \(\sum_{\ell=1}^{L} \omega^{n}_{\ell} = 1\). In \(\mathcal{N}_{+}\), \(\mathcal{N}_{+}\) refers to the truncated Gaussian pdf

\[
\mathcal{N}_{+}(x; \theta, \phi) = \begin{cases} 0 & x < 0 \\ \frac{1}{\sqrt{2\pi} \phi} e^{-\frac{(x-\theta)^2}{2\phi^2}} & x \geq 0 \end{cases}
\]

III. THE HUT-AMP ALGORITHM

A. Message Passing and Turbo Inference

Our overall goal is to jointly estimate the (correlated, non-negative) endmembers \(S\) and (structured sparse, simplex-constrained) abundances \(A\) from noisy observations \(Y\) of the bilinear form \(Z = SA\). Using the mean-removed, augmented probabilistic models from Sec. [II], the joint pdf of all random variables can be factored as follows:

\[
\begin{aligned}
&\log p(Y, \bar{S}, A, E, D) \\
= &\log p(Y|\bar{S}, A) p(\bar{S}|E) p(A, D) \\
= &\log p(Y|\bar{S}) p(\bar{S}|E) p(E) p_{A|D}(A|D) p_D(D)
\end{aligned}
\]

yielding the factor graph in Fig. 1. Due to the cycles within the factor graph, exact inference is NP-hard \([34]\), and so we settle for approximate MMSE inference.

To accomplish approximate MMSE inference, we apply a form of loopy belief propagation that is inspired by the “turbo decoding” approach used in modern communications receivers \([45]\). In particular, after partitioning the overall factor graph into three subgraphs, as in Fig. 1, we alternate between message-passing within subgraphs and message-passing between subgraphs. In our case, BiG-AMP \([20]\) is used for message-passing within the bilinear subgraph and standard methods from \([18], [19]\) are used for message-passing within the other two subgraphs, which involve \(N\) Gauss-Markov chains and \(N\) binary MRFs, respectively. Overall, our proposed approach can be interpreted as a bilinear extension of the “turbo AMP” approach first proposed in \([24]\).

B. Messaging Between Subgraphs

For a detailed description of the message passing within the Gauss-Markov, MRF, and BiG-AMP subgraphs, we refer interested readers to \([18], [19], [20]\), respectively. We now describe the message passing between subgraphs, which relies on the sum-product algorithm (SPA) \([32]\). In our implementation of the SPA, we assume that all messages are scaled to
form valid pdfs (in the case of continuous random variables) or pmfs (in the case of discrete random variables), and we use \( \Delta^b_c(\cdot) \) to represent the message passed from node \( b \) to node \( c \).

As described in [32], the SPA message flowing out of a variable node along a given edge equals the (scaled) product of messages flowing into that node along its other edges. Meanwhile, the SPA message flowing out of a factor node along a given edge equals the (scaled) integral of the product of all incoming messages times the factor associated with that node. Finally, the SPA approximates the posterior of a given random variable as the (scaled) product of messages flowing into that random variable.

As discussed in Sec. II-A, a key property of BiG-AMP is that it certain messages within its sub-graph are approximated as Gaussian. In particular,

\[
\Delta^m_{f_{mn}}(\omega) = N(\omega; \tilde{q}_{mn}, \nu^q_{mn})(23)
\]

\[
\Delta^m_{g_{nt}}(a) = N(a; \tilde{r}_{nt}, \nu^r_{nt})(24)
\]

where the quantities \( \tilde{q}_{mn}, \nu^q_{mn}, \tilde{r}_{nt}, \nu^r_{nt} \) are computed during the final iteration of BiG-AMP. Thus, the SPA approximated posteriors on \( \mathbb{S}_m \) and \( \mathbb{A}_t \) take the form

\[
p_{\mathbb{S}_m | q_m}(s | \tilde{q}_{mn}, \nu^q_{mn}) \sim \Delta^m_{f_{mn}}(s) N(s; \tilde{q}_{mn}, \nu^q_{mn})(25)
\]

\[
p_{\mathbb{A}_t | r_t}(a | \tilde{r}_{nt}, \nu^r_{nt}) \sim \Delta^m_{g_{nt}}(a) N(a; \tilde{r}_{nt}, \nu^r_{nt})(26)
\]

We will use these properties in the sequel.

First, we discuss the message-passing between the bilinear sub-graph and spectral-coherence sub-graph in Fig. I. Given \( (10), (23) \), and the construction of the factor graph in Fig. I the SPA implies that

\[
\Delta^m_{f_{mn}}(e) \propto \int f_m(s, e) \Delta^m_{f_{mn}}(s) ds \propto N(e; \tilde{q}_{mn}, \nu^q_{mn})(27)
\]

\[
\Delta^m_{g_{nt}}(s) = \int f_m(s, e) \Delta^m_{g_{nt}}(e) de \propto N(s; \tilde{r}_{nt}, \nu^r_{nt})(28)
\]

The messages in (28) are used as inputs to the Gauss-Markov inference procedure. By construction, the outputs of the Gauss-Markov inference procedure will also be Gaussian beliefs. Denoting their means and variances by \( \bar{\Omega}_{mn} \) and \( \bar{\Phi}_{mn} \), respectively, we have that

\[
\Delta^m_{f_{mn}}(e) \propto N(e; \bar{\Omega}_{mn}, \bar{\Phi}_{mn})(29)
\]

\[
\Delta^m_{g_{nt}}(s) = \int f_m(s, e) \Delta^m_{g_{nt}}(e) de \propto N(s; \bar{\Theta}_{nt}, \bar{\Phi}_{nt})(30)
\]

When BiG-AMP is subsequently called for inference on the bilinear sub-graph, it will use \( \Delta^m_{f_{mn}}(\cdot) \) as the prior on \( \mathbb{S}_m \).

Next we discuss the message-passing between the bilinear sub-graph and the spatial-coherence sub-graph in Fig. I. The SPA, together with the construction of the factor graph in Fig. I imply

\[
\Delta^m_{g_{nt}}(a, d = +1) = \left(1 + \frac{N(0; \tilde{r}_{nt}, \nu^r_{nt})}{\int \zeta_n(a) N(a; \tilde{r}_{nt}, \nu^r_{nt})} \right)^{-1}(34a)
\]

\[
\Delta^m_{g_{nt}}(a, d = -1) = 1 - \Delta^m_{g_{nt}}(d = +1), (34b)
\]

which implies

\[
\Delta^m_{g_{nt}}(d = +1) = \pi_{nt}(35a)
\]

\[
\Delta^m_{g_{nt}}(d = -1) = 1 - \pi_{nt}(35b)
\]

for some \( \pi_{nt} \in (0, 1) \). The SPA and (15) then imply that

\[
\Delta^m_{g_{nt}}(a) \propto \sum_{d = +1} g_{nt}(a, d) \Delta^m_{g_{nt}}(d)(36)
\]

\[
= (1 - \pi_{nt}) \delta(a) + \pi_{nt} \zeta_n(a)(37)
\]

for \( \zeta_n(\cdot) \) defined in (13). When BiG-AMP is subsequently called for inference on the bilinear sub-graph, it will use \( \Delta^m_{g_{nt}}(\cdot) \) as the prior on \( \mathbb{A}_t \).

C. EM Learning of the Prior Parameters

In practice, we desire that the parameters

\[
\Omega = \{\psi, \{\omega^a_{nt}, \theta^a_{nt}, \phi^a_{nt}\}_{nt}, \{\eta_{nt}, \kappa_{nt}, \sigma^2_{nt}, \alpha_{nt}, \beta_{nt}\}_{nt}\}(38)
\]

used for the assumed likelihood \( p_{\psi, \omega^a_{nt}, \theta^a_{nt}}(y_{mt}|\cdot) \), NNGM abundance prior \( \zeta_n(\cdot) \), Gauss-Markov chain \( p_{\Theta_{nt}}(\cdot) \), and binary MRF \( p_{\mathbb{A}_t}(\cdot) \) are well tuned. With this in mind, we propose an expectation-maximization (EM) [36] procedure to tune \( \Omega \), similar to that used for the GAMP-based sparse-reconstruction algorithms in [37] and [39].

To tune \( \Omega \), the EM algorithm [36] iterates

\[
\Omega^{i+1} = \arg \max \Omega E \left\{ \ln p(E, A, D, Y; \Omega) \mid Y, \Omega^i \right\}(39)
\]

with the goal of increasing a lower bound on the true likelihood \( p(Y; \Omega) \) at each EM-iteration \( i \). In our case, the true posterior distribution used to evaluate the expectation in (39) is NP-hard to compute, and so we use the SPA-approximated posteriors \( \hat{p}_{E|Y}(E|Y) \propto \prod_{m,n} \Delta_{g_{mn}}(e_{mn}) \Delta_{f_{mn}}(e_{mn}) \) from (28)–(29), \( \hat{p}_{D|Y}(D|Y) \propto \prod_{m,n} \Delta_{g_{nt}}(d_{nt}) \Delta_{f_{nt}}(d_{nt}) \) from (34a)–(35), and \( \hat{p}_{A|Y}(A|Y) \propto \prod_{n,t} \Delta_{g_{nt}}(a_{nt}) \Delta_{g_{nt}}(a_{nt}) \) from (24) and (31). Furthermore, since it is difficult to perform the maximization in (39) jointly, we maximize \( \Omega \) one component at a time (while holding the others fixed), which is the well known “incremental” variant of EM [38].

The resulting EM-update expressions for the noise and NNGM parameters \( \psi, \omega^a_{nt}, \theta^a_{nt}, \phi^a_{nt} \) can be found in [39], and those for the Gauss-Markov chain parameters \( \eta_{nt}, \kappa_{nt}, \sigma^2_{nt} \) can be found in [32]. They are all computed in closed-form using readily available quantities, and thus do not add significantly to the complexity of HUT-AMP. The update procedure for the binary MRF parameters \( \alpha_{nt}, \beta_{nt} \) is described in [28] and uses gradient descent. Since a small number of gradient-descent iterations suffice, this latter procedure does not significantly increase the complexity of HUT-AMP.
D. EM Initialization

Since the EM algorithm may converge to a local maximum of the likelihood, care must be taken when initializing the EM-learned parameters. Below, we propose an initialization strategy for HUT-AMP that, based on our empirical experience, seems to work well.

We first initialize the endmembers \( S \). For this, we found it effective to use an off-the-shelf EE algorithm like VCA or FSNMF to recover \( \hat{S}^0 \). Then, as described in (6), we subtract the observation mean \( \mu \) from \( \hat{S}^0 \) to obtain the initialization \( \hat{S}^0 \).

With the aid of \( \hat{S}^0 \), we next run BiG-AMP under

1) the trivial endmember prior

\[
\Delta_{fmn}^0(S) = \delta(S - S_{mn}),
\]

which essentially fixes the endmembers at \( \hat{S}^0 \),

2) the agnostic NNGM abundance initialization from \([30]\):

\[
\Delta_{\text{agn}}^0(a) = (1 - \rho_{nt}^0)\delta(a) + \rho_{nt}^0 \sum_{l=1}^L \omega_{nl}^l \nu_{+}(a; \theta_{nt}^l, \phi_{nt}^l)
\]

with \( \{\omega_{nl}^l, \theta_{nt}^l, \phi_{nt}^l\}_{l=1}^L \) set at the best fit to a uniform distribution on the interval \([0, 1]\), and \( \rho_{nt}^0 = \frac{1}{N} \), and

3) the noise variance initialization from \([30]\):

\[
\psi_0^0 = \frac{\|Y\|_F^2}{(\text{SNR})^2 + 1)MT},
\]

where, without any prior knowledge of the true SNR \( \Delta \equiv E\{\|z_m\|_F^2\}/\psi \), we suggest setting SNR \( 0 = 0 \) dB.

By running BiG-AMP under these settings, we initialize the messages \( \Delta_{fmn}^0(\cdot) \) and \( \Delta_{\text{agn}}^0(\cdot) \) from \([23, 24]\) and we also obtain an initial estimate of \( A \) from the mean of the approximate posterior \( \hat{A}^D \), which we shall refer to as \( \hat{A}^D \).

Finally, we initialize the remaining parameters in \( \Omega \). Starting with the spectral coherence parameters, we set the mean \( \kappa_0^0 \) and variance \( \sigma_{\kappa}^0 \) at the empirical mean and variance, respectively, of the elements in the \( n \)th column of \( \hat{S}^0 \). Then, we initialize the correlation \( \eta_n \) as suggested in \([29]\), i.e.,

\[
\varphi_0 = \frac{\|Y\|_F^2 - MT\psi_0^0}{\|A\|_F^2},
\]

\[
\eta_n = 1 - \frac{1}{M-1} \sum_{m=1}^{M-1} \left| \frac{y_{mt} y_{mnt}^T}{\|A\|_F^2} \right|^2, \quad n = 1, \ldots, N,
\]

where \( y_{mt} \) denotes the \( m \)th row of \( Y \). Lastly, we initialize the spatial coherence parameters as suggested in \([28]\), i.e., \( \beta_0^0 = 0.4 \) and \( \alpha_0^0 = 0.4 \), since \([28]\) shows these values to work well over a wide operating range.

2 With FSNMF (which was used for all of the experiments in Sec. [11, we found that it helped to post-process the observations to reduce the effects of noise. For this, we used the standard PCA-based denoising approach described in [3] the signal subspace was estimated from the left singular vectors of \( Y \) after row-wise mean-removal, and the FSNMF-estimated endmembers were projected onto the signal subspace.

E. HUT-AMP Summary

We now describe the scheduling of turbo-messaging and EM-tuning steps, which together constitute the HUT-AMP algorithm. Essentially, we elect to perform one EM update per turbo iteration, yielding the steps tabulated in Table I. As previously mentioned, the “BiGAMP” operation iterates the BiG-AMP algorithm to convergence as described in (20), the “GaussMarkov” operation performs standard Gauss-Markov inference as described in (13), and the “MRF” operation performs MRF inference via the belief-propagation method described in (19).

F. Selection of Model Order \( N \)

In practice, the number of materials \( N \) present in a scene may be unknown. Thus, we now propose a method to estimate \( N \) from the observed data \( Y \). For this, we use the standard form of penalized log-likelihood maximization \([51]\)

\[
\hat{N} = \arg \max_{N} \text{max} 2 \ln p_T \{ Y | \hat{S}_N, \hat{A}_N; \hat{\psi}_{\text{ML}} \} - \gamma(N),
\]

where \( \hat{S}_N \) and \( \hat{A}_N \) are the estimates of the mean-removed endmembers and abundances returned from \( N \)-material HUT-AMP, \( \hat{\psi}_{\text{ML}} \) is the ML estimate of the noise variance, and \( \gamma(\cdot) \) is a penalty term. As recommended in \([53]\), we choose \( \gamma(\cdot) \) in accordance with the small-sample-corrected Akaike information criterion (AICc) \([31]\), i.e., \( \gamma(N) = 2MT - n(N) - 1 \), where \( MT \) is the number of scalar observations in \( Y \) and \( n(N) \) is the number of scalar degrees-of-freedom (DoF) in our model, which depends on \( N \). In particular, \( n(N) \) comprises \( MN \) DoF from \( S \), \( (N-1)T \) DoF from \( A \), and \( 5N+2NL+N(L-1)+1 \) DoF from \( \Omega \). Plugging the standard form of the ML estimate of \( \psi \) (see, e.g., \([51]\), eq. (7)) into (45), we obtain

\[
\hat{N} = \arg \max_{N} -MT \ln \left( \frac{\|Y - \hat{S}_N \hat{A}_N\|_F^2}{MT} \right) - \frac{2MTn(N)}{MT-n(N)-1}.
\]

To solve the maximization in (46), we first run \( N = 2 \) HUT-AMP to completion and compute the penalized log-likelihood. We then increment \( N \) by 1, and compute the



| TABLE I |
|-----------------------------------------------|
| **HUT-AMP Pseudocode for Fixed Number of Materials** |
| 1: Initialize \( \Delta_{f}^0 \), \( \Delta_{\text{agn}}^0 \), and \( \Omega^* \) as described in Sec. III-D |
| 2: for \( i = 1, 2, \ldots \) do |
| 3: convert \( \Delta_{f}^0 \) to \( \Delta_{f}^0 \) via \([23]\) and \([25]\) |
| 4: convert \( \Delta_{\text{agn}}^0 \) to \( \Delta_{\text{agn}}^0 \) via \([24]\) and \([25]\) |
| 5: \( \Delta_{f}^0 = \text{GaussMarkov}(\Delta_{f}^0, \Omega^*) \) |
| 6: \( \Delta_{f}^0 = \text{MRF}(\Delta_{f}^0, \Omega^*) \) |
| 7: convert \( \Delta_{f}^0 \) to \( \Delta_{f}^0 \) via \([29]\) and \([31]\) |
| 8: convert \( \Delta_{\text{agn}}^0 \) to \( \Delta_{\text{agn}}^0 \) via \([35]\) and \([37]\) |
| 9: \( \Omega^* = \text{EM}(\Delta_{f}^0, \Delta_{\text{agn}}^0, \Delta_{f}^0, \Delta_{f}^0, \Delta_{f}^0, \Delta_{f}^0, \Delta_{f}^0, \Delta_{f}^0, \Omega^*) \) |
| 10: \( \Delta_{f}^0, \Delta_{\text{agn}}^0 = \text{BiGAMP}(\Delta_{f}^0, \Delta_{\text{agn}}^0, \Omega^*) \) |
| 11: end for |
penalized log-likelihood again. If it increases, we increment \( N \) by 1 and repeat the procedure. Once the penalized log-likelihood decreases, we stop the procedure and select the previous model order \( N \), which is the maximizer of the penalized log-likelihood. We refer to the resulting procedure as “HUT-AMP with model-order selection” (HUT-AMP-MOS).

We also note that a similar model-order selection strategy can be implemented to tune the number of NNGM components \( L \) used in [13], and we refer interested readers to [37] for more details. We note, however, that the fixed choice \( L = 3 \) was sufficient to yield the excellent results in Sec. IV.

### IV. Numerical Results

In this section, we report the results of several experiments that we conducted to characterize the performance of our proposed methods on both synthetic and real-world datasets.

In these experiments, we compared the endmembers \( \hat{S} \) recovered from our proposed HUT-AMP and HUT-AMP-MOS\({}^3\) unmixing algorithms to those recovered by the Bayesian unmixing algorithm SCU [17] and the endmember extraction (EE) algorithms VCA [5], FSNMF [8], and MVSA [9]. For FSNMF, we used the PCA post-processing described in footnote 2 to reduce the effects of measurement noise, since this greatly improved its mean-squared estimation error.

We also compared the abundances \( \hat{A} \) recovered by our proposed HUT-AMP and HUT-AMP-MOS unmixing algorithms to those recovered by the Bayesian unmixing algorithm SCU, as well as those recovered by both FCLS [2] (implemented via Matlab’s lsqnonlin) and SU-SAL-Tv [14] using the endmember estimates produced by VCA, FSNMF, and MVSA.

In all cases, algorithms were run using their authors’ implementation and suggested default settings, unless noted otherwise. For SU-SAL-Tv, the regularization weights for the \( \ell_1 \) and TV norms were hand-tuned, because cross-validation tuning was too computationally expensive given the sizes of the datasets.

#### A. Pixel Purity versus Abundance Sparsity

In the first experiment, we aim to assess EE performance as a function of pixel purity and abundance sparsity. Our motivation stems from the fact that the proposed HUT-AMP algorithm aims to exploit sparsity in the columns of the abundance matrix \( A \), while classical EE techniques like VCA and FSNMF aim to exploit the presence of pure pixels, recalling the discussion in Sec. I. Thus, we are interested in seeing how these contrasting approaches fare under varying degrees of pixel purity and abundance sparsity.

For this first experiment, we constructed synthetic data consisting of \( M = 100 \) spectral bands, \( T = 115 \) spatial pixels, and \( N = 10 \) materials. The endmember matrix \( S \in \mathbb{R}^{M \times N} \) was drawn i.i.d such that \( S_{mn} \sim \mathcal{N}(0.5, 0.05) \). The abundance matrix \( A \in \mathbb{R}_+^{N \times T} \) was generated as shown in Fig. 4, where \( P \) of the columns of \( A \) were assigned (uniformly at random) to be pure pixels, and the remaining columns were drawn \( K \)-sparse on the simplex. In particular, for each of these latter columns, the support was drawn uniformly at random, and the non-zero values \( \{ \hat{a}_k \}_{k=1}^K \) were drawn from a Dirichlet distribution, i.e.,

\[
p(\hat{a}_1, \ldots, \hat{a}_K) = \frac{\Gamma(\alpha K)}{\Gamma(\alpha)} \prod_{k=1}^K \hat{a}_k^{\alpha - 1}, \quad \hat{a}_k \in [0, 1]
\]

\[
p(\hat{a}_K | \hat{a}_1, \ldots, \hat{a}_{K-1}) = \delta(1 - \hat{a}_1 - \cdots - \hat{a}_K),
\]

where \( \Gamma(\cdot) \) denotes the gamma function, with concentration parameter \( \alpha = 1 \). Finally, the observation matrix \( Y \) was created by adding white Gaussian noise \( W \) to \( Z = SA \), where the noise variance \( \psi \) was adjusted to achieve \( \text{SNR} \geq \frac{1}{\mathbb{E}[||Z||_2^2]} \).

We emphasize that neither spectral nor spatial coherence were used in this experiment. Thus, we turn off the spectral and spatial coherence exploitation in HUT-AMP, reducing our approach to EM-tuned BiG-AMP [20]. However, we emphasize that this application of EM-AugG-AMP differs from those in [33] in that it involves non-negativity and simplex constraints, i.e., it targets the NMF problem.

Figure 5 shows empirical success probability averaged over \( R = 100 \) realizations, as a function of pixel purity \( P \) and sparsity \( K \), for the HUT-AMP, MVSA, VCA, and FSNMF algorithms. Here, a recovery was considered successful if \( \text{NMSE}_S \leq |S - \hat{S}|_F^2 / \|S\|_F^2 < -40 \text{ dB} \). As seen in Fig. 5(c) and Fig. 5(d), VCA and FSNMF were only successful for the \( K = 1 \) and \( P = 10 \) cases, i.e., the pure-pixel cases. HUT-AMP, on the other hand, was able to successfully recover the endmembers for \( K \leq 6 \)-sparse abundances, even when there was only \( P = 1 \) pure-pixels available. We attribute HUT-AMP’s improved performance to its exploitation of sparsity rather than pure pixels (as with VCA and FSNMF). We also conjecture that sparsity (i.e., \( K > 1 \) and \( P < N \)) is more important in practice, since the spatial resolution of the hyperspectral sensors may not guarantee pixel-purity for all materials, while sparse abundances (i.e., \( K \ll N \)) are more likely to hold. Meanwhile, we note that, although MVSA performs remarkably well for this problem, its performance suffers when real-world endmembers are considered, as demonstrated by the experiments in the sequel.

#### B. Pure-Pixel Synthetic Abundances

The second experiment uses synthetic pure-pixel abundances \( A \) with endmembers \( S \) chosen from the USGS Digital Spectral Library splib06a\({}^4\) which contains laboratory-measured reflectance values for various materials over \( M =

\[3\]Matlab code can be found at http://www.ece.osu.edu/~schniter/HUTAMP

\[4\]See http://speclab.cr.usgs.gov/spectral.lib06/ds2317
224 spectral bands. To construct the data, we partitioned a scene of $T = 50 \times 50$ pixels into $N = 5$ equally sized vertical strips, each containing a single pure material. We then selected endmembers corresponding to the materials Grossular, Alunite, well crystallized (wxl) Kaolinite, Hydroxyl-Apatite, and Amphibole, noting that similar results were obtained in experiments we conducted with other materials. Finally, white Gaussian noise was added to achieve SNR = 30 dB. Figure 6 shows an RGB image constructed from the noiseless measurements $Z$.

Averaging over $R = 50$ noise realizations, Table I shows the normalized mean-squared error (i.e., $\text{NMSE}_S$ and $\text{NMSE}_A \equiv \| A - \hat{A} \|_F^2 / \| A \|_F^2$) of the estimated endmembers and abundances, the runtimes for the individual estimation of $S$ and $A$, and the total runtime. (For HUT-AMP and SCU, the estimation of $S$ and $A$ is done jointly, and thus only the total runtime is reported.)

For this pure-pixel dataset, Table I shows HUT-AMP dominating the other algorithms in both endmember and abundance estimation accuracy. In particular, HUT-AMP outperformed the best competing techniques by 9 dB in $\text{NMSE}_S$ and 60 dB in $\text{NMSE}_A$. We attribute HUT-AMP’s excellent NMSE to several factors. First, it has the ability to jointly estimate endmembers and abundances, to exploit spectral coherence in the endmembers, and to exploit both spatial coherence and sparsity in the abundances (of which there is plenty in this experiment). Furthermore, due to the presence of pure-pixels throughout the scene, the “active” distribution $\zeta_n(\cdot)$ in [18] is simply a Bernoulli distribution, which HUT-AMP is able to learn (via EM) and exploit (via BiG-AMP) for improved performance. Although SCU also performs joint estimation and is able to exploit spectral and spatial coherence, we conjecture that its priors are less well-matched to this highly sparse dataset. Regarding the NMSE advantage of FCLS over SUunSAL-TV, we attribute this to the fact that FCLS enforces the sum-to-one abundance constraint whereas SUunSAL-TV does not.

Table II also shows that the total runtime of HUT-AMP is comparable to that of the “EE-and-inversion” techniques, while being 230 times faster than that of the Bayesian joint unmixing algorithm, SCU. In fact, SCU took advantage of parallel processing over 8 cores, whereas the other algorithms all used a single core. We conjecture that the slower runtime of SCU is due to its use of Gibbs sampling. As for the EE-and-inversion techniques, we note that their total runtime is dominated by the inversion step, which is 2-3 orders-of-magnitude more time-consuming than the EE step.

Although not shown in Table II, we also ran HUT-AMP-MOS on this dataset. The result was that HUT-AMP-MOS correctly estimated the number of materials (i.e., $N = 5$) on every realization and thus gave identical $\text{NMSE}_S$ and $\text{NMSE}_A$ as HUT-AMP. The total runtime of HUT-AMP-MOS was 94.27 seconds, which was about 7 times slower than HUT-AMP but still 33 times faster than SCU.

### C. SHARE 2012 Avon Dataset

Next, we evaluated algorithm performance on the real-world SHARE 2012 Avon dataset [39], which uses $M = 360$ spectral bands, corresponding to wavelengths between 400 and 2450 nm, over a large rural area. To do this, we first cropped down the full image to the scene shown in Fig. 7 which is known to consist of $N = 4$ materials: grass, dry sand, black felt, and white TyVek [40]. This scene was explicitly constructed for use in hyperspectral unmixing experiments, as

---

**Table II**

| Technique          | $S$ time | $A$ time | Total time | $\text{NMSE}_S$ | $\text{NMSE}_A$ |
|--------------------|----------|----------|------------|-----------------|-----------------|
| HUT-AMP            | -        | -        | 13.75 sec  | -52.0 dB        | 91.8 dB         |
| SCU                | -        | -        | 3178 sec   | -39.9 dB        | -30.0 dB        |
| VCA + FCLS         | 0.15 sec | 6.15 sec | 6.31 sec   | -42.0 dB        | 31.5 dB         |
| VCA + SUunSAL-TV   | 0.15 sec | 1.61 sec | 11.76 sec  | -42.2 dB        | 27.6 dB         |
| FSNMF + FCLS       | 0.05 sec | 5.85 sec | 8.09 sec   | -45.0 dB        | 30.8 dB         |
| FSNMF + SUunSAL-TV | 0.05 sec | 1.51 sec | 11.56 sec  | -43.0 dB        | 29.7 dB         |
| MVSA + FCLS        | 0.93 sec | 3.66 sec | 4.59 sec   | -26.2 dB        | 18.9 dB         |
| MVSA + SUunSAL-TV  | 0.42 sec | 7.73 sec | 8.15 sec   | -26.1 dB        | 18.9 dB         |

---

5 The SHARE 2012 Avon dataset can be obtained from [http://www.rit.edu/cos/share2012/](http://www.rit.edu/cos/share2012/)
efforts were made to ensure that the vast majority of the pixels were pure. Also, the data was collected on a nearly cloudless day, implying that shadowing effects were minimal.

To construct “ground truth” endmembers we averaged a 4 × 4 pixel grid of the received spectra in a “pure” region for each material. We then computed the spectral angle distance

\[ \text{SAD}_n = \arccos \left( \frac{\langle s_n, \hat{s}_n \rangle}{\|s_n\|_2 \|\hat{s}_n\|_2} \right) \]  

between each ground-truth endmember \( s_n \), and the estimate \( \hat{s}_n \) produced by each algorithm. Table III shows median SAD over 50 trials, noting that VCA and SCU are random algorithms and thus exhibit variability across trials, even for this deterministic dataset. The table shows that HUT-AMP most accurately extracted all four endmembers. It also shows that SCU, VCA, and FSNMF all had a difficult time estimating the black felt endmember, most likely due to its low reflectivity. HUT-AMP, on the other hand, was able to leverage the spectral coherence of the black felt endmember, resulting in a 4 degree SAD improvement over the next best algorithm. Figure 8 shows an example of the extracted and ground-truth endmembers for visual comparison. It can be seen that HUT-AMP’s estimates are a much better match to the ground-truth for all but the dry-sand material, where VCA performed very well on this trial; on most other trials, VCA performed worse. Figure 8 reveals that MVSA does not always yield non-negative endmembers, which may account for its relatively poor performance in all but our first experiment.

Since ground-truth abundance maps are not available for the SHARE 2012 dataset, we do not present quantitative results on the accuracy of abundance estimation. We do, however, plot the recovered abundance maps in Fig. 9. In interpreting Fig. 9 we reason that the “best” recoveries are the ones that are the most pure within the green, tan, black, and white regions of Fig. 7; given that great care was taken during data collection to keep each region occupied by a single material. Figure 9 shows that, in the case of dry sand and black felt, the abundances recovered by HUT-AMP were the most pure and, in the case of grass and Tyvek, the abundances recovered by HUT-AMP were among the most pure. The other Bayesian approach, SCU, yielded abundance estimates with much less purity, and we conjecture that was due to its priors being less well-matched to this highly sparse scene. Meanwhile, SUnSAL-TV (using both EE techniques) failed to recover the black felt material, which we attribute to its lack of a sum-to-one constraint.

Figure 9 also reports the total runtime of each algorithm. There we see that HUT-AMP’s runtime was 3.6 times slower than the average EE-and-inversion technique but 230 times faster than SCU, the other Bayesian technique. We also ran HUT-AMP-MOS on the SHARE 2012 data and found that it correctly estimated the presence of \( N = 4 \) materials, thus yielding identical recovery performance to HUT-AMP. HUT-AMP-MOS’s runtime was 36.54 seconds, which was 3.5 times slower than HUT-AMP but still 67 times faster than SCU.

---

**TABLE III**  
**MEDIAN SPECTRAL ANGLE DISTANCE (IN DEGREES) BETWEEN RECOVERED AND GROUND-TRUTH ENDMEMBERS IN THE SHARE 2012 EXPERIMENT.**

| Material       | GRASS | DRY SAND | BLACK FELT | WHITE TYVEK |
|----------------|-------|----------|------------|-------------|
| HUT-AMP        | 1.46  | 0.79     | 3.14       | 0.45        |
| VCA            | 1.58  | 2.20     | 11.01      | 2.09        |
| FSNMF          | 1.65  | 1.68     | 7.36       | 1.46        |
| MVSA           | 4.57  | 10.42    | 45.47      | 1.60        |
| SCU            | 2.69  | 2.48     | 32.10      | 1.19        |

---

In real-world HU data, ground truth endmembers are hard to come by, since lab-measured reflectivity can differ dramatically from received radiance at the sensor. In this experiment, we exploit the known purity of the pixels and we minimize noise effects through averaging.

---

Fig. 7: RGB image of the cropped scene of the SHARE 2012 dataset.  
Fig. 8: Examples of recovered and ground-truth endmembers for the SHARE 2012 experiment.
(a) HUT-AMP (average runtime = 10.52 sec):

(b) VCA+FCLS (average runtime = 2.50 sec):

(c) VCA+SUnSAL-TV (average runtime = 4.13 sec):

(d) FSNMF+FCLS (average runtime = 1.67 sec):

(e) FSNMF+SUnSAL-TV (average runtime = 3.36 sec):

(f) MVSA+FCLS (average runtime = 1.82 sec):

(g) MVSA+SUnSAL-TV (average runtime = 4.08 sec):

(h) SCU (average runtime = 2438 sec):

Fig. 9. Examples of recovered abundance maps and average runtimes for the SHARE 2012 experiment. From left to right, the materials are: grass, dry sand, black felt, and white TyVeK.

D. AVIRIS Cuprite Dataset

Our final numerical experiment was performed on the well known AVIRIS Cuprite dataset. Although the original dataset consisted of $M = 224$ spectral bands, ranging from 0.4 to 2.5 $\mu$m, we aimed to replicate the setup in [17], which removed bands 1-10, 108-113, 153-168, and 223-224 to avoid water-absorption effects, resulting in $M = 189$ spectral measurements per pixel. And, like in [17], we considered only the 80x80 pixel scene identified by the black square in Fig. 10. According to the tricorder classification map in Fig. 10, this scene contains 4 materials: Montmorillonite, Alunite, well crystallized (wxl) Kaolinite, and partially crystallized (pxl) Kaolinite. But, as in [3], we allow for the additional presence of Desert Varnish, bringing the total number of materials to $N = 5$. Also, like in [17], we consider both noiseless and white-Gaussian-noise corrupted measurements (at SNR = 30 dB).

Table IV shows the median SAD achieved during endmember extraction over 50 trials. (We used median for consistency with [17], and note that the performance of both VCA and SCU varied over trials due to the random nature of each algorithm.) Table IV shows that, in the noiseless case, HUT-AMP achieved the best median SAD for three materials, while FSNMF achieved the best for two and SCU achieved the best for one. Meanwhile, in the noisy case, HUT-AMP achieved the best median SAD for two materials, while SCU, VCA, and FSNMF each achieved the best for one material. However, the SAD values in the table should be interpreted with caution, since i) they are based on the use of laboratory-measured reflectance spectra from the 2006 USGS library as ground-truth, whereas the Cuprite dataset itself uses reflectance units obtained via atmospheric correction of radiance data; and ii) they are premised on the assumption that these particular materials are actually present in the scene.

Although a lack of ground truth prevents us from quantitatively assessing abundance-map recovery performance for the Cuprite data, we plot the recovered (noiseless) abundance

7The reflectance and radiance versions of the Cuprite dataset can be found at [http://aviris.jpl.nasa.gov/html/aviris.freedata.html](http://aviris.jpl.nasa.gov/html/aviris.freedata.html)
maps in Fig. 11 for visual comparison to the classification map in Fig. 10. Figure 11 shows that the abundance maps produced by $\text{SUnSAL-TV}$ appear more “blurred,” probably as an artifact of TV regularization. The abundances returned by $\text{SCU}$, $\text{MVSA+FCLS}$, and $\text{MVSA+SUnSAL-TV}$ were of much lower contrast (i.e., much less sparse) and suggest different material placements than the maps generated by the other algorithms. For example, $\text{SCU}$ suggests a significant $\text{WxI-Kaolin}$ presence throughout the lower half of the scene, in contrast to other algorithms. However, Table IV shows that $\text{SCU}$ gave the worst SAD for $\text{WxI-Kaolin}$.

Figure 11 also shows the total runtimes of the various approaches. There we see that $\text{HUT-AMP}$ is 5.6 times slower than the average $\text{VCA}$ or $\text{FCLS}$ based approach, while $\text{SCU}$ is 460 times slower. We also ran $\text{HUT-AMP-MOS}$ on the Cuprite data and found that, in both the noiseless and noisy cases, it estimated the presence of $N = 5$ materials, and thus returned identical estimates to $\text{HUT-AMP}$. Meanwhile, $\text{HUT-AMP-MOS}$ had an average runtime of 191.49 seconds, which is 30 times faster than $\text{SCU}$.

V. CONCLUSIONS

In this paper, we proposed a novel Bayesian hyperspectral-unmixing algorithm that jointly estimates endmembers and abundance maps while exploiting the spectral and spatial coherence, as well as the abundance sparsity, that is often present in practice. To perform the overall inference task, we used the “turbol” approach suggested in [24], which breaks up the factor graph into three subgraphs, performs (loopy) BP individually on each subgraph, and then exchanges beliefs between subgraphs. For the spectral and spatial coherence subgraphs, we used standard Gauss-Markov and discrete-Markov methods [18], [19], respectively, while for the non-negative bilinear mixing subgraph, we use the recently proposed BiGAMP algorithm from [20], which exploits the approximate message passing framework from [21], [22]. Furthermore, we tune our prior and likelihood parameters using expectation-maximization, and we estimate the number of materials in the scene using penalized log-likelihood maximization.

Through a detailed numerical study, we demonstrated that our proposed $\text{HUT-AMP}$ algorithm yields excellent recoveries on both synthetic and real-world datasets, and in many cases outperforms the other state-of-the-art algorithms under test. For example, our results suggest that $\text{HUT-AMP}$ can reliably recover endmembers in the absence of pixel purity, unlike those endmember extraction algorithms designed around the pure-pixel assumption (e.g., $\text{VCA}$, $\text{FSNMF}$). Moreover, the runtime of $\text{HUT-AMP}$ is on par with those of other EE-and-inversion techniques (e.g., $\text{VCA+FCLS}$), in contrast to other Bayesian spectral/spatial-coherence exploitation techniques like SCU, whose runtime is 2-3 orders-of-magnitude larger. Our experiments also demonstrated that our model-order selection technique was able to correctly estimate the number of materials in several synthetic and real-world datasets, without requiring a very large increase in runtime.

APPENDIX A

MEAN REMOVAL

We can see that $\mathbf{S}$ from (6) is approximately zero-mean via

$$0 = \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} \mathbf{w}_{mt}$$

$$= \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} \sum_{n=1}^{N} \mathbf{s}_{mn} a_{nt} + \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} \mathbf{w}_{nt}$$

$$\approx O(1) + O(1/N)$$

$$\triangleq \mu_{n}^{(t)}$$

where (49) follows from the definitions (4)-(5). The underbraces in (50) show the scaling on each term in the large-system limit (i.e., as $N \rightarrow \infty$). These particular scalings follow from our assumption that the noise is both zero-mean and white and the convention [20] that both $\mathbf{y}_{mn}$ and the noise variance $\psi$ scale as $O(1)$. Recalling that $\sum_{n=1}^{N} \mu_{n}^{(t)} = 1$ due to the simplex constraint, expression (51) shows that a weighted average of elements in $\mathbf{S}$ is approximately zero, where the approximation becomes exact in the large-system limit.

REFERENCES

[1] J. Vila, P. Schniter, and J. Meola, “Hyperspectral image unmixing via bilinear generalized approximate message passing,” Proc. SPIE, vol. 8743, no. 87430Y, p. 9, 2013.

[2] P. E. Johnson, M. O. Smith, S. Taylor-George, and J. B. Adams, “A semiempirical method for analysis of the reflectance spectra of binary mineral mixtures,” J. Geophysical Research, vol. 88, pp. 3557–3561, 1983.

[3] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, “Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches,” IEEE J. Sel. Topics Appl. Earth Observ., vol. 5, no. 2, pp. 354–379, 2012.

[4] M. E. Winter, “N-FINDR: an algorithm for fast autonomous spectral endmember determination in hyperspectral data,” Proc. SPIE, pp. 266–275, 1999.

[5] J. Nascimento and J. Bioucas Dias, “Vertex component analysis: a fast algorithm to unmix hyperspectral data,” IEEE Trans. Geosci. Remote Sens., vol. 43, no. 4, pp. 898–910, 2005.
Fig. 11. Examples of recovered abundance maps in the noiseless Cuprite experiment. Each row corresponds to an algorithm and each column corresponds to a material. Average runtimes are also listed on the left.

[6] D. L. Donoho and V. Stodden, “When does non-negative matrix factorization give a correct decomposition into parts?” in Proc. Neural Inform. Process. Syst. Conf., 2003.

[7] S. Arora, R. Ge, R. Kannan, and A. Moitra, “Computing a nonnegative
matrix factorization—provably,” in *Proc. Sym. Thy. Computing*, 2012, pp. 145–162.

[8] N. Gillis and S. A. Vavasis, “Fast and robust recursive algorithms for separable nonnegative matrix factorization,” *arXiv:1208.1237v2*, 2012.

[9] J. Li and J. Bioucas-Dias, “Minimum volume simplex analysis: A fast algorithm to umix hyperspectral data,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, 2008, pp. III–250–III–253.

[10] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, “A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing,” *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[11] D. Heinz and C.-I. Chang, “Fully constrained least squares linear spectral unmixing by simplex projection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, 2011.

[12] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, “Sparse unmixing of hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, 2011.

[13] ——, “Total variation spatial regularization for sparse hyperspectral unmixing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[14] A. Chambolle, “An algorithm for total variation minimization and applications,” *J. Math. Imaging Vis.*, vol. 20, pp. 89–97, 2004.

[15] N. Dobigeon, S. Moussaoui, M. Coulon, J. Y. Tourneret, and A. Hero, “Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery,” *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, 2009.

[16] R. Mittelman, N. Dobigeon, and A. Hero, “Hyperspectral image unmixing using a multiresolution sticky hdp,” *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1656–1671, 2012.

[17] C. A. Bouman, “Markov random fields and stochastic image models,” in *IEEE Int. Conf. Image Processing Tutorial*, Oct. 1995.

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

[19] J. T. Parker, P. Schniter, and V. Cevher, “Bilinear generalized approximate message passing—Part I: Derivation,” *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5854–5867, Nov. 2014, (See also *arXiv:1310.2632*).

[20] M. I. Jordan, Ed. *MIT Press*, 1999, pp. 355–368.

[21] 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. MIT Press, 1999, pp. 355–368.

[22] K. Canham, D. Goldberg, J. Kerekes, N. Raqueno, and D. Messinger, “SHARE 2012: Large edge targets for hyperspectral imaging applications,” *Proc. SPIE*, vol. 8743, no. 87430G, p. 9, 2013.

[23] R. N. Clark, G. A. Swayeze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R. McDougall, and C. A. Gent, “Imaging spectroscopy: Earth and planetary remote sensing with the USGS tetraekord and expert systems,” *J. Geophys. Res.*, vol. 108, no. E12, p. 5131, 2003.