# ROBUST SPECTRAL UNMIXING FOR ANOMALY DETECTION

*Gregory E. Newstadt, Alfred O. Hero III*

University of Michigan
Electrical Engineering and Computer Science
Ann Arbor, Michigan 48109
Email: {newstage},{hero}@umich.edu

*Jeff Simmons*

Air Force Research Laboratory
Wright Patterson Air Force Base, OH 45433
Email: jeff.simmons@wpafb.af.mil

## ABSTRACT

This paper is concerned with a joint Bayesian formulation for determining the endmembers and abundances of hyperspectral images along with sparse outliers which can lead to estimation errors unless accounted for. We present an inference method that generalizes previous work and provides a MCMC estimate of the posterior distribution. The proposed method is compared empirically to state-of-the-art algorithms, showing lower reconstruction and detection errors.

## 1. INTRODUCTION

A large amount of recent research has focused on the spectral unmixing problem, wherein images collected at multiple frequencies are decomposed into a product form of so-called *endmembers* and *abundances*. The endmembers represent the basic spectra building blocks from which any pixel is constructed, while the abundances represent the mixing proportions. There are methods for estimating the endmembers first, such as with N-FINDER [1], followed by estimating the abundances using least squares or Bayesian methods. Fully Bayesian methods have been proposed as well [2].

There has also been great interest in the so-called robust principal component analysis (RPCA) problem [3–5] that decomposes high-dimensional signals into low-rank, sparse, and noise components. The sparse component can generally not be represented easily by the low-rank model and thus contributes to estimation errors when unaccounted for. In [3, 4], inference in this model is done by maximizing an objective function that promotes a sparse number of factors (i.e. endmembers) through the nuclear norm, a sparse number of outliers through the $l_1$ norm, and robustness to noise through the Frobenius norm. However, one drawback of these methods involves finding the tradeoff parameters between these three objectives, which in general may depend on the given signal. Bayesian methods by Ding et al. [5] have been proposed that simultaneously learn the noise statistics and infer the low-rank and sparse components. Moreover, they show that their method can be generalized to richer models, e.g. Markov dependencies on the target locations.

This work provides a generalization of [5] to the case where the low-rank component can be given either by a singular value decomposition (SVD) or a non-negative matrix factorization (as in [2]). Moreover, we consider spectral learning simultaneously with anomaly detection by including a sparse component. Additionally, we extend the previous work by explicitly allowing for correlated and group sparse anomalies over local pixels (spatial) and local frequency bins. We provide an inference algorithm based on MCMC methods which provides an estimate of the posterior distribution. Finally, we compare this algorithm on a simulated dataset to [2,5], and achieve lower reconstruction and detection errors.

The Bayesian spectral unmixing algorithm proposed here has been applied to anomaly detection in energy-dispersive X-ray spectroscopy (EDS) images for the purpose of detecting defects in materials. The authors of [6, 7] demonstrate the benefits of using spectral unmixing to decompose EDS images into a few characteristic spectra. Our anomaly-driven algorithm was developed to improve on previous EDS unmixing methods by: (a) simultaneously learning both the endmembers and the anomalies and (b) providing uncertainty characterization through the posterior distribution. Due to lack of space, we have not included the real-data EDS application in this paper. The application to EDS is reported in an extended paper [8].

## 2. BAYESIAN MODEL

Similar to [5], we propose a decomposition of the observed high-dimensional signal $\boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{S} + \boldsymbol{N}$, where $\boldsymbol{L}$ is a low-rank matrix, $\boldsymbol{S}$ is a sparse component, and $\boldsymbol{N}$ is dense low-amplitude noise. Each of these components belongs to the space $\mathbb{R}^{F\times P}$, where $P$ is the number of pixels and $F$ is the number of frequencies. Note that many signals can be decomposed this way. This includes (a) video processing where $\boldsymbol{L}$ represents the stationary background and $\boldsymbol{S}$ represents sparse moving targets, and (b) material analysis where $\boldsymbol{L}$ are the basic constituents of a material while $\boldsymbol{S}$ are defects.

### 2.1. Low-rank component, $\boldsymbol{L}$

The low-rank component can be modeled in many ways, depending on the application. In this work, we consider two basic models for $\boldsymbol{L}$: (a) singular value decomposition and (b) non-negative matrix factorization. In (a), we can state the model as $\boldsymbol{L} = \boldsymbol{D}\boldsymbol{\Lambda}\boldsymbol{W}^T$ where $\boldsymbol{D} = [d_1 d_2 \cdots d_R] \in \mathbb{R}^{F\times R}$ and $\boldsymbol{W} = [w_1 w_2 \cdots w_R] \in \mathbb{R}^{P\times R}$ are matrices of the left- and right-singular vectors, respectively, and $\boldsymbol{\Lambda} = \text{diag}\{\lambda_r\}_{r=1}^R$ is a diagonal matrix consisting of the singular values. We follow the model proposed by [5] so that $\lambda_r = z_r\delta_r, \quad r = 1, 2, \ldots, R$, where $z_i \in \{0, 1\}$ $\delta_i \in \mathbb{R}$, and $\boldsymbol{Z}$, $\boldsymbol{\Delta}$ are the vector quantities. As in [5], this decouples learning the rank structure (i.e., $\|\boldsymbol{Z}\|_0$) from the learning of the singular vectors.

We also consider the case where the underlying structure has non-negativity constraints, and sum-to-one constraints on the factor loadings. Consider the linear mixing model (LMM) for the observed spectrum of the $p$-th pixel, $\boldsymbol{l}_p$ [2]:

$$\boldsymbol{l}_p = \sum_{r=1}^{R} \boldsymbol{m}_r a_{p,r}, \quad p = 1, 2, \ldots, P, \tag{1}$$

$$a_{p,r} \geq 0, \sum_{r=1}^{R} a_{p,r} = 1, \quad \forall r = 1, 2, \ldots, R, \tag{2}$$

$$m_{r,f} \geq 0, \quad \forall r = 1, 2, \ldots, R, \forall f = 1, 2, \ldots, F \tag{3}$$

where $a_{p,r}$ are the factor loadings and $\boldsymbol{m}_r$ are the factors. For either of these models, it is possible to propose a fully Bayesian model from which we can derive a MCMC sampling strategy. Details are given for both models in [5] and [2], respectively. For simplicity, we will provide details only for the former case and refer the reader to [2] for the other. In particular, we assume Bayesian priors on the variables in the following way for $r = 1, 2, \ldots, R$:

$$z_r \sim \text{Bernoulli}\left(\pi_r^z\right), \quad \boldsymbol{d}_r \sim \text{Normal}\left(0, \boldsymbol{I}_{F\times F}/F\right), \tag{4}$$

$$\delta_r \sim \text{Normal}\left(0, 1/\tau^z\right) \quad \boldsymbol{w}_r \sim \text{Normal}\left(0, \boldsymbol{I}_{P\times P}/P\right) \tag{5}$$

In this model, we assume that the parameters $\{\pi_r^z\}_r$ and $\tau^z$ are also random (and hence estimated from the observed data). We assume priors of the form for $\{\pi_r^z\}_{r=1}^R$ and $\tau^z$:

$$\pi_r^z \sim \text{Beta}\left(a_0, b_0\right), \tau^z \sim \text{Gamma}\left(c_0, d_0\right) \tag{6}$$

Note that these parameters form conjugate pairs with the distributions in equations (4)-(5). Thus, the posterior distributions are known in closed-form, which leads to efficient sampling strategies. The parameters $\{a_0, b_0, c_0, d_0\}$ are chosen in accordance with [5]. In particular, we choose $a_0/(a_0 + b_0) \ll 1$ to promote sparsity in the number of factors, and let $c_0 = d_0 = 10^{-6}$ to yield a non-informative prior on the precision $\tau^z$.

### 2.2. Sparse component, $\boldsymbol{S}$

We consider two models for the sparse component $\boldsymbol{S}$. In particular, we begin with the model in [5], where $\boldsymbol{S} = \boldsymbol{B} \circ \boldsymbol{X}$ with $\boldsymbol{B} \in \{0,1\}^{F\times P}$ and $\boldsymbol{X} \in \mathbb{R}^{F\times P}$, and $\circ$ represents the element-wise product of two-matrices. This separates the learning of the locations of the sparse components from their values. Moreover, we consider a Bayesian model wherein the elements $b_{f,p}$ and $x_{f,p}$ have conjugate distributions for $f = 1, \ldots, F$ and $p = 1, \ldots, P$.

$$b_{f,p} \sim \text{Bernoulli}\left(\pi_{f,p}\right), \quad \pi_{f,p} \sim \text{Beta}\left(a_1, b_1\right) \tag{7}$$

$$x_{f,p} \sim \text{Normal}\left(0, 1/\tau^x\right), \quad \tau^x \sim \text{Gamma}\left(c_1, d_1\right). \tag{8}$$

As in [5], it is easy to impose additional structure on $\pi_{f,p}$ such as including a Markov property to account for the fact that pixels near detected anomalies are more likely to be anomalous as well. We extend this idea in two ways: (a) we include group sparsity directly (both in the space and frequency domains), and (b) we include the ability to model correlated anomalies across that group. Figure 1 illustrates the group sparsity model compared to the independent anomaly model.

For clarity, assume that we can partition the pixel set $\{1, 2, \ldots, P\}$ into $\kappa_P$ disjoint sets of size $L$, and the frequency set $\{1, 2, \ldots, F\}$ into $\kappa_F$ disjoint sets of size $M$. Then we have

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{1,1} & \cdots & \boldsymbol{S}_{1,\kappa_P} \\ \vdots & \ddots & \vdots \\ \boldsymbol{S}_{\kappa_F,1} & \cdots & \boldsymbol{S}_{\kappa_F,\kappa_P} \end{bmatrix} \tag{9}$$

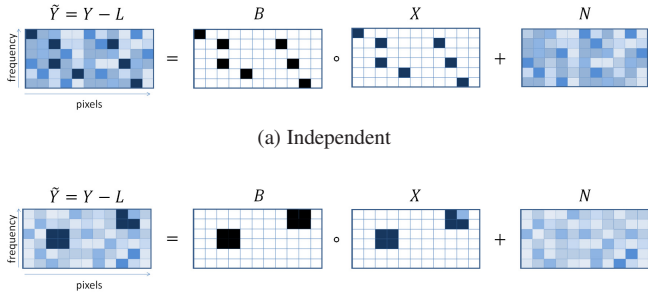
where for each $i = 1, \ldots, \kappa_F$, $j = 1, \ldots, \kappa_P$, we have

$$b_{i,j} \sim \text{Bernoulli}\left(\pi_{i,j}^b\right), \quad \pi_{i,j}^b \sim \text{Beta}\left(a_1, b_1\right) \tag{10}$$

$$\tilde{\boldsymbol{X}}_{i,j} \sim \text{Normal}\left(0, \boldsymbol{\Sigma}^X\right), \quad S_{i,j} = b_{i,j}\boldsymbol{X}_{i,j} \tag{11}$$

where $\tilde{\boldsymbol{X}}_{i,j}$ is a vectorized version of $\boldsymbol{X}_{i,j}$ and

$$\boldsymbol{\Sigma}^X = \left[(1-\rho^x)\boldsymbol{I}_{LM\times LM} + \rho^x \mathbf{1}_{LM}\mathbf{1}_{LM}^T\right]/\tau^x \tag{12}$$




(a) Independent




(b) Group sparse

**Fig. 1**. This figure displays the group sparsity model for anomalies in (b), compared to the standard independent model in (a).

**procedure** $\{\boldsymbol{\Theta}\}_{i=1:N_{samples}} = \text{RSU}(\boldsymbol{\Theta}_0, \boldsymbol{Y})$
  $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta}_0$
  **for** $iteration = 1$ to $N_{burnin} + N_{samples}$ **do**
    Sample $\sim f\left(\boldsymbol{D}, \boldsymbol{Z}, \boldsymbol{\Delta}, \boldsymbol{W} | \boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{\eta}\right)$ *//*$\boldsymbol{L} = \boldsymbol{D}(\boldsymbol{Z}\boldsymbol{\Delta})\boldsymbol{W}$
    Sample $\sim f\left(\boldsymbol{X}, \boldsymbol{B} | \boldsymbol{Y}, \boldsymbol{L}, \boldsymbol{\eta}\right)$ *//*$\boldsymbol{S} = \boldsymbol{B} \circ \boldsymbol{X}$
    Sample $\sim f\left(\{\pi_r^z\}_r | \boldsymbol{Y}, \boldsymbol{L}, \tau^n\right)$ *//SVD factor probs.*
    Sample $\sim f\left(\tau^z | \boldsymbol{Y}, \boldsymbol{L}, \tau^n\right)$ *//SVD precision*
    Sample $\sim f\left(\{\pi_{i,j}^b\}_{i,j} | \boldsymbol{Y}, \boldsymbol{S}\right)$ *//Anom. probs.*
    Sample $\sim f\left(\tau^x | \boldsymbol{Y}, \boldsymbol{S}, \tau^n\right)$ *//Anom. precision*
    Sample $\sim f\left(\rho^x | \boldsymbol{Y}, \boldsymbol{S}, \tau^n, \tau^x\right)$ *//Anom. correlation*
    Sample $\sim f\left(\tau^n | \boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{L}\right)$ *//Noise precision*
    $\boldsymbol{\Theta}_{iteration-N_{burnin}} \leftarrow \boldsymbol{\Theta}$ **if** $iteration > N_{burnin}$
  **end for**
**end procedure**

**Fig. 2**. Gibbs Sampling Pseudocode using SVD model for $\boldsymbol{L}$

where $\rho^x \sim \text{Beta}\left(a_2, b_2\right)$ is the correlation coefficient, which may be application-specific. Note that $\rho^X = 0$ represents an IID model, while $\rho^X \approx 1$ promotes highly correlated anomalies. It is assumed that $a_2$ and $b_2$ are chosen to reflect the application, and $\tau^x$ is given by (8). By including group structure over the indicator and/or amplitude variables, we improve the power of detecting the anomalies.

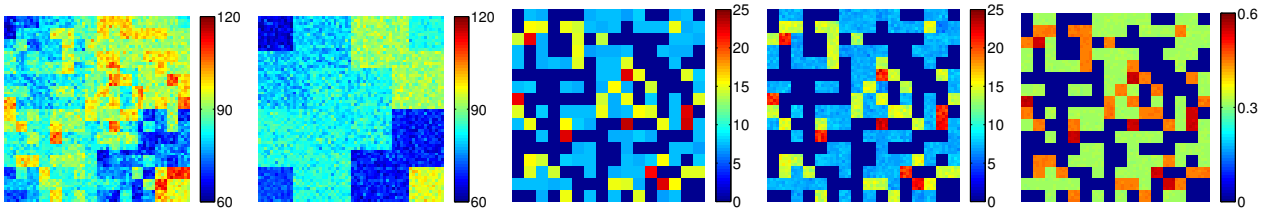
### 2.3. Noise component, $\boldsymbol{N}$

We model the noise $\boldsymbol{N}$ as being IID zero-mean noise, with each element $n_{f,p}$ distributed as

$$n_{f,p} \sim \text{Normal}\left(0, 1/\tau^n\right), \quad \tau^n \sim \text{Gamma}\left(c_2, d_2\right) \tag{13}$$

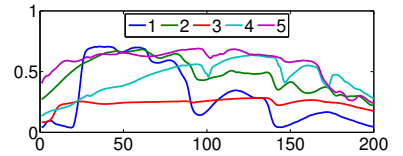
## 3. INFERENCE ALGORITHM

In this section, we provide details on estimating the posterior distribution of the model parameters given the observations. It should be noted that given the posterior distribution, one can perform many appropriate inference tasks, such as providing the maximum a posteriori (MAP) estimate, confidence regions, and probabilities for the existence of the anomaly. These are in contrast to point estimates that are available from standard maximum likelihood estimates, and this is one of the primary reasons for using the Bayesian method. For example, in Fig. 3(d) and (e), we show the estimated mean and standard deviations of the sparse component.

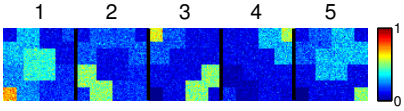



(a) Sample $\boldsymbol{Y}$ (RMS) (b) Sample $\boldsymbol{L}$ (RMS) (c) Sample $\boldsymbol{S}$ (RMS) (d) $\mathbb{E}\{\hat{\boldsymbol{S}}\}$ (RSU-LMM) (e) $\mathrm{STD}\{\hat{\boldsymbol{S}}\}$ (RSU-LMM)

**Fig. 3**. We show the observed signal (a) (RMS over frequency) for the simulations in Section 4, the true low dimensional component (b), and the true sparse component in (c). The mean estimate from RSU-LMM is given in (d), and its associated standard deviation in (e). Note that the standard deviation can be used to quantify a confidence on the estimates in (d), which is one key benefit to using the Bayesian method.




(a) Factors




(b) Loadings

**Fig. 4**. We show the factors (a) and loadings (b) used to construct the low dimensional component for Section 4. There are very few loadings near 1, which indicates that this is a highly mixed model.

Generally, estimating the posterior distribution on this model would be a very difficult task due to the large number of variables and the dependence among them. In particular, we use a Markov Chain Monte Carlo (MCMC) algorithm in the form of a Gibbs sampler to iteratively estimate the full joint posterior. In MCMC, this distribution is approximated by drawing samples iteratively from the conditional distribution of each (random) model variable given the most recent estimate of the rest of the variables. Let $\Theta = \{\boldsymbol{D}, \boldsymbol{Z}, \boldsymbol{\Delta}, \boldsymbol{W}, \boldsymbol{B}, \boldsymbol{X}, \boldsymbol{\eta}\}$ represent a current estimate of all of the model variables, where $\boldsymbol{\eta} = \left\{\{\pi_r^z\}_r, \tau^z, \{\pi_{i,j}^b\}_{i,j}, \rho^x, \tau^x, \tau^n\right\}$ is the set of all hyperparameters[1]. Given measurements $\boldsymbol{Y}$, the robust spectral unmixing (RSU) inference algorithm is given in Figure 2. We denote RSU-LMM and RSU-SVD for the LMM and SVD models for $\boldsymbol{L}$, respectively. Note that MCMC algorithms require a burn-in period, after which the Markov chain has become stable. The duration of the burn-in period depends on the problem. After the Markov chain has become stable, we collect $N_{samples}$ samples that represent the full joint distribution.

We refer the reader to [5] and [2] for the full sampling details for the low-dimensional component. For the sparse component, we note the following decomposition:

$$\begin{aligned} f(\boldsymbol{X}, \boldsymbol{B}|\boldsymbol{Y}, \boldsymbol{L}, \boldsymbol{\eta}) &= \prod_{i,j} f(\boldsymbol{X}_{i,j}, b_{i,j}|\boldsymbol{Y}, \boldsymbol{L}, \boldsymbol{\eta}) \\ &= \prod_{i,j} f(\boldsymbol{X}_{i,j}|b_{i,j}, \boldsymbol{Y}, \boldsymbol{L}, \boldsymbol{\eta})\, f(b_{i,j}|\boldsymbol{Y}, \boldsymbol{L}, \boldsymbol{\eta}) \end{aligned} \tag{14}$$

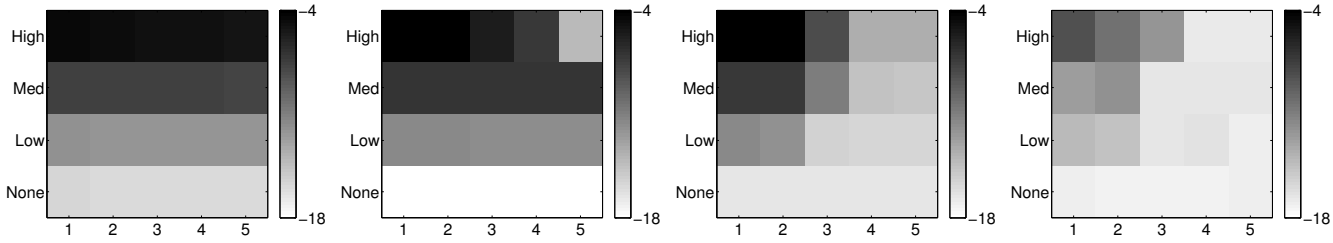
The latter part is just a Bernoulli distribution, which can be computed by noting the $f(\boldsymbol{Y}_{i,j}|\boldsymbol{L}, \boldsymbol{\eta}, b_{i,j})$ is normally-distributed. Moreover, the former component conditioned on $b_{i,j}$ is either zero (if $b_{i,j} = 0$) or normally distributed with known covariance (since we condition on $\boldsymbol{\eta}$). By jointly sampling $\boldsymbol{B}$ and $\boldsymbol{X}$, we accelerate convergence (and reduce the number of burn-in samples needed).

For space considerations, we do not provide full sampling procedures for the hyperparameters, but note that each sampling step can be described by one of three tasks: (1) sampling from a Beta distribution (for probabilities); (2) sampling from a Gamma distribution (for precisions); and (3) sampling from a Metropolis-Hastings step (for the correlation coefficient). The first two follow simply from using conjugate distribution pairs, where the posterior distribution has a closed-form expression. However, there is no closed-form expression for the posterior distribution of $\rho^x$, though it can be shown that the distribution of $\boldsymbol{\Sigma}^x$ follows a Inverse-Wishart distribution when conditioned on $\rho^x, \boldsymbol{Y}$ and $\tau^x$. Thus, we can use a Metropolis-Hastings step using a random walk proposal distribution in order to update $\rho^x$. Since this is an update of a single scalar, the additional computational load is minimal.

## 4. APPLICATION: SIMULATED DATASET

We compare the performance of the proposed inference algorithms to BLU [2] and BRPCA [5] over a simulated dataset. In particular, the dataset consisted of multiple samples of size 64x64 spatial locations and 200 frequencies. Each sample (as shown in Fig. 3) had an identical low-dimensional component, which was constructed using a LMM with factors/loadings given in Fig. 4. The sparse component was constructed with $X_{i,j} \sim \mathrm{Normal}(\mu_x, \sigma_x^2)$ and $b_{i,j} \sim \mathrm{Bernoulli}(0.1)$, where group sparsity was imposed with $L = 25$ and $M = 4$. Moreover, we considered three cases where $\mu_x = \{0.1, 0.3, 0.6\}$ and $\sigma_x^2 = 0.0025$. Note that increasing $\mu_x$ leads to higher detectability, which eases inference in models that include an sparse component. We also generated samples without any anomalies. Finally, all samples were corrupted with zero-mean noise with variance 0.004.

[1]Note that this definition would be slightly different if we used the non-negative matrix factorization instead of the SVD formulation.




(a) BLU (b) BRPCA (c) RSU-SVD (d) RSU-LMM

**Fig. 5**. These figures show the reconstructive error (in dB) while varying the number of samples (1 to 5) and the type of anomaly (none, low, med., high). In (a), results are given for BLU which does not account for anomalies. (b) compares to BRPCA, while (c) and (d) provide results for RSU-SVD and RSU-LMM, respectively. Lighter colors indicate better performance.

| Samples | Anom. Type | BRPCA | | RSU-SVD | | RSU-LMM | |
|---|---|---|---|---|---|---|---|
| | | $e(\hat{\boldsymbol{S}})$ | $e(\hat{\boldsymbol{B}})$ | $e(\hat{\boldsymbol{S}})$ | $e(\hat{\boldsymbol{B}})$ | $e(\hat{\boldsymbol{S}})$ | $e(\hat{\boldsymbol{B}})$ |
| 1 | Low | 1.96 | 0.99 | 0.97 | 0.96 | 0.59 | 0.55 |
| | Med | 1.15 | 0.95 | 0.88 | 0.87 | 0.31 | 0.22 |
| | Low | 1.02 | 0.96 | 0.90 | 0.89 | 0.33 | 0.28 |
| 3 | Low | 1.88 | 0.99 | 0.46 | 0.42 | 0.42 | 0.33 |
| | Med | 1.14 | 0.95 | 0.41 | 0.30 | 0.21 | 0.02 |
| | Low | 0.50 | 0.61 | 0.37 | 0.26 | 0.17 | 0.07 |
| 5 | Low | 1.92 | 0.99 | 0.43 | 0.38 | 0.41 | 0.29 |
| | Med | 1.12 | 0.95 | 0.22 | 0.07 | 0.21 | 0.02 |
| | Low | 0.31 | 0.08 | 0.14 | 0.01 | 0.11 | 0.00 |

**Table 1**. Sparse reconstruction errors (red: error less than 10%)

In Figure 5, we show the reconstruction error in the low-dimensional component $\|\boldsymbol{L} - \hat{\boldsymbol{L}}\|_2/\|\boldsymbol{L}\|_2$ for the inference algorithms described above. We show performance for various sparse component amplitudes (Y-axes) and number of samples (X-axes). Note that using more samples just increased the number of total pixels, but we did not include any additional information between samples. The plots indicate that BLU (a) suffered greatly in performance when large anomalies were present. Moreover, comparing BRPCA (b) and RSU-SVD (c), we see that using the group-sparse model in (c) improves the ability to estimate $\boldsymbol{S}$ which in turns provides additional fidelity in estimating $\boldsymbol{L}$. Finally, RSU-LMM, which matches the simulated data the best, provides the best results in (d).

In Table 1, we show the reconstruction error in the sparse component $e(\hat{\boldsymbol{S}}) = \|\boldsymbol{S} - \hat{\boldsymbol{S}}\|_2/\|\boldsymbol{S}\|_2$, and the detection error $e(\hat{\boldsymbol{B}}) = \|\boldsymbol{B} - \hat{\boldsymbol{B}}\|_2/\|\boldsymbol{B}\|_2$ for the three algorithms that include the sparse component. It is seen that RSU-LMM performs the best in both metrics over all anomaly types and number of samples. Since this algorithm is best matched to the simulated data, it is the most capable of detecting the anomalies either when there are fewer samples or the anomalies have lower means (i.e., medium vs. high types). Moreover, the RSU-SVD algorithm outperforms BRPCA because of its ability to detect group sparse anomalies which improves its detection power.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we generalize previous work [2, 5] by providing a formulation for robust spectral unmixing. In particular, we directly estimate sparse outliers in spectral data simultaneously with the endmembers/abundances of the low-rank component. We provide a fully Bayesian model, along with a MCMC inference method that provides an estimate of the posterior distribution. Finally, we demonstrated the validity of the algorithm on a simulated dataset and showed that it performs better in terms of reconstruction and detection errors in comparison to state-of-art algorithms.

## 6. ACKNOWLEDGMENT

The research in this paper was partially supported by Air Force Office of Scientific Research award FA9550-06-1-0324, by Air Force Research Laboratory award FA8650-07-D-1221-TO1, and by USAF/AFMC award FA8650-9-D-5037/04.

## 7. REFERENCES

[1] M. E. Winter, "N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*. International Society for Optics and Photonics, 1999, pp. 266–275.

[2] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4355–4368, 2009.

[3] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," *Journal of the ACM*, 2009.

[4] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, 2011.

[5] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. Image Process*, vol. 20, 2011.

[6] P. G. Kotula, M. R. Keenan, and J. R. Michael, "Automated analysis of sem x-ray spectral images: a powerful new microanalysis tool," *Microscopy and Microanalysis*, vol. 9, no. 01, pp. 1–17, 2003.

[7] M. R. Keenan and P. G. Kotula, "Accounting for poisson noise in the multivariate analysis of tof-sims spectrum images," *Surface and Interface Analysis*, vol. 36, no. 3, pp. 203–212, 2004.

[8] G. Newstadt, A. Hero III, and J. Simmons, "Bayesian robust spectral unmixing for anomaly detection," University of Michigan, Communications and Signal Processing Lab., Tech. Rep., 2014.