PERFORMANCE GUARANTEES FOR SPARSE REGRESSION-BASED UNMIXING

Yuki Itoh, Marco F. Duarte, and Mario Parente

Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003

ABSTRACT

Sparse regression-based unmixing has received much attention in recent years; however, its theoretical performance has not been explored in the literature. In this work, we present theoretical guarantees for the performance of a sparse regression based unmixing (in short, sparse unmixing) implemented in the form of a Lasso optimization with non-negativity constraints. We provide a sufficient condition required for the exact recovery of the endmembers and validate it both theoretically and through experiments. In cases in which the condition is not verified, we explore the performance of sparse unmixing in relation to the exact recovery coefficient (ERC).

Index Terms— sparse regression, unmixing, non-linear mixing, Hapke model

1. INTRODUCTION

Spectral unmixing (SU) aims at detecting the pure components (i.e., *endmembers*) of mixed spectra and to estimate endmember fractional abundances. A significant amount of research has been developed in unmixing methods [1, 2] and, among them, sparse regression-based unmixing (SRU) approaches have received much attention in recent years [3–6]. SRU takes the benefit of the availability of a large library (or *dictionary*) of candidate endmembers, and then recovers those contributing to the observed mixed spectra.

Previous attempts at predicting the performance of SRU algorithms [3] have relied on theoretical results linking the likelihood of obtaining useful sparse representations of mixed signals to the low degree of coherence (i.e., the largest normalized correlation) between the columns of the dictionary matrix and the high degree of sparseness of the abundance vectors (e.g., [7, 8]). While coherence can be easily computed, it provides pessimistic guarantees in most cases due to the fact that such guarantees consider all sparse signals. This pessimism is particularly strong in hyperspectral applications, where the spectral signatures of the materials tend to be highly correlated. As a result, to the best of our knowledge, theoretical guarantees for the performance of SRU algorithms have not yet been established.

In this paper, we will establish theoretical guarantees on the performance of SRU algorithms by extending the re-

sults in [9]. The method is based on the evaluation of the *exact recovery condition* (ERC), which is used to provide performance guarantees of sparse regression algorithms (e.g., lasso [10]) that are specific to a given support set. In our application, each choice of support for the decomposition of the observed spectra corresponds to a specific mixture of endmembers being present. We will extend the performance guarantees in [9] to sparse regression with nonnegative constraints on mixed spectra obtained by nonlinear combinations of endmembers. While we find that the exact recovery condition (ERC) in the proposed theorems are highly demanding, we also easily find example cases in which the ERC is not met and SRU causes errors in endmember selection. Nonetheless, we find that the degree to which the ERC is met is indicative of the performance of SRU.

2. BACKGROUND: PERFORMANCE GUARANTEE FOR LASSO

The lasso is an unconstrained optimization algorithm that regularizes a least-squares fit penalty with a sparsity-inducing cost on the abundance vector. When applied to spectral unmixing, lasso searches for the endmembers that best describe the input signal as a linear combination in a least square sense while reducing the number of contributing endmembers. The lasso is written as

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|_2^2 + \gamma\|\boldsymbol{x}\|_1, \qquad (1)$$

where $\boldsymbol{y} \in \mathbb{R}^p$ is an input signal, $\mathbf{A} \in \mathbb{R}^{p \times N}$ is a matrix whose columns corresponds possible endmembers in the dictionary, and $\boldsymbol{x} \in \mathbb{R}^N$ is an abundance vector. An analysis of the lasso was provided by Tropp [9] and hinges on the exact recovery coefficient, defined as

$$\text{ERC}(\Lambda) := 1 - \max_{n \notin \Lambda} \|\mathbf{A}_\Lambda^\dagger \mathbf{A}_n\|_1, \qquad (2)$$

where $\Lambda \subseteq \{1, \ldots, N\}$ are indices for a subset of the columns of $\mathbf{A}$, $\mathbf{A}_\Lambda$ denotes the submatrix of $\mathbf{A}$ containing those columns, $\mathbf{M}^\dagger$ denotes the pseudoinverse of the matrix $\mathbf{M}$, and $\mathbf{A}_n$ denotes the nth column of $\mathbf{A}$. Note that it is implicitly assumed that the columns of $\mathbf{A}_\Lambda$ are linearly independent so that the pseudoinverse exists. When the columns of $\mathbf{A}$ have unit ℓ_2 norm, the condition considers the minimum angle between endmembers outside of Γ and the subspace spanned by

$\mathbf{A}_\Gamma$. Intuitively, a larger ERC is preferred because it reduces correlation between $\mathbf{A}_\Gamma$ and endmembers outside the set. The following theorem provides performance guarantees for the lasso that are specific to a particular support Λ.

Theorem 1 *[9, Theorem 8] Let Λ index a linearly independent collection of columns of $\mathbf{A}$ for which $\mathrm{ERC}(\Lambda) \geq 0$. Suppose that $\boldsymbol{y}$ is an input signal whose ℓ_2 best approximation $\boldsymbol{a}_\Lambda = \mathbf{A}_\Lambda \mathbf{A}_\Lambda^\dagger \boldsymbol{y}$ over $\mathbf{A}_\Lambda$ satisfies the correlation condition*

$$\|\mathbf{A}^T(\boldsymbol{y} - \boldsymbol{a}_\Lambda)\|_\infty \leq \gamma \mathrm{ERC}(\Lambda). \tag{3}$$

Let $\boldsymbol{x}^*$ be the solution of the lasso with parameter γ, We may conclude the following.

- *The support of $\boldsymbol{x}^*$, denoted $\mathrm{supp}(\boldsymbol{x}^*)$, is contained in Λ;*

- *the distance between $\boldsymbol{x}^*$ and the optimal coefficient vector $\boldsymbol{c}_\Lambda = \mathbf{A}_\Lambda^\dagger \boldsymbol{y}$ (appropriately zero-padded) satisfies*

$$\|\boldsymbol{x}^* - \boldsymbol{c}_\Lambda\|_\infty \leq \gamma \|(\mathbf{A}_\Lambda^T \mathbf{A}_\Lambda)^{-1}\|_{\infty,\infty}; \tag{4}$$

- *and $\mathrm{supp}(\boldsymbol{x}^*)$ contains the indices $\lambda \in \Lambda$ for which*

$$|\boldsymbol{c}_\Lambda(\lambda)| > \gamma \|(\mathbf{A}_\Lambda^T \mathbf{A}_\Lambda)^{-1}\|_{\infty,\infty}. \tag{5}$$

In words, the theorem states that if the approximation error of the input over the group Λ of columns of $\mathbf{A}$ is sufficiently uncorrelated with all other columns of $\mathbf{A}$, then the solution of the lasso does not pick any columns outside Λ, while picking columns corresponding to all sufficiently large entries of the approximation coefficients for $\boldsymbol{y}$ in $\mathbf{A}_\Lambda$. In terms of unmixing, this means that the algorithm commits no false alarms while successfully identifying all endmembers with *sufficiently strong* abundances.

3. PERFORMANCE GUARANTEE FOR SPARSE UNMIXING

In the hyperspectral remote sensing scenario, we use a constrained version of Lasso (CLasso) since the abundances to be estimated are constrained to be non-negative. The optimization problem is defined by

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|_2^2 + \gamma\|\boldsymbol{x}\|_1 \\ \text{subject to} \quad & \boldsymbol{x} \succeq \mathbf{0} \end{aligned} \tag{6}$$

Note that when the solution to problem (1) is nonnegative then it is also the solution to problem (6). We exploit this equivalence and the guarantee for the Lasso to provide the following guarantee for CLasso, proven in [11].

Theorem 2 *Let $\boldsymbol{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{e}$ denote the input to CLasso, where the abundance vector $\boldsymbol{\theta} \succeq \mathbf{0}$ ($\succeq$ denoting entry-wise inequality), $\Gamma = \mathrm{supp}(\boldsymbol{\theta})$ indexes a linearly independent collection of columns of $\mathbf{A}$, and $\boldsymbol{e}$ represents a departure from linearity (e.g., the effect of noise or nonlinear mixing). Let $\boldsymbol{x}^*$*

be the solution of CLasso with parameter γ. If $\mathrm{ERC}(\Gamma) \geq 0$, the vector $\boldsymbol{e}$ obeys

$$\|\mathbf{A}^T \mathbf{P}_{\mathbf{A}_\Gamma^\perp} \boldsymbol{e}\|_\infty \leq \gamma \mathrm{ERC}(\Gamma), \tag{7}$$

where $\mathbf{P}_{\mathbf{A}_\Gamma^\perp}$ is the projector onto the orthogonal complement of the span of $\mathbf{A}_\Gamma$, and

$$\boldsymbol{\theta}_\Gamma \succeq \gamma \|(\mathbf{A}_\Gamma^T \mathbf{A}_\Gamma)^{-1}\|_{\infty,\infty} - \mathbf{A}_\Gamma^\dagger \boldsymbol{e}, \tag{8}$$

then we have that $\mathrm{supp}(\boldsymbol{x}^*) = \mathrm{supp}(\boldsymbol{\theta}) = \Gamma$.

The theorem can be interpreted as follows. Under the assumption of the ERC condition being met, it is possible to find a value of $\gamma > 0$, dependent on the disturbance, for which CLasso will successfully identify all of the endmembers in the mixture that are sufficiently large.

The correlation condition (7) provides an insight into the types of errors that can violate the assumptions required for the exact recovery of endmembers. From the right hand side in (7), the endmembers in $\boldsymbol{A}_\Gamma$ are correctly recovered as long as the projection of the error on the orthogonal complement of $\mathrm{span}(\boldsymbol{A}_\Gamma)$ is small enough. Even if the error has a significant components orthogonal to this subspace, the assumption still holds as long as no other endmember in $\boldsymbol{A}$ is sufficiently correlated to the error projection.

Noting that no assumption is made on the distribution of errors in Theorem 2, a signal $\boldsymbol{y} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{e}$ obtained by non-linearly mixing the endmembers can be modeled via a departure $\boldsymbol{e}$ from a linear combination of the endmembers $\mathbf{A}\boldsymbol{\theta}$. If such deviation is small enough to meet the conditions (7) and (8), the endmembers are correctly recovered by CLasso, as we will demonstrate in the next section.

4. EXPERIMENTAL RESULTS

In this section, we will verify Theorem 2 and study the general performance of CLasso. To solve the CLasso problem, the sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) [6] is employed.

In our experiments we are not concerned with the estimation of abundances. The reason is that we will use a linear model to reconstruct non-linearly mixed spectra. As a result, the estimated abundances will be unequivocally biased. We focus instead on the detection of endmembers and we evaluate algorithm performance using common metrics used in detection problems. In particular, we use recall, defined by $\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$ and false alarm rate (FAR), defined by $\mathrm{FAR} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TN})$, where TP, FP, TN, and FN denote true/false positive and true/false negative counts, respectively.

4.1. Spectrum Database: Base Library

We constructed a base dataset using RELAB database [12]. We choose 360 spectra from 14 mineral classes. Each spectrum was acquired in the VNIR-SWIR range (300-2600 nm) with 5 nm spectral resolution. In order to obtain the same
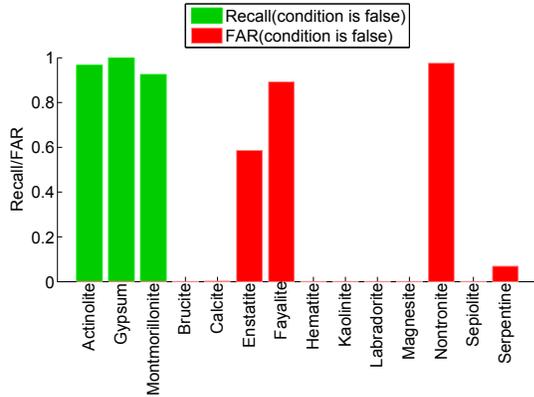
Fig. 1. Performance on the dataset with 14 endmembers from different classes.

Fig. 2. Performance on a dataset with positive ERC.

number of spectra in each class, we mixed the elements of each class according to a nonlinear mixing model introduced by Hapke [13]. The resulting dataset contains 43 spectra in each class and 14 classes.

In order to test the performance of CLasso, we select a spectrum from the classes actinolite, gypsum, and montmorillonite to form the submatrix $\mathbf{A}_\Gamma$ that generates the observed mixtures. In the following three sections, we construct the complement dictionary $\mathbf{A}_{\Gamma^c}$ in three different ways to illustrate our results. In all experiments, we construct test signals $\boldsymbol{y}$ using the Hapke model. The abundances for the three endmembers used in the mixtures are sampled uniformly from the probability simplex. In total, we generate 1000 test mixtures.

4.2. Experiment 1: Dataset with one endmember per class

In the first experiment, we construct the $\mathbf{A}_{\Gamma^c}$ library matrix by selecting one spectrum from each of the 11 classes not included in $\mathbf{A}_\Gamma$. The dictionary $\mathbf{A}$ constructed in this example exhibits ERC $= -0.6$ for the support set Γ, which means that no test spectrum $\boldsymbol{y}$ meets the assumption of the theorem because the left hand side of the inequality (7) requires ERC to be positive.

SUnSAL is applied to this data set with its trade-off parameter set at $\gamma = 0.001$ for best performance. Figure 1 shows the performance on this data set by showing in green the recall and in red the FAR over the 1000 test spectra. The first three selected endmember spectra (from the left) corresponds to the correct endmembers in $\mathbf{A}_\Gamma$. Although ERC < 0, almost all true endmembers are correctly detected even though there are some false alarms. In addition, the false alarms appear to occur with endmembers that show higher correlation with the spectra in $\mathbf{A}_\Gamma$ as is the case of enstatite and fayalite (similar to actinolite) and nontronite (similar to montmorillonite).

4.3. Experiment 2: Dataset with positive ERC

In this experiment, we construct the dictionary matrix and the synthetic mixtures so that they satisfy the sufficient condition
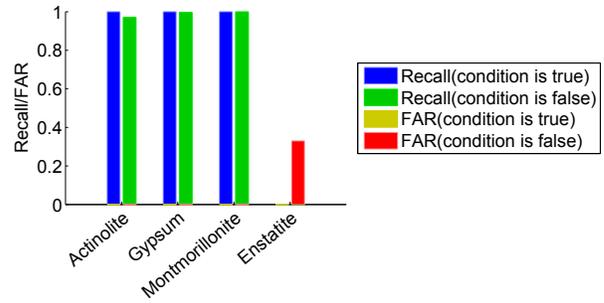
required by Theorem 2. We first consider the library $\mathbf{A}$ for which ERC becomes positive. Unfortunately, only spectra from the enstatite class can be added to $\mathbf{A}_{\Gamma^c}$ to obtain reasonable ERC, as the spectra in the base library are highly correlated to each other. We include one enstatite spectrum in $\mathbf{A}_{\Gamma^c}$ and consider the same three endmembers from the actinolite, gypsum, and montmorillonite classes in $\mathbf{A}_\Gamma$, resulting in four endmembers in the library $\mathbf{A}$. First, we assess whether each tested spectrum fulfills the assumptions in Theorem 2. The ERC for this dataset is 0.06 and 310 points (out of 1000) satisfy the conditions (7) and (8). We apply SUnSAL with $\gamma = 0.001$, and the performance is evaluated in the same way as the last experiment and reported in Figure 2. All the spectra that fulfill the sufficient condition of Theorem 2 are correctly retrieved without any false alarm, while for the other points the recall is 99% and the FAR is 1%. This result validates the sufficiency of Theorem 2 in guaranteeing perfect detections. Furthermore, it suggests a positive correlation between ERC and detection performance. Even if for some spectra the (strict) conditions in the theorem are violated, a positive ERC does result in good performance on them as well.

4.4. Experiment 3: Dataset with multiple endmembers per class

In this experiment, we augment the library considered in Section 4.2 by including in $\mathbf{A}_{\Gamma^c}$ one additional spectrum from each of the classes present in $\mathbf{A}_\Gamma$. 17 members are contained in the new library. Obviously, the first condition (7) is violated because the addition of new spectra with respect to the library from Experiment 1 cannot increase the already negative ERC. We discuss the performance of the CLasso on the incremented dictionary when SUnSAL is applied with $\gamma = 0.001$. Figure 3 illustrates the performance of the CLasso. The three endmembers in the fourth to sixth columns from the left are the augmented spectra. It is easy to see that the these three additional endmembers are falsely detected with high probability. However, the trend of the false detection for the other classes is similar to that in the first experiment shown in Figure 1. These results indicate that if we include spectra from the same classes used to generate the test mixtures, the CLasso tends to preferentially identify them as present in the mixtures as opposed to members of the other classes.
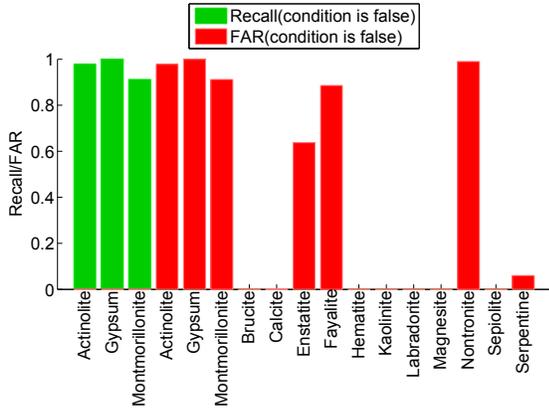
Fig. 3. Performance on the augmented dataset with three classes with multiple endmembers.
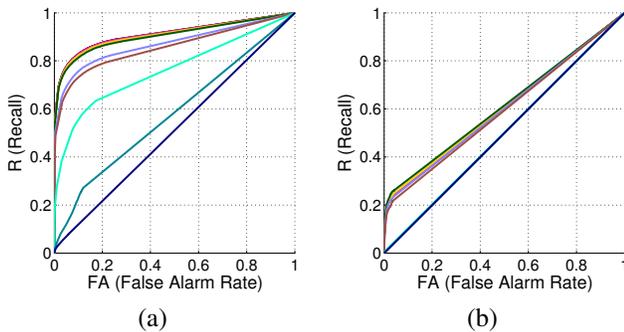

Fig. 4. Performance of the CLasso (a) class by class, and (b) element by element. Different curves are associated with different trade-off parameters γ over $[1.0 \times 10^{-5}, 100]$ at almost fixed intervals in log-scale.

4.5. Experiment 4: Entire Base Library

The observation from Experiment 3 is important because it can be used to explain the performance of CLasso in more general examples with multiple endmembers per class. In particular, we conducted an experiment using the whole base dictionary where all possible ternary class combinations were considered to produce test mixtures. Fifty mixed spectra were generated for each set of three endmembers. Figure 4 shows the overall performance of the SUnSAL algorithm. The performance is measured for different values of γ corresponding to each curve. SUnSAL exhibits high performance when the detection metrics are considered on a class-by-class basis, while the performance is noticeably poorer when it is measured on individual elements. This result reinforces the argument that most of the false alarms are committed with members of the same classes used to generate the test mixtures.

5. CONCLUSION

In this work, we presented a theoretical performance guarantee for CLasso. The performance theorem is verified mathematically and experimentally on mixtures from a fixed set of endmembers. As shown in the experiments, ERC plays

an important role to infer if we can correctly detect endmembers. Future work includes the development of the more relaxed condition for exact recovery of the endmembers for more practical situations, and theoretical performance when multiple endmembers are included in the library.

6. REFERENCES

[1] J. M. 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 Observations Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] R. Heylen, D. Burazerovic, and P. Scheunders, "Non-linear spectral unmixing by geodesic simplex volume maximization," *IEEE J. Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 534–542, Mar. 2011.

[3] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse Unmixing of Hyperspectral Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, June 2011.

[4] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total Variation Spatial Regularization for Sparse Hyperspectral Unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[5] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative Sparse Regression for Hyperspectral Unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.

[6] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, June 2010, pp. 1–4.

[7] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 27, pp. 9446–9451, July 2005.

[8] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the Uniqueness of Nonnegative Sparse Solutions to Underdetermined Systems of Equations," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4813–4820, Nov. 2008.

[9] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[11] Y. Itoh, M. F. Duarte, and M. Parente, "Performance guarantees for sparse regression-based unmixing," 2015, Available at `http://www.ecs.umass.edu/~mduarte/images/CLassoTR.pdf`.

[12] NASA Reflectance Experiment Laboratory, "Relab spectral database," available at `http://www.planetary.brown.edu/relabdocs/relab.htm`, 2014.

[13] B. Hapke, *Theory of Reflectance and Emittance Spectroscopy*, Cambridge University Press, Jan. 2012.