Performance Guarantees for Sparse Regression-Based Unmixing

Yuki Itoh
Marco Duarte, University of Massachusetts - Amherst
Mario Parente, University of Massachusetts - Amherst
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 results 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

$$\min_{\mathbf{x}} \frac{1}{2} \| \mathbf{y} - \mathbf{A} \mathbf{x} \|_2^2 + \gamma \| \mathbf{x} \|_1,$$  (1)

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

$$\text{ERC}(A) := 1 - \max_{n \notin \Lambda} \| \mathbf{A}^\dagger \mathbf{A}_n \|_1,$$  (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 \( n^{th} \) 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 \( \ell_2 \) norm, the condition considers the minimum angle between endmembers outside of \( \Gamma \) 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 A for which ERC(Λ) ≥ 0. Suppose that y is an input signal whose ℓ2 best approximation \(a_Λ = A_Λ A_Λ^T y\) over \(A_Λ\) satisfies the correlation condition

\[
\|A^T(y - a_Λ)\|_∞ ≤ γ_{ERC}(Λ). \tag{3}
\]

Let \(x^*\) be the solution of the lasso with parameter γ. We may conclude the following.

- The support of \(x^*\), denoted \(\text{supp}(x^*)\), is contained in \(Λ\);
- the distance between \(x^*\) and the optimal coefficient vector \(e_Λ = A_Λ^T y\) (appropriately zero-padded) satisfies

\[
\|x^* - e_Λ\|_∞ ≤ γ\|A_Λ^T A_Λ\|^{-1}\|\|∞,∞; \tag{4}
\]
- and \(\text{supp}(x^*)\) contains the indices \(\lambda \in Λ\) for which

\[
|c_Λ(\lambda)| > γ\|A_Λ^T A_Λ\|^{-1}\|\|∞,∞. \tag{5}
\]

In words, the theorem states that if the approximation error of the input over the group \(Λ\) of columns of \(A\) is sufficiently uncorrelated with all other columns of \(A\), 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 \(y\) in \(A_Λ\). 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

\[
\text{minimize } x \quad \frac{1}{2} \|y - Ax\|_2^2 + γ \|x\|_1 \tag{6}
\]

subject to \(x ≥ 0\).

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 \(y = Aθ + e\) denote the input to CLasso, where the abundance vector \(θ ≥ 0\) (≥ denoting entry-wise inequality), \(Γ = \text{supp}(θ)\) indexes a linearly independent collection of columns of \(A\), and \(e\) represents a departure from linearity (e.g., the effect of noise or nonlinear mixing). Let \(x^*\) be the solution of CLasso with parameter γ. If ERC(Γ) ≥ 0, the vector \(e\) obeys

\[
\|A^T P_{A_Γ^⊥} e\|_∞ ≤ γ_{ERC}(Γ), \tag{7}
\]

where \(P_{A_Γ^⊥}\) is the projector onto the orthogonal complement of the span of \(A_Γ\), and

\[
θ_Γ ≥ γ\| (A_Γ^T A_Γ)^{-1}\|_{∞,∞} - A_Γ^T e, \tag{8}
\]

then we have that \(\text{supp}(x^*) = \text{supp}(θ) = Γ\).

The theorem can be interpreted as follows. Under the assumption of the ERC condition being met, it is possible to find a value of γ > 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. In the right hand side in (7), the endmembers in \(A_Γ\) are correctly recovered as long as the projection of the error on the orthogonal complement of span(\(A_Γ\)) is small enough. Even if the error has a significant component orthogonal to this subspace, the assumption still holds as long as no other endmember in \(A\) is sufficiently correlated to the error projection.

Noting that no assumption is made on the distribution of errors in Theorem 2, a signal \(y = Aθ + e\) obtained by nonlinearly mixing the endmembers can be modeled via a departure \(e\) from a linear combination of the endmembers \(Aθ\). 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 Recall = TP/(TP + FN) and false alarm rate (FAR), defined by FAR = FP/(FP + 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 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.
Recall/FAR

In order to test the performance of CLasso, we select a spectrum from the classes actinolite, gypsum, and montmorillonite to form the submatrix \( A_\Gamma \) that generates the observed mixtures. In the following three sections, we construct the complement dictionary \( A^c \) in three different ways to illustrate our results. In all experiments, we construct test signals \( y \) 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 \( A_{\Gamma^c} \) library matrix by selecting one spectrum from each of the 11 classes not included in \( A_\Gamma \). The spectral signatures in this library are shown in [11]. The dictionary \( A \) constructed in this example exhibits \( ERC = -0.6 \) for the support set \( \Gamma \), which means that no test spectrum \( y \) meets the assumption of the theorem because the left hand side of the inequality (7) requires \( ERC \) to be positive.

SUxSAL 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) correspond to the correct endmembers in \( A_\Gamma \). 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 \( A_\Gamma \) 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 \( A \) for which \( ERC \) becomes positive. Unfortunately, only spectra from the enstatite class can be added to \( A_{\Gamma^c} \) to obtain reasonable \( ERC \), as the spectra in the base library are highly correlated to each other. We include one enstatite spectrum in \( A_{\Gamma^c} \) and consider the same three endmembers from the actinolite, gypsum, and montmorillonite classes in \( A_\Gamma \), resulting in four endmembers in the library \( A \), which are shown in [11].

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 \( A_{\Gamma^c} \) one additional spectrum from each of the classes present in \( A_\Gamma \). In total, 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. The spectral signatures in this modified augmented library are shown in [11]. 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
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


