# UNMIXING SPARSE HYPERSPECTRAL MIXTURES

*Marian-Daniel Iordache[1], José Bioucas-Dias[1], António Plaza[2]*

[1]Instituto de Telecomunicações, Instituto Superior Técnico, TULisbon,
Av. Rovisco Pais, 1049-001, Lisbon, Portugal
[2]Department of Technology of Computers and Communications, University of Extremadura,
Avda. De la Universidad S/N, 10.071, Cáceres, Spain

**ABSTRACT**

Finding an accurate sparse approximation of a spectral vector described by a linear model, when there is available a library of possible constituent signals (called *endmembers* or *atoms*), is a hard combinatorial problem which, as in other areas, has been increasingly addressed. This paper studies the efficiency of the sparse regression techniques in the spectral unmixing problem by conducting a comparison between four different approaches: Moore-Penrose Pseudoinverse, Orthogonal Matching Pursuit (*OMP*) [1], Iterative Spectral Mixture Analysis (*ISMA*) [2] and $l_2 - l_1$ sparse regression techniques, which are of widespread use in compressed sensing. We conclude that the $l_2 - l_1$ sparse regression techniques, implemented here by Iterative Shrinkage/Thresholding (*TwIST*) algorithm [3], yield the state-of-the-art in the hyperspectral unmixing area.

***Index terms*** *– sparse regression, hyperspectral unmixing,* $l_2 - l_1$ *norm minimization, convex optimization*

## 1. INTRODUCTION – HYPERSPECTRAL MIXING/UNMIXING PROCESSES

The development of the hyperspectral sensors opened new horizons in the exploitation of the sub-pixel details. These sensors are able to sample the electromagnetic spectrum in tens or hundreds of contiguous spectral bands (from the visible to the near-infrared region). Often, they have spatial resolutions of tens of meters. As a consequence, the efforts of the researchers moved forward from improving the spatial resolution toward the exploiting of the spectral resolution.

The relative low spatial resolution of the hyperspectral sensors implies the existence of several materials (called *endmembers*) inside the same pixel. These pixels, which contain more than one endmember, are called *mixed pixels*.

Any material is characterized by a specific spectrum (called *spectral signature*), which is obtained by measuring the reflectance of that material for a specific range of wavelengths of the incident light. The data acquired by a hyperspectral sensor is a data cube (two spatial coordinates and one coordinate corresponding to the wavelengths of the incident light) of high dimension, containing the values of the measured reflectance of all the pixels for specific spectral bands. The measurements are affected by errors, due, on one hand, to the electronic noise and, on the other hand, to the speed of the sensor flying at high altitudes. The *unmixing* process consists in finding what are the materials inside a mixed pixel, their respective spectra and what percentage of the area occupy each material inside the pixel (called *fractional abundance*).

There are two models used in the unmixing problem:

- the linear mixing model (LMM): the endmembers occupy distinct areas inside the pixel;
- the nonlinear mixing model (NLMM): the endmembers form an intimate mixture inside the pixel.

In our paper, we adopt the LMM, which is the most used in the unmixing problem, as it corresponds to a reasonable balance between accuracy and model complexity. The NLMM is far more complex and it is used in specific applications, being parameterized with scene parameters, quite often very expensive, or impossible, to obtain.

## 2. THE LINEAR MIXING MODEL (LMM)

The LMM assumes that the endmembers are surface distributed occupying distinct areas inside each pixel (e.g., checkerboard-type scenes). The acquired spectrum from a given pixel is a linear combination of the spectra of materials present in the pixel. For each pixel, it can be expressed as follows:

$$r_i = \sum_{j=1}^{q} m_{i,j}\alpha_j + n_i \qquad (1)$$

where $r_i$ is the measured value of the reflectance at band *i*, *q* is the total number of endmembers present in the pixel, $m_{i,j}$ is the reflectance of the $j^{th}$ endmember at band *i*, $\alpha_j$ is the fractional abundance of the $j^{th}$ endmember, and $n_i$ represents the error term for the spectral band *i* (e.g., the noise affecting the measurement process).

The spectral signatures of the endmembers can be collected in an *L*-by-*q* matrix (*L* being the number of spectral bands) called *the mixing matrix.* The expression (1) can be written in a compact form:

$$R = M \cdot \alpha + n \qquad (2)$$

where *R* is and *L*-by-*1* column vector (the *measured spectrum* of the pixel), $\alpha$ is a *q*-by-*1* vector containing the respective fractional abundances of the endmembers and *n* is an *L*-by-*1* vector collecting the errors affecting the measurements at each spectral band.

The fractional abundances of the endmembers sum to one and can not be negative. These constraints are known as *the sum-to-one* and *the non-negativity* constraints:

$$\sum_{j=1}^{q} \alpha_j = 1 \ \ (\textit{sum-to-one}) \tag{3}$$

$$\alpha_j \geq 0 \ , \ \forall j \ (\textit{non-negativity}). \tag{4}$$

We use the LMM in the unmixing problem under a semi-supervised approach, which consists in searching for the spectral signatures of the endmember in a large dictionary called *spectral library*, which contains $p$ (usually $p << L$) members and will be denoted by $S$. At this point, the system (2) can be written as follows:

$$R = S \cdot f + n \tag{5}$$

where $f$ is an $L$-by-$1$ vector containing the fractional abundances of the members contained in $S$.

The goal of the unmixing process is to find the mixing matrix $M$ and the vector of fractional abundances $f$ , given $R$ and $S$.

As the number of actual endmembers $q$ is much smaller than the number $p$ of spectra contained in $S$, the vector of fractional abundances $f$ is **sparse**. This is a combinatorial problem which calls for efficient sparse regression techniques.

## 3. UNMIXING ALGORITHMS

This section describes the four unmixing algorithms tested in the paper.

### 3.1. Moore-Penrose Pseudoinverse (MPP)

As $S$ is not a square matrix (so it is not invertible), the unmixing problem is ill-posed and we can not find an estimate $\hat{f}$ of $f$ by multiplying the inverse of $S$ with $R$: $\hat{f} = S^{-1}R$. Given the characteristics of $S$ ( $L >> p$ and linearly independent columns), the product $S^T S$ is square with full rank, thus it is invertible. An estimate of $f$ can be found as follows:

$$\hat{f} = \left(S^T \cdot S\right)^{-1} \cdot S^T \cdot R = S^{\#} \cdot R \tag{6}$$

where $S^{\#} = \left(S^T \cdot S\right)^{-1} \cdot S^T$ is *the Moore-Penrose pseudoinverse* of the matrix $S$. This solution solves the minimization

$$\hat{f} = \arg\min_{f} \left\| R - S \cdot f \right\|^2 \tag{7}$$

and $\hat{f} = f$ only if the measurements are not affected by noise. Otherwise, the estimate includes an error term, which strongly depends on the condition number of $S$, becoming more important when $S$ is bad-conditioned (the condition number of $S$ is big). Moreover, the solution obtained by the MPP is unconstrained and it is plausible that it will not satisfy the sum-to-one and the non-negativity constraints.

### 3.2. Orthogonal Matching Pursuit (OMP)

OMP was introduced in 1993 in [1] as an alternative to Matching Pursuit [4] and it is an iterative technique which searches, at each iteration, the spectral signature from $S$ which best explains a predetermined residual.

At the first iteration, the initial residual is equal to the observed spectrum of the pixel, the vector of fractional abundances is null and the matrix of the indices of selected endmembers is empty. Then, at each iteration, the algorithm finds the member of $S$ which is best correlated to the actual residual, adds this member to the endmembers matrix, updates the residual and computes the estimate of $f$ using the selected endmembers. The algorithm stops when a stop criterion is satisfied. A member from $S$ can not be selected more than once, as the residual is orthogonalized to the members already selected.

### 3.3. Iterative Spectral Mixture Analysis (ISMA)

ISMA is an iterative technique derived from Spectral Mixture Analysis [5]. It finds the optimal endmember set by examining the change in the root-mean-squared (RMS) error along the iterations.

ISMA consists in two parts. In the first one, ISMA computes, initially, an unconstrained solution of the unmixing problem, using all the members of $S$. Then, removes the member with the lowest abundance fraction and repeats the process with the remaining endmembers, until one endmember remains. The second part of ISMA consists in finding *the critical iteration*, which is the iteration corresponding to the first abrupt change in the RMS error, computed as follows:

$$\Delta RMS = 1 - \left( RMS_{j-1} / RMS_j \right) \tag{8}$$

$RMS_j$ is the RMS error corresponding to the $j^{th}$ iteration and it is computed as

$$RMS = \left( \frac{1}{q} \cdot \sum_{b=1}^{q} \left( r_b' - r_b \right)^2 \right)^{1/2} \tag{9}$$

where $r_b'$ and $r_b$ are the modeled and image spectrum values at band $b$.

The critical iteration corresponds to the optimal set of endmembers. The idea of recovering the true endmember set by analyzing the change in the RMS error is based on the fact that, before finding the optimal set of endmembers, the RMS error varies in certain (small) limits and it has a bigger variation when one endmember from the optimal set is removed, as the remaining endmembers are not enough to model with good accuracy the actual observation. ISMA computes, at each iteration, an unconstrained solution instead of a constrained one, as it is predictable that, when the endmember set approaches the optimal one, the abundance fractions will approach the true ones.

### 3.4. Two-Step Iterative Shrinkage/Thresholding (TwIST) algorithm

As shown in section 3.1, the estimate $\hat{f}$ of $f$ obtained by MPP contains an error term, when the observation is affected by noise:

$$\hat{f} \Rightarrow S^{\#} R = \hat{f} + S^{\#} \cdot n \tag{10}$$

If $S$ is bad-conditioned, the error term $S^{\#} n$ becomes more significant, as the small eigenvalues in the singular value decomposition (SVD) of $S$ lead to the amplification of noise and also to loss of sparseness. At this point, enforcing sparseness becomes not only a desirable approach, but a necessary one.

Many approaches to linear inverse problems (as the spectral unmixing problem is) define a solution as a minimizer of a convex objective function:

$$g(x) = \frac{1}{2} \left\| y - K \cdot x \right\|^2 + \lambda \cdot \Phi(x) \tag{11}$$

where $y$ is the observed data, $K$ is the linear operator, $\lambda$ is *the regularization parameter* and $\Phi(x)$ is the *regularizer*. By minimizing the objective function $g$, we are able to find the best compromise between the lack of fitness of a candidate $x$ to the observed data (which is given by $\|y - K \cdot x\|^2$) and its degree of undesirability (given by $\Phi(x)$). The relative weight of the two terms is controlled by the regularization parameter $\lambda$.

TwIST is an iterative algorithm with general applicability in inverse problems which solves the optimization problem given in (11), combining the advantages of iterative shrinkage/thresholding (IST) [6] and iterative reweighted shrinkage (IRS) [7] algorithms (good accuracy for not very ill-conditioned operators and very fast convergence for ill-conditioned operators, respectively). At each iteration, TwIST computes a solution that depends on the two previous estimates (this being the reason underlying the name *Two-Step*).

In the spectral unmixing problem described by (11), the regularizer $\Phi$ is the $l_0$ norm of $f$ (denoted by $\|f\|_0$), which gives the number of non-zero elements (in order to penalize non-sparse solutions), and the objective function becomes:

$$g(f) = \frac{1}{2}\|R - S \cdot f\|^2 + \lambda \cdot \|f\|_0 \qquad (12)$$

The objective function in (12) is a non-convex one, difficult to solve. It is, however, known [8] that, for matrices $S$ with certain properties of incoherence and sparse vectors $f$, the $l_0$ norm can be replaced by the $l_1$ norm $\|f\|_1$:

$$g(f) = \frac{1}{2}\|R - S \cdot f\|^2 + \lambda\|f\|_1 \qquad (13)$$

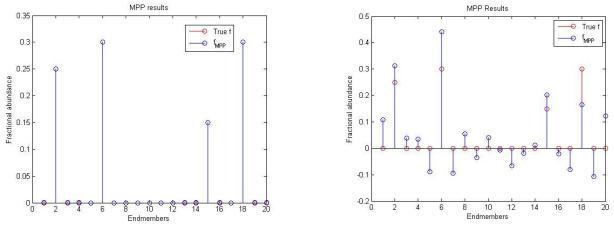
where $\|f\|_1 \equiv \sum_{i=1}^{p} |f_i|$. TwIST solves the unmixing problem by minimizing the $l_2 - l_1$ objective function given in (13), which has the advantage of being convex, by opposite to the one given in (12).

## 4. TEST DATA

The four unmixing algorithms were tested for different configurations of the unmixing problem. We considered three different spectral libraries differentiated by their condition number: $S_1$ having the condition number $C_1 \approx 10$ (*well-conditioned*), $S_2$ with the condition number $C_2 \approx 90$ (*medium-conditioned*) and $C_3$ with its respective condition number $C_3 \approx 2.45 \cdot 10^3$ (*bad-conditioned*). Each spectral library contains $p = 20$ simulated spectral signatures, defined for $L = 220$ spectral bands. The observations are affected by zero-mean Gaussian noise and signal-to-noise ratios ($SNR \equiv \|S \cdot f\|^2 / \|n\|^2$) of *20*, *40*, *60* and *80 dB*. We generated observations containing 4, 5 or 6 endmembers. The algorithms were tested for all possible combinations of these simulation conditions. We adopted a Monte Carlo methodology – the results presented in section 5 are medium results obtained for 100 simulations of the respective configurations.

## 5. RESULTS

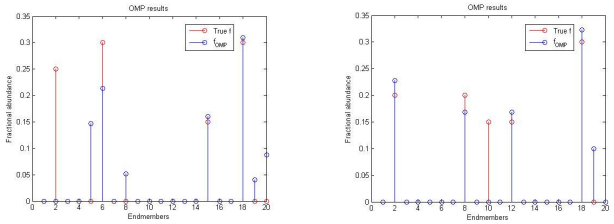
The tests showed poor results returned by MPP most of the time. Even for medium-conditioned $S$ and not too low SNR ($SNR = 40dB$), MPP returns dissatisfactory results. MPP can be successfully used only in good conditions (well-conditioned $S$ and high SNR), as shown in Fig. 1.a. In the other cases, the results are physically unrealistic, which means that they do not respect the non-negativity and the sum-to-one constraints (Fig. 1.b).




a) $q = 4$ ; $S = S_1$ ; $SNR = 60dB$ b) $q = 4$ ; $S = S_2$ ; $SNR = 20dB$

Fig. 1. MPP results

OMP returns, generally, better results than MPP, but it encounters difficulties when the spectral library $S$ is bad ($S = S_3$) or medium-conditioned ($S = S_2$) and/or the SNR is low as shown in Fig. 2, for $q = 4$ (Fig. 2.a) and $q = 5$ (Fig. 2.b). For the OMP algorithm, the stop criteria used was "the residual falls below a preset threshold". If the threshold is too small, the method doesn't converge, as the residual never attends it.




a) $q = 4$ ; $S = S_3$ ; $SNR = 40dB$ b) $q = 5$ ; $S = S_2$ ; $SNR = 20dB$

Fig. 2. OMP results

ISMA is a much more powerful method than MPP and OMP. The critical iteration is found when the RMS variation is bigger than a preset threshold $t$. By selecting an appropriate threshold, ISMA finds the correct set of endmembers, as shown in Fig. 3, for $q = 5$, $S = S_2$ and $SNR = 20dB$. Although, the fractional abundances differ from the true ones, their accuracy being affected by the noise and by the condition number of $S$.

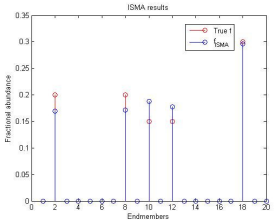
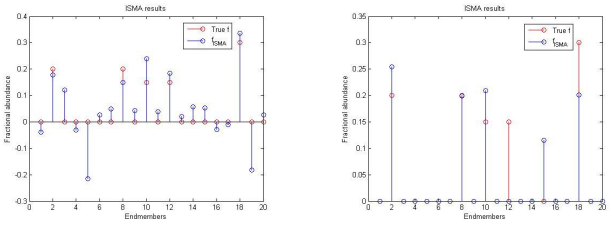



Fig. 3. ISMA results by choosing the optimal threshold $t$

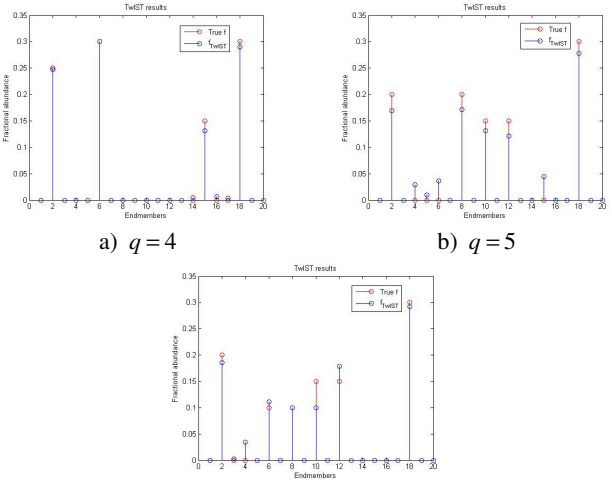
ISMA encounters difficulties when $S = S_2$ or $S = S_3$ and $SNR < 40dB$, due to the difficulty of choosing the appropriate threshold *t*. When the spectral library *S* becomes bad-conditioned and the *SNR* decreases, the change in the RMS error becomes smoother and the range in which *t* can be chosen becomes very tight. If the chosen threshold is smaller than the optimal one, the critical iteration is found after a smaller number of iterations and the number of endmembers found by ISMA can be much larger than the true one (Fig. 4.a). This behavior implies unrealistic fractional abundances, as ISMA computes, at each iteration, an unconstrained solution. On the other hand, if the imposed threshold is higher than the optimal one, the computational process continues to remove endmembers from the optimal set (the critical iteration is found later), which means that the number of endmembers is underestimated (Fig. 4.b). In Fig. 4, note the very small difference between the values of *t*.




a) $t = 2.8e-3$ b) $t = 4.3e-3$

Fig. 4. ISMA results for $q = 5$, $S = S_2$, $SNR = 20dB$

TwIST is a more robust method than ISMA and it returns better results than any of the previous methods, even in bad conditions ($S = S_3$ and $SNR = 20dB$), as shown in Fig. 5 for different numbers of endmembers present in the pixel. In our tests, the regularization parameter $\lambda$ was hand-tuned. An important advantage of TwIST is that $\lambda$ has a relatively wide range of variation which doesn't influence in a significant way the results, which means that it is easier to choose than the ISMA parameter *t*.




a) $q = 4$ b) $q = 5$

c) $q = 6$

Fig. 5. TwIST results in bad conditions

## 6. CONCLUSIONS

The tests showed the superiority of TwIST compared to the other methods. The good results of TwIST in the unmixing problem are in line with the success of the $l_2 - l_1$ minimization methods in signal processing and, particularly, in the Sparse-Land applications [8]. Applied to the unmixing problem, TwIST algorithm is able to infer the actual endmembers with a much higher accuracy than the other three methods tested. ISMA has difficulties in finding the correct set of endmembers when the spectral library is bad-conditioned and/or the SNR is low, due to the difficulty of finding the appropriate stop criteria. MPP has good results only when the spectral library is well-conditioned and the SNR is low, but this is a rare situation in practice, being almost an ideal situation. OMP has better results than MPP, but fails when the spectral library is medium or bad-conditioned or the SNR is low.

## 7. REFERENCES


[1] Y. C. Pati, R. Rezahfar, P.S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition", in *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems and Computers*, Los Alamitos, CA, USA, 2003.

[2] D. M. Rogge, B. Rivard, J. Zhang, J. Feng, "Iterative Spectral Unmixing for Optimizing Per-Pixel Endmember Sets", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, December 2006.

[3] J. M. Bioucas-Dias, M. A. T. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration", *IEEE Transactions on Image Processing*, vol. 16, Issue 12, 2007.

[4] S. Mallat, Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries", *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397-3415, 1993.

[5] J. B. Adams, M. O. Smith, A.R. Gillespie, "Imaging Spectroscopy: Interpretation Based on Spectral Mixture Analysis", *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, Cambridge Univ. Press, pp. 145-166, Cambridge, U.K., 1993.

[6] R. Nowak, M. A. T. Figueiredo, "Fast Wavelet-Based Image Deconvolution using the EM Algorithm", *Proc. 35th Asilomar Conf. Signals, Systems, Computers,* vol. 1, pp. 371-375, 2001.

[7] J. M. Bioucas-Dias, "Fast GEM Wavelet-Based Image Deconvolution Algorithm", *Proc. IEEE Int. Conf. Image Processing,* vol. 2, pp. 961-964, 2003.

[8] A. M. Bruckstein, D. L. Donoho, M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images", *SIAM Review*, vol. 51, no. 1, pp. 34-81, 2009



**Acknowledgements**

This work has been supported by the European Community's Marie Curie Research Training Networks Programme under contract MRTN-CT-2006-035927, "Hyperspectral Imaging Network (HYPER-I-NET)".