SPARSE HYPERSPECTRAL UNMIXING

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

[1]Instituto de Telecomunicações, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001, Lisbon, Portugal – (miordache, bioucas)@lx.it.pt
[2]Department of Technology of Computers and Communications, University of Extremadura, Avda. de la Universidad S/N, 10.071, Cáceres, Spain – aplaza@unex.es

KEY WORDS: Hyperspectral Unmixing, Sparse Regression, Convex Optimization

ABSTRACT:

Given a set of mixed spectral vectors, spectral mixture analysis (or spectral unmixing) aims at estimating the number of reference materials, also called endmembers, their spectral signatures, and their fractional abundances. A semi-supervised approach to deal with the linear spectral unmixing problem consists in assuming that the observed spectral vectors are linear combinations of a small number of spectral signatures known in advance. Unmixing then amounts to find a small number of materials in the spectral library that best represent the observed data. Finding a small number of signatures in a large library is a combinatorial problem which calls for efficient sparse regression techniques. In this study, we compare four unmixing algorithms with the ultimate goal of analyzing their potential in solving sparse hyperspectral unmixing problems. The algorithms compared are:
1) Moore-Penrose pseudoinverse;
2) Orthogonal Matching Pursuit algorithm (Y. Pati, *et al.* ,1993);
3) ISMA – Iterative Spectral Mixture Analysis (D. Rogge, *et al*., 2006);
4) TwIST (Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration) algorithm (J. Bioucas-Dias and M. Figueiredo, 2007).
After conducting a quantitative and comparative analysis of the above-mentioned algorithms, we conclude that the $l_2 - l_1$ sparse regression techniques and the respective algorithms, of which TwIST is an example, yield state-of-the-art performance in hyperspectral sparse unmixing; this conclusion is in line with the success that these optimization methods have achieved in the area of compressed sensing.

1. INTRODUCTION

The spectral images usually contain mixed pixels, as numerous disparate substances (called *endmembers*) can contribute to the measured spectra of a single pixel. The characteristics of a mixed pixel are determined by the low spatial resolution of the spectral sensor flying at high altitudes and by the combination of different materials in a (homogeneous) mixture inside the pixel (Keshava, 2003).

The mixing models used to solve the unmixing problem are either *linear (LMM - linear mixing model)* or *non-linear*. The non-linear models are far more complex, often depending on scene parameters difficult to obtain. For these reasons, the non-linear models are not of widespread use in hyperspectral applications.
In this paper, we adopt the LMM, which, despite its simplicity, models with reasonable approximation many real situations.

* Corresponding author

2. THE LINEAR MIXING MODEL

The LMM assumes that the spectral response of a pixel is a linear combination of the endmembers signatures present in that pixel, i.e.,

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

where r_i is the reflectance of the pixel measured by the satellite sensor at band *i*, $m_{i,j}$ is the reflectance of the j^{th} endmember at band *i*, α_j is the fractional abundance of the j^{th} endmember, n_i is the error term for spectral band *i* (*e.g.*, the noise affecting the measuring process), and *q* is the total number of endmembers present in the pixel. If we consider that the hyperspectral sensor used in data acquisition has *L* spectral bands, the system (1) can be written in matrix form as follows:

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

where *R* is an L-by-1 column vector (the *measured spectrum* of the pixel), *M* is an L-by-q matrix called *the mixing matrix* (the spectral signatures of the endmembers), α is a q-by-1 column-vector (the respective fractional abundances of the endmembers) and *n* is an L-by-1 column-vector collecting the errors affecting the measurements. The fractional abundances of the constituent endmembers are subject to two constraints: i) the non-negativity constraint: the fractional abundances can not be negative; ii) the sum-to-one constraint: the fractional abundances of the endmembers should sum to one.

In a semi-supervised approach, the unmixing problem is solved by searching for the endmembers in a given large database – a spectral library which will be denoted by the L-by-p matrix *S*, where *p* is the number of spectral signatures contained in *S* and $p>>q$; we assume that $L>>p$. The system (2) can be written:

$$R = S \cdot f + n \qquad (3)$$

As the number of endmembers contained in a pixel is usually small, the vector of fractional abundances *f* is **sparse** and the unmixing process requires solving a combinatorial problem, which demands to significant time and computational efforts. These obstacles call for efficient sparse regression techniques.

3. THE ALGORITHMS

This section describes the four algorithms tested in this paper.

3.1 Moore-Penrose Pseudoinverse

Given system (3), the first simple idea coming into mind, in order to find an estimate $\hat{f}$ of f, is to multiply the inverse of S by the observation vector R ($\hat{f} = S^{-1}R$). However, this is not possible, as the matrix S is not square, thus it is not invertible and we have to deal with an ill-posed problem. Based on the fact that L>>p and considering that the columns of *S* (the spectra) are linearly independent, the product $S^T S$ is square with full-rank (so, invertible) and it can be shown that the estimate $\hat{f}$ can be obtained as follows:

$$\hat{f} = \left(S^T \cdot S\right)^{-1} \cdot S^T \cdot R = S^{\#} \cdot R \qquad (4)$$

where $S^{\#} = \left(S^T \cdot S\right)^{-1} \cdot S^T$ is called *the Moore-Penrose pseudoinverse* of the matrix S. This is the unconstrained solution of system (3) in the least-squares sense:

$$\hat{f} = \arg\min_{f} \left\| R - S \cdot f \right\|^2 \qquad (5)$$

If the observation is not affected by noise, this solution is the exact one. When the observation is affected by noise, the estimate includes an error term. This error term depends on the condition number of the matrix $S^T S$, becoming more important when *S* is bad-conditioned. The details about how this inexact solution leads to a new approach are shown in section 3.4. However, as it is an unconstrained solution, it is plausible that it will not satisfy the non-negativity and the sum-to-one constraints.

3.2 Orthogonal Matching Pursuit (OMP)

OMP (Pati *et al.*, 1993) is an iterative technique derived from Matching Pursuit

(Mallat & Zhang, 1993). It searches for the spectral signature contained in the spectral library which best represents a predetermined residual. The algorithm starts with three initializations: i) the initial residual is equal to the observed spectrum of the pixel: $RES \leftarrow R$; ii) the vector of fractional abundances is null: $\hat{f} \leftarrow 0$; iii) the matrix of the indices of selected endmembers is empty: $\Lambda = \Phi$.

Further, at each step, the algorithm finds the member of the library which is best correlated to the actual residual, adds this member to the endmember matrix and memorizes in $\hat{f}$ the coefficients of the selected members. The residual is then updated and the entire process repeats until a stop criterion is satisfied. At any iteration, the residual is orthogonalised to the members which are already selected, so the next selected member is always a new one (a member can not be selected more than once).

3.3 Iterative Spectral Mixture Analysis (ISMA)

ISMA (Rogge *et al.*, 2006) was introduced as an alternative to Spectral Mixing Analysis (Adams *et al.,* 1993). As its name suggests, it is an iterative algorithm, which exploits the change in the root-mean-squared (RMS) error as a function of the selected endmembers. It is plausible that, when the set of selected endmembers approaches the actual one, the predicted fractional abundances are similar to the actual ones and they should sum to one. This is why ISMA computes, at each iteration, an unconstrained solution instead of a constrained one.

ISMA removes iteratively the members from the spectral library, until one member remains, and then finds the optimal endmember set by analyzing the change in the RMS as a function of the number of endmembers.

The member that is removed, at each iteration, is the one with the minimum fractional abundance. The process repeats with the remaining members.

The second part of ISMA consists in determining the critical iteration, which represents the "limit" between "too many" and "too few" endmembers and corresponds to the optimal endmember set. The critical iteration is found by examining the change in the RMS error, which, for the j^{th} iteration, is:

$$\Delta RMS = 1 - \left(RMS_{j-1} / RMS_j \right). \quad (6)$$

At some iteration, the endmember set found by ISMA is the optimal one and it is related to a certain RMS error (not very different from the previous ones). When an endmember from the optimal set is removed, the error increases dramatically, as the remaining endmembers are not enough to model with good accuracy the observed spectra; the critical iteration is, then, detected, when the changing in RMS is above a preset threshold. The set of endmembers remaining in the spectral library at the critical iteration is the actual mixing matrix.

3.4 Two-Step Iterative Shrinkage/Thresholding (TwIST) algorithm (Bioucas-Dias & Figueiredo, 2007)

TwIST is an iterative technique representing an improved version of *IST* (*Iterative Shrinkage/Thresholding*) class algorithms, which solve an unconstrained optimization problem arising from combining a linear observation model with a nonquadratic regularizer. The IST algorithms have a convergence rate which depends strongly on the linear observation operator. If this operator is ill-conditioned or ill-posed, the convergence speed of this class of algorithms is very slow. On the other hand, a different algorithm, proposed by Bioucas-Dias, under the generalized expectation-maximization framework, proved to be much faster than IST when the linear operator is strongly ill-conditioned. This algorithm is known as *IRS* (*Iterative Re-Weighted Shrinkage*) algorithm (Bioucas-Dias, 2003).

As we mentioned in the Section 3.1, the solution obtained by using the Moore-Penrose pseudoinverse, for a noisy observation, is affected by an error term:

$$\hat{f} \Rightarrow S^{\#}R = \hat{f} + S^{\#} \cdot n \qquad (7)$$

The error term $S^{\#}n$ is strongly influenced by the condition number of the matrix *S*. In terms of the SVD decomposition, the matrix *S* can be written as $S = U\Sigma_S V^H$ and its Moore-Penrose pseudoinverse as $S^{\#} = V\Sigma_S^{-1}U^H$.

This expression shows that the small eigenvalues in its SVD decomposition lead to a large error term in $\hat{f}$ (due to the amplification of noise) and also to loss of sparsity. This means that the ill-conditioned nature of *S* has a strong influence on the accuracy of the results.

In order to limit the influence of small eigenvalues, we can search for a sparser solution by introducing a sparsity enforcing term in the objective function (5). A possible solution is $\|f\|_0$, the so-called l_0 norm, which gives the number of non-zero elements of f. The new optimization problem is then

$$\hat{f} = \arg\min_f \|R - S \cdot f\|^2 + \lambda \|f\|_0 \ , \quad (8)$$

where $\lambda > 0$ is a regularization parameter controlling the relative weight between the data misfit term $\|R - S \cdot f\|^2$ and the regularizer.

The optimization (8) is a non-convex combinatorial problem, which is very difficult to solve. It is, however, known (Bruckstein *et al.,* 2009) 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. This is of great relevance because the new optimization problem is convex and thus manageable. Based on this result, we use TwIST to minimize efficiently the $l_2 - l_1$ objective function

$$g(f) = \|R - S \cdot f\|^2 + \lambda \|f\|_1 \qquad (9)$$

where $\|f\|_1 \equiv \sum_{i=1}^{p} |f_i|$ is the l_1 norm of f.

4. TEST DATA

As the results are strongly influenced by the condition number of the spectral library, this study considered three simulated spectral libraries, S_i , *i=1,2,3*, with corresponding condition numbers $C_1 = 10.84$ ("well-conditioned"), $C_2 = 91.7$ ("medium-conditioned" or "not too bad conditioned") $C_3 = 2.47 \cdot 10^3$ ("ill-conditioned"). Each of these spectral libraries contains *p = 20* materials. The number of spectral bands is *L = 220*. The mixtures contain 4, 5 or 6 endmembers. 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*. The algorithms were tested for all the possible combinations of these simulation conditions.

5. RESULTS

The tests showed that all the algorithms have good results in good conditions (well-conditioned mixing matrix and high SNR).

When the mixing matrix is bad- or medium-conditioned, the Moore-Penrose Pseudoinverse, being an unconstrained method, returns accurate solutions only if the SNR is high (*60dB* or *80dB*); otherwise, the results are physically unrealistic (the fractional abundances do not satisfy the non-negative constraint and the sum-to-one constraint).

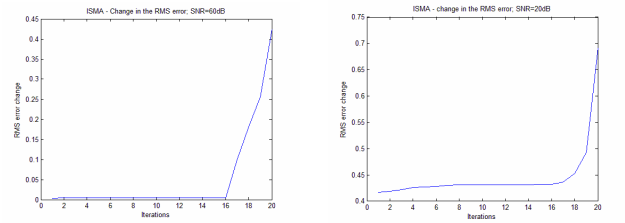
OMP returns, usually, better results than the Moore-Penrose Pseudoinverse. The results are accurate when the mixing matrix is well- or medium-conditioned and the noise level is low. In the case of medium-conditioned *S*, the OMP algorithm yields good results, even for $SNR = 40dB$, but, in hard conditions (bad-conditioned matrix and low SNR), it fails in inferring accurate solutions. In our tests, the stop criteria used was "the residual is below an imposed threshold". The method doesn't converge if this threshold is too low, as the residual never attends it. If the imposed residual is too high, the method underestimates the size of the set of endmembers.

The results obtained by ISMA are accurate most of the time. Although, the method has problems in finding the endmembers set when

the mixing matrix is medium- or bad-conditioned and the noise has a high level.

As it was described, ISMA finds the set of endmembers by examining the change in the RMS error along the iterations. The problems encountered by ISMA in hard conditions appear because of the high sensitivity of the method at the imposed threshold *t*.

Fig. 1 shows the evolution of the RMS error along the iterations for $S = S_2$, $q = 5$, SNR= $60db$ (Fig. 1.a) or SNR= $20dB$ (Fig. 1.b).

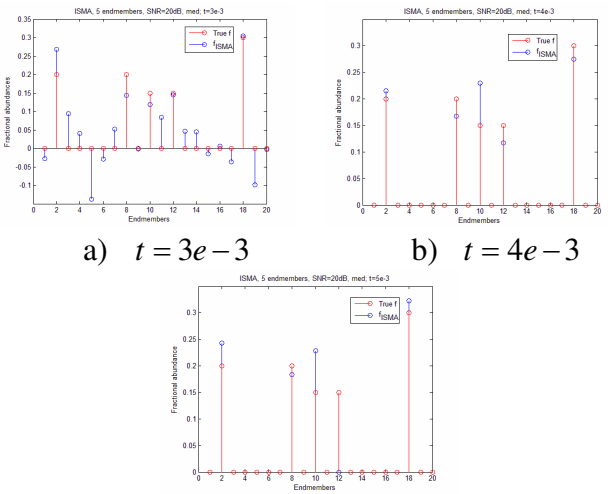


a) b)

Fig. 1 ISMA – The evolution of the RMS error for: a) $SNR = 60dB$; b) $SNR = 20dB$

From fig. 1, it is obvious that the stop criteria is very flexible in good conditions, but it is more difficult to fine tune in bad conditions.

Fig. 2 shows the results obtained by ISMA for different imposed thresholds. The configuration of the simulated system is the same as the one in Fig. 1.b).



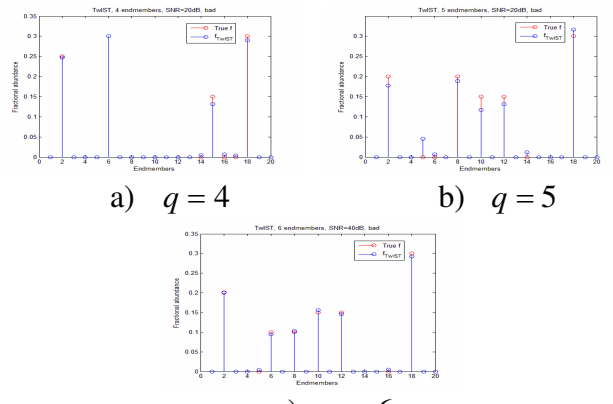
a) $t = 3e-3$ b) $t = 4e-3$

c) $t = 5e-3$

Fig. 2 ISMA - the reconstructed fractional abundances for different thresholds *t*.

In Fig. 2.a), the imposed threshold *t* is too small and the iterative process finds the "critical iteration" – the "boundary" between "too many" and "too few" endmembers – after only two steps. The algorithm returns an inaccurate solution. This solution contains both positive and negative abundance fractions, as the solution computed by ISMA at each step is an unconstrained one. For Fig. 2.b), the threshold is the optimal one and the inaccuracies of the solution are due to the high level of the noise. In Fig. 2.c), the imposed threshold is too high and the algorithm finds the critical iteration later than it should. Not all the endmembers are found and the endmember set becomes incomplete, because each additional iteration means the removal of one endmember. The major disadvantage of ISMA is that, in order to obtain an accurate solution, the imposed threshold has a very small range of values in which can be considered optimal. In harder conditions, it is even more difficult to find the optimal threshold.

TwIST proves to be more powerful than any of the previous methods, having the advantage of finding more accurate results even in bad conditions (bad-conditioned mixing matrix and high noise). Fig. 3 shows the results obtained by TwIST in bad conditions ($S = S_3$, $SNR = 20dB$ in Fig. 3.a and 3.b and $SNR = 40dB$ in Fig. 3.c) and different dimensions of the actual set of endmembers (4, 5 and 6 endmembers, respectively).



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

c) $q = 6$

Fig. 3 TwIST – Results in bad conditions

From Fig. 3, it can be seen that TwIST finds all the endmembers present in the mixture. The results have good accuracy even in bad conditions.

6. CONCLUSIONS

The paper conducted a study over the performances of different spectral unmixing algorithms. The tests showed that TwIST is the most efficient algorithm, having high performances in any conditions. TwIST combines in an efficient way the advantages of IST and IRS class algorithms. ISMA is efficient when the mixing matrix is not bad-conditioned and the noise affecting the observations is not very high. The Moore-Penrose Pseudoinverse is a method strongly influenced by the noise and, because of its unconstrained character, returns unrealistic fractional abundances (which do not satisfy the non-negative and sum-to-one constraints) when the observations are noisy or the mixing matrix is not well-conditioned. OMP has better performances than the Moore-Penrose Pseudoinverse, but fails in returning accurate results in medium or bad conditions.

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)".

Bibliography

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.

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

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

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.

N. Keshava, "A Survey of Spectral Unmixing Algorithms", *Lincoln Laboratory Journal*, vol. 14, no. 1, 2003.

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.

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

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.