SPARSE PIXEL-WISE SPECTRAL UNMIXING - WHICH ALGORITHM TO USE AND HOW TO IMPROVE THE RESULTS

Jakub Bieniarz, Rupert Müller, Xiao Xiang Zhu, Peter Reinartz

German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Wessling, Germany.

ABSTRACT

Recently, many sparse approximation methods have been applied to solve spectral unmixing problems. These methods in contrast to traditional methods for spectral unmixing are designed to work with large *a-prori* given spectral dictionaries containing hundreds of labelled material spectra enabling to skip the expensive endmember extraction and labelling step. However, it has been shown that sparse approximation methods sometimes have problems with selection of correct spectra from the dictionary when these are similar. In this paper we study the detection and approximation accuracy of different sparse approximation methods as well as the influence of the proposed modifications.

Index Terms— Hyperspectral image, sparse unmixing, spectral dictionary.

1. INTRODUCTION

Spectral unmixing is a method for quantitative analysis of hyperspectral images. This analysis includes detection of *endmembers* (unique spectra) and estimation of their *abundance* in each pixel of the hyperspectral image. The usual approach for spectral unmixing is first, to either extract endmembers from the image or to use a preselected set of available known endmembers and second, to estimate their abundance. The abundance estimation step can be done by computing the inverse of the spectral mixing process e.g. using the linear mixing model (LMM) which states that the spectrum of a mixed pixel can be expressed as the weighted integrated sum of the spectra of all endmembers within the pixel [1].

The detection of endmembers especially in highly mixed scenarios is often very hard and requires supervision and expert knowledge to assign the endmember to a real material. To obey this step, recently *sparse spectral unmixing* (SSU) has been proposed [2, 3, 4], a method which uses large spectral dictionaries containing hundreds of labelled endmembers for abundance estimation. These methods, additionally to the LMM assume the sparsity of the abundance vector, i.e., the fact that hyperspectral pixels usually contain only few endmembers when compared to the size of the dictionary.

The most common way to introduce the sparsity prior in SSU is by regularising the abundance estimation error with either ℓ_0 or ℓ_1 norm of the abundance vector for every pixel in the image known as *pixel-wise* SSU. Researchers approached this new problem by applying different algorithms based on e.g. the alternating direction method of multipliers (ADMM) [2], Least Absolute Shrinkage and Selection Operator (LASSO) [3] or Orthogonal matching pursuit (OMP) [5].

As it has been shown in previous works the results of SSU are often negatively influenced by the high mutual similarities between spectra of many materials and noise leading to unstable solutions. To deal with these problems several approaches have been proposed in recent literature. The first one discussed in this paper is the removal of unnecessary elements from the dictionary, i.e., *dictionary pruning*. Pruning is usually done by selecting the most reliable endmembers from a redundant dictionary before performing actual abundance estimation [6, 4]. However, one should note that selecting endmembers from a dictionary, e.g. those the most correlated with the data, does not guarantee selecting the correct ones. Therefore we propose a method which instead of selecting only one one enedmember selects the complete class to which it belongs. The second approach to improve results of the SSU discussed in this paper is the derivative transformation of the dictionary to decrease the coherences.

Besides of the possible improvements, this paper also presents a comprehensive analysis of the three different algorithms for sparse approximation applied to different SSU problem formulations. These findings can help researchers in the decision which algorithm is suitable for their problems.

2. METHODOLOGY

Let us assume a LMM

$$y = Ax + \epsilon \tag{1}$$

where $y \in \mathbb{R}^m$ is the measured spectrum from a hyperspectral pixel, $A \in \mathbb{R}^{[m \times n]}$ is the mixing matrix containing endmembers and $x \in \mathbb{R}^n$ is the abundance vector. When A is overcomplete i.e. it contains more endmembers than spectral channels $n > m$ or even $n >> m$ one could expect x to be sparse.

The expected sparsity of the abundance vector implies the use of sparse approximation methods. Hence, the abundance

estimation problem can be formulated as the non-negative version of the OMP (nOMP)

$$\min\|Ax - y\|_2^2 \text{ s.t. } \|x\|_0 \leqslant \delta_0 \text{ and } x \geqslant 0 \qquad (2)$$

where δ_0 is the upper bound set on the number of endmembers in the spectrum y. The OMP algorithm is efficient with very sparse problems because it can recover a $k-$sparse solution in only k iterations. To solve nOMP problems a slight modification of the OMP Algorithm is required (for details please refer to [7]).

Alternatively, the sparsity can be controlled using the non negative LASSO (nLASSO) formulation

$$\min\|Ax - y\|_2^2 \text{ s.t. } \|x\|_1 \leqslant \delta_1 \text{ and } x \geqslant 0 \qquad (3)$$

where δ_1 is the upper bound set on the sum of all elements in vector x. An algorithm able to solve the nLASSO problem using a modified version of the Least Angle Regression (LARS/LASSO algorithm) has been reported in [8].

The nLASSO problem can be expressed in its equivalent Lagrangian version, i.e., non negative Basis Pursuit Denoising (nBPDN)

$$\min\frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x\|_1 \text{s.t. } x \geqslant 0 \qquad (4)$$

where lambda is a penalty parameter for the ℓ_1 norm of x. Note that this formulation does not provide as intuitive control of the sum of all elements in vector x as the one in (3). The nBPDN problem can be solved using e.g. an ADMM based sparse unmixing via variable splitting augmented Lagrangian (SUnSAL) described in [2].

The probability of reliable sparse unmixing using above algorithms highly depends on the properties of the matrix A and the noise present in the spectrum [2, 3]. Therefore, we propose the following improvements.

To reduce the size of a dictionary and remove unused endmembers a new algorithm for Two Step Group Unmixing (TSGU) can be used. The algorithm in an initial step clusters the dictionary A into a predefined number of clusters k using the *k-means* algorithm. Then the algorithm operates in a pixel-wise manner. First, the initial unmixing using nLASSO or nOMP is performed. Next, a new dictionary is created containing clusters of endmembers in which at least one endmember has positive abundance. Using the pruned dictionary a second unmixing using nLASSO is done and the final abundance vector is approximated. The pseudo code for this algorithm is shown in Algorithm 1.

The use of spectral derivatives to reduce the correlations in the dictionary has been first proposed in [3] and also applied in [5]. The differentiation of spectra does not result in more information contained in the original bands but decreases the background reflectance and can therefore considerably improve the detection of convoluted weaker absorption features [3, 5, 9]. It is also easy to see that applying the spectral derivative to both sides of the equation $\mathcal{S}y = \mathcal{S}Ax$, where

Algorithm 1: TSGU

Input: A, $Y^{[m\times p]}$, δ_1, δ_{TSGU}, ϵ, c

begin

  Clustering: using *K-means* algorithm cluster dictionary A with n elements $\{a_1, a_2, ..., a_n\}$ in to a k sets of clusters where $\mathcal{L} = \{l(a)|a = 1, 2, ..., n\}$ is the set of clusters labels of A;

  initialize: set all $\hat{x}_{j,i}$ for $j \in 1, ..., m, p \in 1, ..., p$ to 0;

  for i = *1,...,p* **do**

    compute the initial $x_j^{(0)}$ abundance using the nLASSO or nOMP;

    find set of clusters $\mathcal{C}$ and labels of active endmembers $\mathcal{L}_\mathcal{C} = \{l(j)|j : x_j^{(0)} \neq 0\}$;

    construct a new pruned dictionary $A_\mathcal{C}$ containing endmembers labeled $\mathcal{L}_\mathcal{C}$;

    compute abundances $\hat{x}_i^\mathcal{C}$ using dictionary with selected clusters $A_\mathcal{C}$ using nLASSO;

    set $[\hat{X}(j,i)|j \in C] \longleftarrow \hat{x}_i^\mathcal{C}$;

return $\hat{X}$;

$\mathcal{S}$ is the spectral derivative operator, does not change the properties of the abundance vector. Therefore, the so transformed image and dictionary can be unmixed using any method based on the LMM.

3. RESULTS

For the experiments the spectral dictionary was created using 432 selected spectra from the USGS spectral library [10] (splib06). Every spectrum has been resampled to match the original spectral response function of the AVIRIS sensor using formulas described in [11]. To simulate spectral mixtures the following simulation scenario was used. First, form a mixing matrix $M \in \mathbb{R}^{[m\times n']}$ by selecting n' endmembers at random from A. Then, an abundance vector $x'^{[n']}$ by randomly assigning elements such as $x_i > 0.01$ and $\|x\|_1 = 1$ was created. Finally, a mixed spectrum by means of LMM $y'^{[m]} = Mx' + \varepsilon'$, where $\varepsilon'^{[m]}$ is Gaussian noise with SNR $[dB] = 10\log_{10}\frac{\|y'[m]\|_2^2}{\|\varepsilon'[m]\|_2^2}$, has been generated. The simulated data consisted of 10 000 spectra containing mixtures of 2 to 10 equally distributed endmembers. In such a way 5 sets of data sets were generated with SNRs of 20dB, 30dB, 40dB, 50dB and 60dB. The experiments were run in Matlab 7b on Intel(R) Core(TM) i5-2520M CPU @ 2.50GHz processor with 8GB of RAM using algorithms reported in [2, 12] as well as Matlab build in functions.

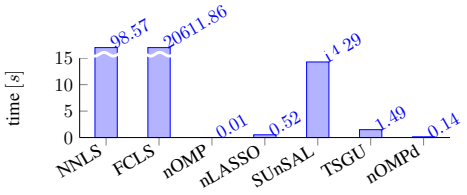

Fig. 1. Execution time of tested algorithms for dataset with 10000 pixels, 224 channels and SNR of 30dB.

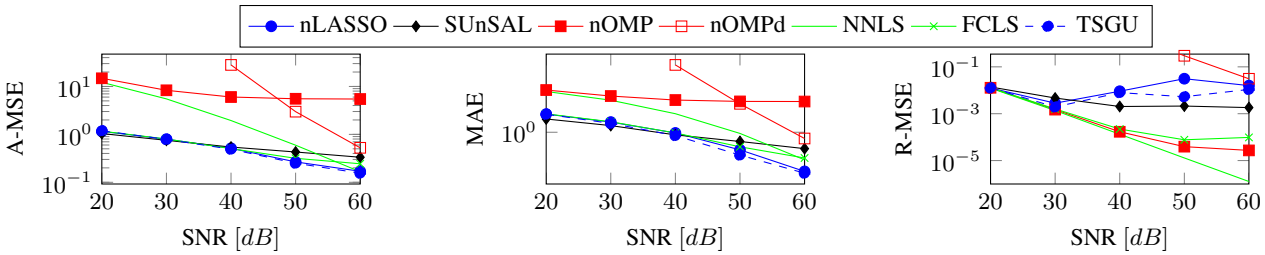

Fig. 2. Plots of mean errors for datasets with different SNR values. The plots present (from left to right) A-MSE, MAE and R-MSE.

We have tested the TSGU and derivative transform applied in nOMP (nOMPd) by comparison with other state of the art algorithms i.e. Nonnegative Least Squares (NNLS), Fully-Constrained Least Squares (FCLS) [1], nLASSO, nOMP and SUnSAL [2]. The algorithms were compared globally by calculating mean values of: (1) the Absolute Mean Square Error (A-MSE) defined as A-MSE $= \frac{\|x-\hat{x}\|_2^2}{\|x\|_2^2}$, (2) the Mean Absolute Error (MAE) defined as MAE $= \frac{\|x-\hat{x}\|_1}{\|x\|_1}$ and (3) the Reflectance Mean Squared Error (R-MSE) defined as R-MSE $= \frac{\|y-A\hat{x}\|_2^2}{\|y\|_2^2}$. The errors has been calculated for all estimated abundance vectors in each data set witch different SNR. The results are shown in Fig. 2. Additionally, we have assessed the results for (4) the endmember detection accuracy (ACC) using ACC $= \frac{\sum \mathrm{TP}+\sum \mathrm{TN}}{\sum \mathrm{P}+\sum \mathrm{N}}$ measure and for (5) the sensitivity of the detection (SNT) using SNT $= \frac{\sum \mathrm{TP}}{\sum \mathrm{P}}$ measure, where TP, TN denote respectively true positive and true negative detection of an endmember, P and N denotes active and inactive endmember in the ground truth.

Considering the A-MSE measure, the algorithms nLASSO, TSGU, SUnSAL and FCLS perform similarly for SNR values equal 20dB and 30dB witch slightly lower mean A-MSE when using SUnSAL. For SNRs from 40dB to 60dB our algorithm TSGU exhibit the lowest A-MSE followed by nLASSO and FCLS. The nOMP method performs worst showing the highest A-MSE values for all test cases. However, the use of nOMP together with the TSGU (TSGU-OMP) significantly improves the results considering all measures. Using the R-MSE measure shows that the algorithm NNLS provide the best estimation of $\hat{y}$ followed by nOMP and FCLS (the results are shown in Fig. 3). Note, that the algorithms NNLS and nOMP have low R-MSE but high A-MSE and MAE values indicating very good reconstruction of spectrum but poor estimation of abundance vector. The first and second column in Fig. 3 shows the dependency of the number of mixed endmembers to A-MSE and MAE. In both cases for all algorithms the error rises with the increase of the contributing endmembers. Here the benefit of using the derivative transform can bee seen in the high SNR scenario e.g. for SNR=60dB results of nOMPd are comparible to other methods. The ACC of endmember detection depending on the number of mixed spectra in the ground truth an the SNR value is plotted in the third column in Fig. 3. All algorithms exhibit high values of ACC ranging from 0.93 for spectra mixed of 10 spectra to 0.98 for mixtures of only two endmembers. The ACC appears to remain similar for all SNR values. Algorithms TSGU, nOMP and nOMPd for all SNR are among three with the highest ACC. The SNT of the detection for all tested algorithms is shown in the fourth column in Fig. 3. The LASSO based algorithms together with NNLS and FCLS exhibit the best detection of endmembers for all algorithms.

Comparing the execution time of tested algorithms the FCLS perform the worst out of all algorithms needing 20611.86 seconds. NNLS converged in 98.57 seconds. The ℓ_0 regularized method nOMP and nOMPd performed the fastest with 0.01 second for the whole dataset. The second fastest algorithm was nLASSO with 0.53 seconds. Both algorithms with preselection step TSGU and TSGU-OMP performed slower than nOMP or nLASSO but were still significantly faster than conventional approaches. SUnSAL was the slowest algorithm out of the pixel based sparse regularized methods. The execution times for all algorithms are reported in Fig. 1.

4. CONCLUSIONS

In this paper we presented a comparison of SSU approaches using different algorithms and their modifications. The following has been found: (1) SSU methods with ℓ_1 minimization are at least as good as FCLS and better than NNLS with the execution time reduced by several orders of magnitude. (2) With noisy data SNR<40dB nLASSO, SUnSAL, and FCLS exhibit the best performance from which LASSO is the fastest. (3) The endmember preselection used in our TSGU algorithm can significantly increase the accuracy and sensitivity of detection of endmembers. (4) For data with SNR>50dB the nOMP with derivative transformed dictionary is the fastest option with similar performance to other algorithms. For future work we address similar tests with real data.

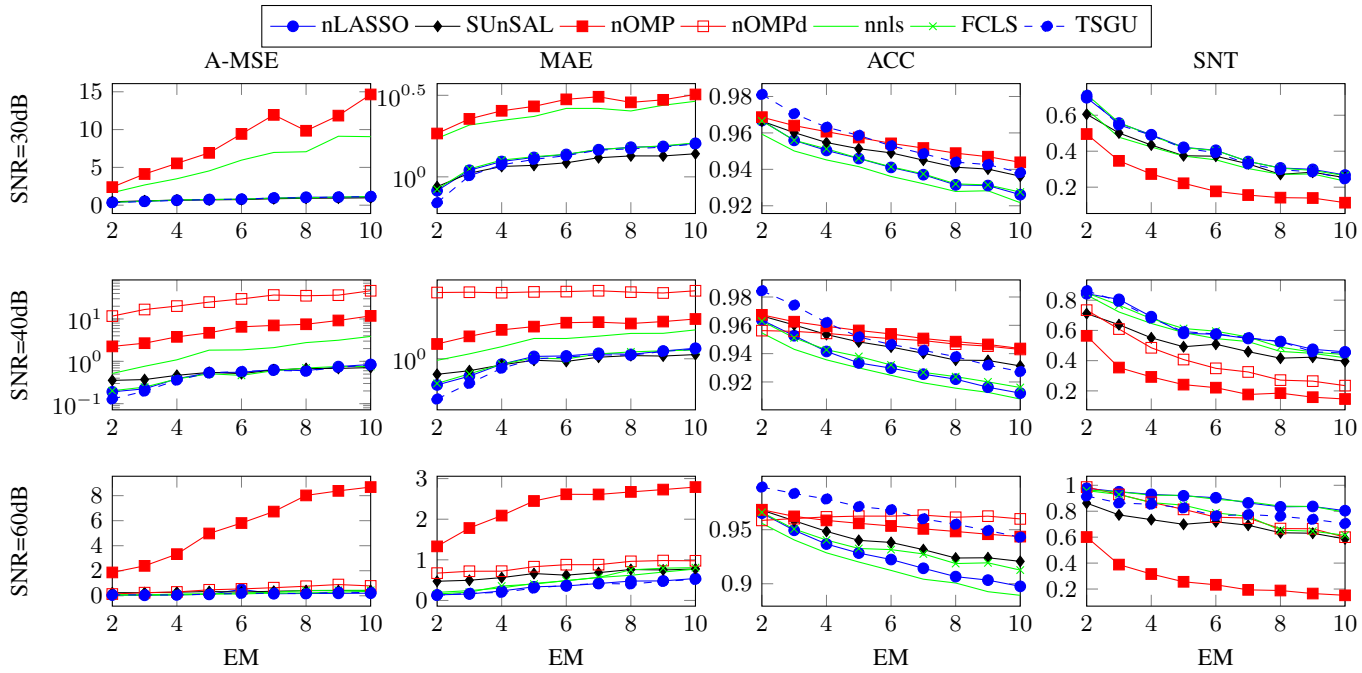

Fig. 3. Results of the assessment of unmixing experiments plotted for different SNR values in rows and using different measures in columns. The results for nOMPd are plotted only when the results were comparable to other methods, otherwise nOMPd performed much worse than the competitors.

5. REFERENCES

[1] N. Keshava, "A survey of spectral unmixing algorithms a survey of spectral unmixing," *Lincoln Laboratory Journal*, vol. 14, pp. 55–78, 2003.

[2] D. M. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens*, vol. 49, no. 6, pp. 2014–2039, 2011.

[3] J. Bieniarz, R. Müller, X. X. Zhu, and P. Reinartz, "On the use of overcomplete dictionaries for spectral unmixing," in *4th Workshop on Hyperspectral Image and Signal Processing. WHISPERS 2012*, 2012.

[4] W. Tang, Z. Shi, and Y. Wu, "Regularized simultaneous forward-backward greedy algorithm for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens*, vol. 52, no. 9, pp. 5271–5288, 2014.

[5] N. Akhtar, F. Shafait, and A. Mian, "Futuristic greedy approach to sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens*, PP, 1–18, 2015.

[6] M.-D. Iordache, J.M. Bioucas-Dias, A Plaza, and B. Somers, "Music-csr: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression," *IEEE Trans. Geosci. Remote Sens*, vol. 52, no. 7, pp. 4364–4382, 2014.

[7] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "Sparse non-negative solution of a linear system of equations is unique," in *Proc. 3rd Int. Symp. Communications, Control and Signal Processing ISCCSP 2008*, 2008, pp. 762–767.

[8] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of statistics*, vol. 32(2), pp. 407499, 2004.

[9] G. Talsky, *Derivative Spectrophotometry of First and Higher Orders*, VCH Verlagsgesellschaft mbH, Weinheim, VCH Publishers, Inc., New York, NY (USA), 1994.

[10] R.N. Clark, G.A. Swayze, R. Wise, E. Livo, T. Hoefen, R. R. Kok, and S.J. Sutley, "USGS digital spectral library splib06a," Digital Data Series 231, 2007.

[11] J. Bieniarz, D.Cerra, J. Avbelj, P. Reinartz, and Rupert Müller, "Hyperspectral image resolution enhancement based on spectral unmixing and information fusion," in *ISPRS Hannover Workshop 2011*, 2011.

[12] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.