SPECTRAL UNMIXING WITH SPARSITY AND STRUCTURING CONSTRAINTS

Ramzi Ben Mhenni[1]   *Sébastien Bourguignon*[1]   *Jordan Ninin*[2]   *Frédéric Schmidt*[3]

[1] École Centrale de Nantes, LS2N, 1 rue de la Noë, F-44321 Nantes, France
[2] ENSTA-Bretagne, Lab-STICC, 2 rue François Verny, F-29806 Brest cedex 9, France
[3] Université Paris-Sud, GEOPS, 1 rue du Belvédère, F-91405 Orsay, France

ABSTRACT

This paper addresses the linear spectral unmixing problem, by incorporating different constraints that may be of interest in order to cope with spectral variability: sparsity (few nonzero abundances), group exclusivity (at most one nonzero abundance within subgroups of endmembers) and significance (non-zero abundances must exceed a threshold). We show how such problems can be solved *exactly* with mixed-integer programming techniques. Numerical simulations show that solutions can be computed for problems of limited, yet realistic, complexity, with improved estimation performance over existing methods, but with higher computing time.

Index Terms— sparse spectral unmixing, ℓ_0-norm optimization, structured sparsity, mixed-integer programming.

1. INTRODUCTION

Spectral unmixing consists in decomposing an observed reflectance spectrum into a linear mixture of elementary spectra (*endmembers*) and estimating their proportions (*abundances*) [1, 2]. We focus on abundance estimation, and endmembers are supposed to be known, *e.g.,* obtained by laboratory spectroscopic measurements or estimated from the hyperspectral data set, forming a *dictionary* used for decomposition. In order to cope with spectral variability, spectral libraries may contain an increasingly high number of spectra—the minerals data base provided by the United States Geological Survey (USGS), for example, increased from 504 spectra in 2003 [3] to 1276 spectra in 2017 [4]. Dealing with a large number of endmembers is acknowledged as a critical issue, especially because many spectra are highly correlated [5, 6]. In such cases, the standard approach based on least-squares fitting under positivity and possibly sum-to-one constraints [7], may perform badly, and stronger constraints are required, some of which are considered in this paper.

The first one is *sparsity*: usually, each mixture contains only a few components. Second, imposing structural constraints on the dictionary is also of interest in order to cope with variability. In particular, experimental data concerning a mineral are usually acquired in various conditions, and the dictionary may include such different spectra, although at most one is actually used in the decomposition (*exclusivity* constraint). Last, imposing a minimum value to each non-zero abundance (*significance* constraint) also promotes sparsity: conjugated with positivity and sum-to-one constraints, it concentrates the solution into few components, without imposing a maximum number of non-zero abundances.

To our knowledge, such constraints have never been addressed *exactly*, probably because of their numerical difficulty. Cardinality constraints resort to combinatorial optimization. Sparse spectral unmixing has been proposed *e.g.,* with ℓ_1-norm [8, 9] or $\ell_{p,\,p<1}$-norm [10] regularization, or with a greedy Orthogonal Matching Pursuit procedure [9]. While computationally attractive, such approaches usually do not solve the original sparse problem, and they face difficulties in considering the sum-to-one constraint and non-normalized dictionaries. Both exclusivity and significance constraints are basically *logical* constraints, that cannot be addressed by standard optimization (exclusivity is mentioned *e.g.* in [6], but no algorithmic solution is proposed). This paper considers the exact resolution of such problems through mixed-integer programming (MIP) techniques, which mix continuous and discrete variables and offers a natural framework for such constraints. Although MIP optimization is usually NP-hard, it has been investigated for long in operations research and very efficient solvers were developed [11].

In this paper, we first show how the above constraints can be exactly translated into MIPs; their exact resolution is then performed through dedicated software. Second, we evaluate the obtained solutions through numerical simulations in their ability to correctly solve unmixing problems, as a function of the problem complexity (cardinality and noise level).

2. CONSTRAINED LINEAR SPECTRAL UNMIXING

We consider the standard linear mixture model [1, 2]:

$$\boldsymbol{y} = \sum_{n=1}^{N} a_n \boldsymbol{s}_n + \boldsymbol{\epsilon} = \mathbf{S}\boldsymbol{a} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y} \in \mathbb{R}^L$ is a reflectance spectrum acquired in L spectral bands, $\boldsymbol{s}_n \in \mathbb{R}^L$ represents the n-th endmember, $a_n \in \mathbb{R}$

is the associated abundance, $\boldsymbol{a} = [a_1, \ldots, a_n]^T$, $\boldsymbol{\epsilon}$ represents noise and model errors and $\mathbf{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N]$ is the dictionary of N endmembers, with usually $N \gg L$. In order to account for physical considerations (abundances are percentages), positivity and sum-to-one constraints are usually added. The resulting *Fully Constrained Least-Squares* (FCLS) estimate [7] reads:

$$\min_{\boldsymbol{a} \in \mathbb{R}^N} \|\boldsymbol{y} - \mathbf{S}\boldsymbol{a}\|^2 \text{ such that (s.t.) } \boldsymbol{a} \geq \boldsymbol{0}, \sum_{n=1}^{N} a_n = 1, \quad (1)$$

where vector inequality is taken componentwise. In the presence of noise, and if the dictionary $\mathbf{S}$ contains highly correlated spectra, FCLS may not be efficient, as shows the example in Figure 1, Section 4: estimated abundances are spread over many non-zero, but small, values, which does not correspond to physical reality and lead to erroneous interpretation. In such cases, imposing more constraints may be necessary.

2.1. Cardinality constraint (ℓ_0-norm sparsity)

In most cases, only a small number of elementary spectra are used for representing a mixture: $\boldsymbol{a}$ should be sparse, that is, only a few a_n are non-zero. Let us remark that formulation (1), through the positivity constraint, produces solutions for which some abundances are zero. However, on difficult problems, FCLS generally produces a large number of small-valued abundances and may fail in locating the true components (see simulation results in Section 4). Enforcing more sparsity may contribute to improve abundance estimation [9, 12, 5]. Introducing the ℓ_0 "norm" $\|\boldsymbol{a}\|_0 = \text{Card}\{n | a_n \neq 0\}$, the sparse linear unmixing problem can be written as the cardinality-constrained optimization problem:

$$\min_{\boldsymbol{a} \geq 0} \|\boldsymbol{y} - \mathbf{S}\boldsymbol{a}\|^2 \text{ s. t. } \|\boldsymbol{a}\|_0 \leq K, \sum_{n=1}^{N} a_n = 1. \quad (2)$$

Since solving (2) is essentially a combinatorial problem, suboptimal estimation strategies are usually preferred for computational efficiency. Substituting the ℓ_0 norm with the convex ℓ_1 norm has been a standard in sparsity-enhancing methods for a long time. However, it is not fully appropriate here: since abundances are positive and sum to one, their ℓ_1 norm is fixed to 1! Therefore, the ℓ_1 norm cannot be used for enforcing sparsity in problem (1). Indeed, we note that the works in [8, 9] exploiting ℓ_1-norm strategies discard the sum-to-one constraint. While we acknowledge that imposing a lower value on $\sum a_n$ may act as a sparse regularizer, in our experiments, however, the best ℓ_1-norm-based solution was always found with $\sum a_n \simeq 1$, that is, the FCLS solution (1). The non-convex $\ell_{p,p<1}$-norm regularization recently proposed in [10] may be more suited for enforcing sparsity, although optimization strategies in [10] also discard the sum-to-one constraint. Standard greedy iterative algorithms [13]

are also unadapted to cope with the sum-to-one constraint and with non-normalized dictionaries: since endmembers are reflectance spectra, the dictionary cannot be normalized without affecting the abundance estimation [12]. Finally, the backward procedure in [12] iteratively removes one by one the small-but-non-zero components in the FCLS solution. Although this method has no optimality guarantee with respect to problem (2), it provides a sparse solution by running iterations until a few abundances are non-zero. Therefore, it will be used for comparison in our experiments.

2.2. Group exclusivity constraint

In order to improve the model accuracy in dictionary-based spectral unmixing, a given mineral can be represented by several possible spectra, depending on grain size, acquisition geometry, impurities, etc. Also, mineral samples used for endmember laboratory measurements may come from different places on Earth, resulting in slight differences in impurities. In addition, some minerals (*e.g.,* olivine or pyroxene) are actually solid-state solutions with a continuum of possible compositions, creating a variability in the spectral shape. In such cases, the spectral library may include different possible variations of a given mineral, although at most one may be activated in a given mixture [6]. Let the component indices be split[1] into J groups $G_1, \ldots, G_J$, and let $\boldsymbol{a}^{(j)}$ denote the corresponding abundance subset. We address the following optimization problem, which imposes that the mixture contains at most one spectral element in each group G_j:

$$\min_{\boldsymbol{a} \geq 0} \|\boldsymbol{y} - \mathbf{S}\boldsymbol{a}\|^2 \text{ s. t. } \begin{cases} \|\boldsymbol{a}^{(j)}\|_0 \leq 1, j = 1, \ldots, J \\ \sum_{n=1}^{N} a_n = 1, \end{cases} \quad (3)$$

Note that such *group exclusivity* (GE) constraints also enhance more sparsity in the FCLS solution (1).

2.3. Significant abundances only

If the spectral dictionary is highly correlated and in the presence of noise, the FCLS solution usually shows (possibly many) small-amplitude components. Such artifacts can be easily eliminated after some post-processing step; however, they may also perturbate detection of the true components (see Section 4, Figure 1, left column). Therefore, one may want to impose a minimum value to non-zero abundances, in order to detect only significant contributions. We call these constraints *significant abundances* (SA). The last optimization problem we are interested in therefore reads:

$$\min_{\boldsymbol{a} \geq 0} \|\boldsymbol{y} - \mathbf{S}\boldsymbol{a}\|^2 \text{ s. t. } \begin{cases} (a_n \neq 0 \Rightarrow a_n \geq \tau), n = 1, \ldots, N \\ \sum_{n=1}^{N} a_n = 1. \end{cases}$$
$$(4)$$

---

[1] We consider here a partition of the dictionary, but similar constraints operating on overlapping groups could also be accounted for.

It represents another sparsity-promoting formulation (less than $1/\tau$ abundances are non-zero), that may be preferred to the ℓ_0-norm problem (2). In particular, tuning the related threshold τ can be easier in practice than specifying the number of non-zero abundances that are searched in the mixture.

3. REFORMULATIONS AS MIXED-INTEGER PROGRAMS

We now formulate the three constrained spectral unmixing problems (with sparsity, exclusivity and significance constraints) as mixed-integer programs (MIPs). Since these constraints are discrete or logical ones, we introduce binary variables b_n indicating the presence or absence of a given endmember in the mixture: $b_n = 1 \Leftrightarrow a_n \neq 0$. Remarking that abundances $a_n \in [0,1]$, one can easily show that the former equivalence can be exactly translated as $0 \leq a_n \leq b_n$, giving N linear inequality constraints mixing binary and real variables, that can be efficiently tackled by MIP optimization.

The ℓ_0 norm can now be simply written as the sum of the binary variables: $\|a\|_0 = \sum_n b_n$. Therefore, the cardinality-constrained problem (2) is equivalently reformulated as:

$$\min_{a \in \mathbb{R}^N, b \in \{0,1\}^N} \|y - Sa\|^2 \ \text{ s. t. } \ \begin{cases} 0 \leq a \leq b \\ \sum_n b_n \leq K \\ \sum_n a_n = 1 \end{cases}.$$

Similarly, the exclusivity-constrained problem (3) reads:

$$\min_{a \in \mathbb{R}^N, b \in \{0,1\}^N} \|y - Sa\|^2 \ \text{s. t.} \ \begin{cases} 0 \leq a \leq b \\ \sum_{i \in G_j} b_i \leq 1, j = 1, \ldots, J \\ \sum_{n=1}^N a_n = 1. \end{cases}$$

Finally, one can show that the logical constraints in (4) can be equivalently written as $\tau b_n \leq a_n \leq b_n$, so that the problem with significance constraints (4) reads:

$$\min_{a \in \mathbb{R}^N, b \in \{0,1\}^N} \|y - Sa\|^2 \ \text{ s. t. } \ \begin{cases} \tau b_n \leq a_n \leq b_n, \ \forall n \\ \sum_{n=1}^N a_n = 1 \end{cases}.$$

These three optimization problems share a common structure: a quadratic cost function, linear equality and inequality constraints, and binarity of some variables. Efficient resolution methods have been developed in operations research for such *quadratic MIPs*, based on branch-and-bound tree exploration strategies. In this paper, resolution is performed with IBM ILOG CPLEX, which is unanimously considered among the best MIP solvers [11].

Finally, although considered separately in the former description, the three constraints can obviously be mixed in order to build more complex estimates, and resulting problems still form quadratic MIPs. Note also that alternate formulations that would switch the cost function and the constraints (*e.g.*, minimize the ℓ_0 norm under a bounded approximation error constraint) result in *quadratically-constrained MIPS*, whose resolution by MIP solvers is much less efficient [14].

4. SIMULATION RESULTS

We study the performance of these approaches on simulated spectral unmixing problems. The dictionary is composed of reflectance spectra of minerals taken in the USGS database [3], possibly present on planetary surfaces [15]. It contains $N = 481$ spectra, among which $J = 85$ groups of spectra with same mineral name contain from 2 to 18 endmembers, and 146 endmembers are not clustered into groups. Spectral sampling represents $L = 113$ wavelengths from 1 to 2.5 microns. Abundances are uniformly distributed above $\tau = 0.1$ and sum to one. Spectral components are randomly chosen, with at most one endmember in each group. Gaussian noise ϵ is added with $\text{SNR}_{\text{dB}} = 10 \log(\|Sa\|^2/\|\epsilon\|^2)$ equal to 55 and 40 dB. Due to the positivity of the spectra (which have therefore nonzero mean), such values seem particularly high, but they actually correspond to low and medium noise levels (see Figure 1 for an example). In both cases, data are sufficiently noisy so that the FCLS estimate almost always fails in retrieving the true abundance set, as will be seen later.

Abundances are estimated by FCLS, by the backward method in [12] and by the following constrained estimates: ℓ_0-norm sparsity (ℓ_0), exclusivity (GE), ℓ_0-norm sparsity and exclusivity (ℓ_0+GE), significance and exclusivity (SA+GE). The best ℓ_1-norm-based solutions were obtained by tuning the ℓ_1 norm close to 1 (see § 2.1), and therefore gave similar results to FCLS. The threshold τ for significance constraints is set to 0.1 and cardinality constraints are set to their true value. Resolution is performed on a standard desktop computer.

Figure 1 shows an example of results, with SNR = 40 dB and $K = 4$. FCLS correctly detects only 3 spectra, with underestimated amplitudes, and produces many erroneous detections. In this example, adding the GE constraint improves the detection since the 4 spectra are correctly detected. However, some artifacts can still be found, and abundances of true components are therefore badly estimated. The backward method correctly detects 3 spectra, with slightly misestimated amplitudes. On the other hand, the three ℓ_0, ℓ_0+GE and SA+GE estimates perfectly detect the four components, with fairly estimated amplitudes (we only show the first estimate, since the three are identical).

Figure 2 shows the quadratic error between true and estimated abundances as a function of K, averaged over 30 data realizations, with low (55 dB) and medium (40 dB) noise levels. GE improves over simple FCLS, and the backward performs better FCLS and GE at high SNR and low K. The three ℓ_0, ℓ_0+GE and SA+GE estimates always give the best results, with perfect detections at high SNR (even for $K = 7$, whereas FCLS and GE fail from $K = 2$ and the backward method fails from $K = 4$). Of course, their performance decreases as the problem complexity increases: then, the ℓ_0 minimizer, even correctly computed, does not find the true solution. For SNR = 55 dB and $K > 6$, the problem complexity does not make it possible to compute the solution in the maximum time al-
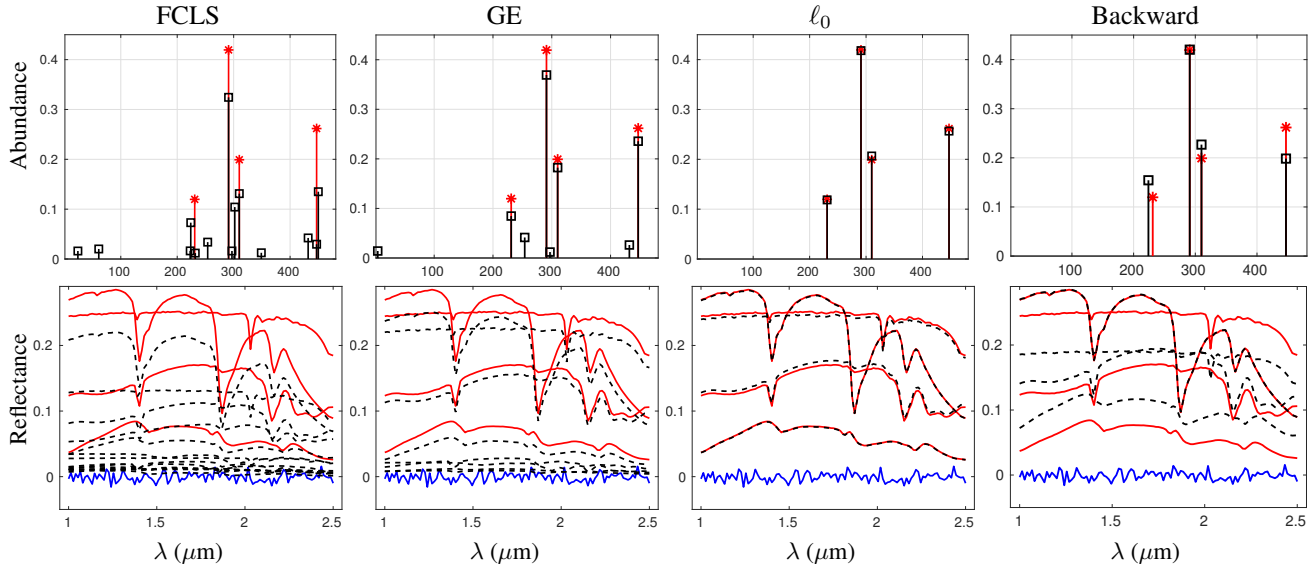
Fig. 1. Example of result, $K = 4$ spectra, SNR = 40 dB. Top: estimated (black squares) and true (red stars) abundances. Bottom: true (solid red line) and estimated (dashed black line) endmembers weighted by their abundances, and noise (blue solid line).
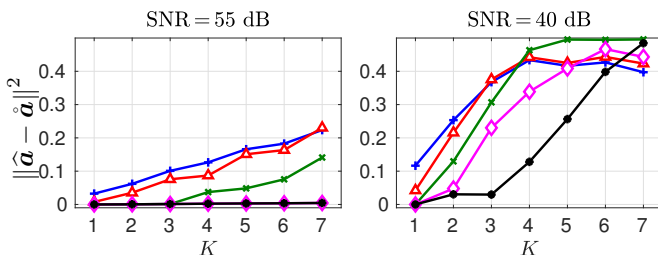

Fig. 2. Error on estimated abundances for FCLS (+), GE (△), ℓ_0 (o), SA+GE (◇), and backward [12] (x).

SNR	Method	K=3	K=4	K=5	K=6	K=7
55 dB	GE	1.3	2.2	3.3	5.9	9.8
	ℓ_0	1.5	3	11.8	$124^{(3)}$	$468^{(9)}$
	ℓ_0+GE	1.6	3.1	7.1	$82^{(1)}$	$320^{(7)}$
	SA+GE	2.5	11.5	29	76	$264^{(5)}$
40 dB	GE	1.3	2.2	2.1	3.2	2.9
	ℓ_0	9.2	$107^{(1)}$	$574^{(13)}$	$963^{(27)}$	$1000^{(30)}$
	ℓ_0+GE	8.1	75	$515^{(11)}$	$929^{(25)}$	$982^{(28)}$
	SA+GE	$282^{(6)}$	$426^{(8)}$	$746^{(19)}$	$936^{(26)}$	$997^{(29)}$

Table 1. Computing times (s) for optimization of MIP problems, averaged over 30 instances. In parentheses: number of instances for which optimization did not terminate in 1000 s. The computing time of the backward method is almost constant same for all problems, around 32 s.

lowed for MIP resolution (1000 s).

Table 1 gives average computing times obtained on the former instances, for GE, ℓ_0, ℓ_0+GE and SA+GE constraints. Including GE constraints is never computationally prohibitive, although it requires the resolution of a MIP. The computing times of the three ℓ_0-norm-based estimates are still reasonable for low noise level or small K, but strongly increase with these two factors. In particular, the optimal solution of ℓ_0+GE is obtained in less than 1000 s, on all instances up to $K = 6$ for SNR = 55, and up to $K = 4$ for SNR = 40. Problems with significance constraints are much more difficult to solve, as show the rows 'SA+GE' in Table 1. In all our simulations, ℓ_0 and ℓ_0+GE always gave the same solution. However, adding GE constraints allows better numerical resolution for the most difficult problems. Finally, the computing time of the backward method in [12] is rather high (approximately 32 s for all problems), since it requires the resolution of many (up to $N - K$) FCLS problems.

5. CONCLUSION

Complex constraints for spectral unmixing (cardinality, group exclusivity, significant abundances) were taken into account *via* MIP formulations. When such constraints are valid in practice, they may be advantageously used for improving abundance estimation. Exact optimization of the resulting problems was shown to be possible for problems of limited complexity. In particular, accounting for group exclusivity only reasonably increases the computing time. More involved sparsity-inducing constraints (cardinality or significant abundances) are computationally more demanding. In our simulations with 481 endmembers, exact resolution of ℓ_0-norm problems was still possible up to $K = 6$ at high SNR and $K = 4$ at medium SNR, which is still realistic in many cases.

Incorporating binary indicating variables offers a flexible framework, which may be used for considering other high-level structuring constraints. Adding more constraints to the problems studied in this paper is also a lever for improving their numerical resolution. Last, dedicated MIP optimization strategies are currently under investigation in order to reduce the computing time.

6. REFERENCES

[1] R. B. Singer and T. B. McCord, "Mars - Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Lunar and Planetary Science Conference Proceedings*, 1979, vol. 10.

[2] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE signal processing magazine*, vol. 19, no. 1, 2002.

[3] R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. Hoefen, R. F. Kokaly, and S. J. Sutley, "USGS digital spectral library splib05a," *US Geological Survey, Digital Data Series*, vol. 231, 2003.

[4] R.F. Kokaly, R.N. Clark, G.A. Swayze, K.E. Livo, T.M. Hoefen, N.C. Pearson, R.A. Wise, W.M. Benzel, H.A. Lowers, R.L. Driscoll, and 2017 Klein, A.J., "USGS Spectral Library Version 7 Data," *U.S. Geological Survey data release*, 2017.

[5] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, April 2012.

[6] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, Jan 2014.

[7] D. C. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, Mar 2001.

[8] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," in *Proc SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, 05 2009.

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

[10] D. Tuia, R. Flamary, and M. Barlaud, "Nonconvex regularization in remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6470–6480, Nov 2016.

[11] R. E. Bixby, "A brief history of linear and mixed-integer programming computation," *Doc. Math.*, vol. Optimization Stories, 2012.

[12] J. B. Greer, "Sparse demixing of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 219–228, Jan 2012.

[13] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[14] S. Bourguignon, J. Ninin, H. Carfantan, and M. Mongeau, "Exact sparse approximation problems via mixed-integer programming: Formulations and computational performance," *IEEE Trans. Signal Process.*, vol. 64, no. 6, March 2016.

[15] F. Schmidt, M. Legendre, and S. Le Mouëlic, "Minerals detection for hyperspectral images using adapted linear unmixing: LinMin," *Icarus*, vol. 237, 2014.