Constrained reflect-then-combine methods for unmixing hyperspectral data

Paul Honeine, Henri Lantéri

CONSTRAINED REFLECT-THEN-COMBINE METHODS FOR UNMIXING HYPERSPECTRAL DATA

Paul Honeine

Institut Charles Delaunay (CNRS)
Université de technologie de Troyes
Troyes, France
paul.honeine@utt.fr

Henri Lantéri

Laboratoire Lagrange (CNRS)
Université de Nice Sophia-Antipolis
Nice, France
henri.lanteri@unice.fr

ABSTRACT

This paper deals with the linear unmixing problem in hyperspectral data processing, and in particular the estimation of the fractional abundances under sum-to-one and non-negativity constraints. For this purpose, we propose to adapt the reflect-then-combine iterative technique, initially derived by Cimmino. Several strategies are studied in order to handle the constraints, and experimental results are analyzed.

***Index Terms*—** Constrained optimization, hyperspectral data, unmixing problem, parallel projection, Cimmino's method

1. INTRODUCTION

In hyperspectral data processing, spectral unmixing is one of the most fundamental and challenging problems. The spectral unmixing problem consists of breaking down a spectrum into a set of pure spectra, a.k.a. endmembers, and their fractional abundances. Assuming that the endmembers were extracted using any off-the-shelf technique (see [1] for a survey), the estimation of the abundances brings new opportunities and challenges to both linear [6] as well as non-linear unmixing problems [3]. To adopt a physical interpretation, the estimation problem requires the fulfillment of two constraints: the fractional abundances must add up to 100% (a.k.a. sum-to-one constraint), and must be additive (a.k.a. non-negativity constraint). See for instance [5]. This paper deals with constrained fractional abundances estimation, by suitably adapting parallel orthogonal projections/reflections.

Methods of orthogonal projections onto subspaces have been bridging the words of geometry and algebra, successfully applied in solving optimization problems. While the alternating projection algorithm goes back at least to Hermann A. Schwarz in the 1870's, it was revisited in the late 1930's by Kaczmarz's cyclic projections [7] and Cimmino's parallel reflections [4]. More recently, there has been a modern revival of projection algorithms, namely with projections onto convex sets within adaptive filtering and machine learning [9, 8]. These methods open tremendous opportunities in various domains in signal and image processing.

Hyperspectral data processing still do not take advantage of this increasing research activity, even in the fundamental linear spectral unmixing problem. In this paper, we propose to tackle the major obstacle, which is the fulfillment of the physical interpretation, namely by enforcing non-negativity and sum-to-one constraints within an iterative projection scheme. To this end, we revisit the *parallel reflect then combine* update rule, initially proposed by Cimmino [4]. We consider two strategies to handle the sum-to-one constraint, either by normalization or by an asymptotic convergence. The non-negativity constraint is enforced by two different strategies, either by relaxing the reflection or by projecting onto the non-negative orthant.

This work was supported by the French ANR, grant HYPANEMA: ANR-12-BS03-0003.

2. THE LINEAR (UN-)MIXING MODEL

Given a spectrum with L wavelength $\boldsymbol{x} = [x_1 \; x_2 \; \cdots \; x_L]^\top$ (e.g. a pixel in a hyperspectral image), the linear mixing model takes the form

$$\boldsymbol{x} = \sum_{k=1}^{K} \alpha_k \, \boldsymbol{m}_k + \boldsymbol{\epsilon},$$

where $\boldsymbol{m}_k$ is the spectral signature of a pure material, i.e., endmember, and $\boldsymbol{\epsilon}$ is the vector of fitness error. In matrix form, we get

$$\boldsymbol{x} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{M} = [\boldsymbol{m}_1 \; \boldsymbol{m}_2 \; \cdots \; \boldsymbol{m}_K]$, and $\boldsymbol{\alpha} = [\alpha_1 \; \alpha_2 \; \cdots \; \alpha_K]^\top$ is the vector of abundances to be determined. In this work, it is assumed that the endmembers have been identified, using any off-the-shelf endmember extraction technique. See for instance [1, 6] and references therein. The (unconstrained) optimization problem, i.e., $\arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{x} - \boldsymbol{M}\boldsymbol{\alpha}\|^2$, leads to the optimal solution $\boldsymbol{\alpha} = \boldsymbol{M}^\top(\boldsymbol{M}\boldsymbol{M}^\top)^{-1}\boldsymbol{x}$, with optimality in the least-squares sense.

In order to adopt a physical interpretation in the unmixing model, the above linear combination must be convex, namely by satisfying the sum-to-one and the non-negativity constraints, respectively $\mathbf{1}^\top \boldsymbol{\alpha} = 1$ and $\boldsymbol{\alpha} \geq 0$, where $\mathbf{1}$ denotes the unit column-vector of K entries, and the inequality is taken component-wise. An iterative scheme is required to solve this constrained optimization problem, with essentially two steps at each iteration: an approximation step that determines an intermediate solution by minimizing the fitness error; and a step that constrains the solution. The latter step sets to zero the negative values in order to impose the non-negativity constraint, and normalizes to one to enforce the sum-to-one constraint.

Due to its nature, the sum-to-one constraint can be easily handled in the least-squares optimization problem. To this end, the following augmented model is substituted for (1):

$$\begin{bmatrix} 1 \\ \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \mathbf{1}^\top \\ \boldsymbol{M} \end{bmatrix} \boldsymbol{\alpha} + \begin{bmatrix} \epsilon \\ \boldsymbol{\epsilon} \end{bmatrix}. \tag{2}$$

This model has been recently investigated in the literature for unmixing hyperspectral data, as studied for instance in [6] using a simplex-based approach. The solution is given by analogy to the above unconstrained problem. Still, the solution is optimal in the least-squares sense, thus leading to a fitness error ϵ within the sum-to-one, as opposed to the explicit constraint enforced in the above iterative scheme. The distinction between these two strategies (explicitly imposed vs. accepting fitness error) is explored in this work in order to derive several constrained reflect-then-combine methods.

3. CONSTRAINED REFLECT-THEN-COMBINE METHODS

3.1. Unconstrained reflect-then-combine method

Back to the unconstrained problem defined in (1), we consider the set of L equations:

$$x_\ell = \widetilde{\boldsymbol{m}}_\ell^\top \boldsymbol{\alpha}, \qquad \text{for } \ell = 1, 2, \ldots, L \tag{3}$$

where $\widetilde{\boldsymbol{m}}_\ell^\top$ denotes the vector of the K endmember spectral signatures at the ℓ-th wavelength band. In other words, the ℓ-th row of $\boldsymbol{M}$ is denoted by $\widetilde{\boldsymbol{m}}_\ell^\top$. The solution of the unmixing problem is the unique[1] intersection of the L (affine) hyperplanes defined by the above equations.

Cimmino's method consists of two steps to refine the approximate solution $\boldsymbol{\alpha}^{(t)}$ at iteration t, as described by the following update equations:

$$\boldsymbol{\alpha}_{/\ell}^{(t)} = \boldsymbol{\alpha}^{(t)} + 2\,\frac{x_\ell - \widetilde{\boldsymbol{m}}_\ell^\top \boldsymbol{\alpha}^{(t)}}{\|\widetilde{\boldsymbol{m}}_\ell\|^2}\,\widetilde{\boldsymbol{m}}_\ell \tag{4}$$

$$\boldsymbol{\alpha}^{(t+1)} = \sum_{\ell=1}^{L} \gamma_\ell^{(t)}\, \boldsymbol{\alpha}_{/\ell}^{(t)} \tag{5}$$

The first step involves intermediate solutions, by taking the reflection (i.e., mirror image) of $\boldsymbol{\alpha}^{(t)}$ with respect to each of the hyperplanes defined by (3). The second step combines the resulting reflections through weights $\gamma_\ell^{(t)}$ that satisfy the the convexity constraints, i.e., $\sum_{\ell=1}^{L} \gamma_\ell^{(t)} = 1$ and $\gamma_1^{(t)}, \gamma_2^{(t)}, \ldots, \gamma_L^{(t)} \geq 0$ at each iteration t.

The motivation of this "reflect-then-combine" rule is illustrated in Figure 1. In fact, any convex combination of the reflections provides a better solution that the initial one. In particular, as often applied by Cimmino, one may take the center of gravity of the intermediate solutions, namely by setting $\gamma_\ell^{(t)} = 1/L$ for all $\ell = 1, 2, \ldots, L$. It is worth noting that this strategy converges faster than a "project-then-combine" strategy, as shown in Figure 1 (dotted red line vs blue line).

By merging the reflection and combination steps into a single expression, one can write the following update rule:

$$\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)} + 2\,\boldsymbol{M}^\top \boldsymbol{\Gamma}(\boldsymbol{x} - \boldsymbol{M}\,\boldsymbol{\alpha}^{(t)}), \tag{6}$$

where $\boldsymbol{\Gamma}$ is a L-by-L diagonal matrix with entries $\gamma_1/\|\widetilde{\boldsymbol{m}}_1\|^2$, $\gamma_2/\|\widetilde{\boldsymbol{m}}_2\|^2$, ..., $\gamma_L/\|\widetilde{\boldsymbol{m}}_L\|^2$. This method has several desirable properties, such as parallelization as given by the update rule (4) which can be processed separately on a L-core machine, as well as multiple pixel processing. For the latter case, by collecting the $\boldsymbol{x}$ and the corresponding $\boldsymbol{\alpha}^{(t)}$ of each pixel into $\boldsymbol{X}$ and $\boldsymbol{A}^{(t)}$ respectively, we get the following update rule:

$$\boldsymbol{A}^{(t+1)} = \boldsymbol{A}^{(t)} + 2\,\boldsymbol{M}^\top \boldsymbol{\Gamma}(\boldsymbol{X} - \boldsymbol{M}\,\boldsymbol{A}^{(t)}).$$

3.2. Sum-to-one constraint

Next, we explore two strategies to handle the sum-to-one constraint, namely $\sum_{k=1}^{K} \alpha_k^{(t+1)} = 1$ at each iteration t. In the first strategy, the system converges to this constraint by using the augmented model, while in the second one its is verified by normalization each iteration.

In the first strategy, we consider the augmented model (2), which can be easily investigated using the "reflect-then-combine" update rule. The resulting rule can be written as

$$\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)} + 2\,[\,\mathbf{1}\;\; \boldsymbol{M}^\top\,]\,\boldsymbol{\Gamma}\left(\begin{bmatrix} 1 \\ \boldsymbol{x} \end{bmatrix} - \begin{bmatrix} \mathbf{1}^\top \\ \boldsymbol{M} \end{bmatrix} \boldsymbol{\alpha}^{(t)}\right), \tag{$\mathcal{S}_{\text{augment}}$}$$

where $\boldsymbol{\Gamma}$ is here a $(L+1)$-by-$(L+1)$ diagonal matrix. In this paper, the augmented model extends naturally the "reflect-then-combine" update rule. This strategy can be viewed as

[1]The intersection is not unique, due to the presence of noise as given by $\boldsymbol{\epsilon}$ in (1). However, nothing prevents us from applying the reflection principle, since the affine hyperplane defined by $x_\ell = \widetilde{\boldsymbol{m}}_\ell^\top \boldsymbol{\alpha} + \epsilon_\ell$ is parallel to the one given in the noiseless expression (3).

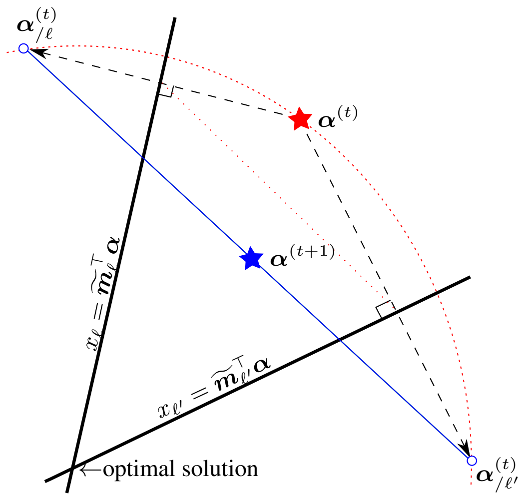

Fig. 1. Illustration in two-dimensions of the "reflect-then-combine" update rule, with two hyperplanes (black lines). The final solution $\boldsymbol{\alpha}^{(t+1)}$ (★) is obtained by a convex combination of the reflections (○) of the initial solution $\boldsymbol{\alpha}^{(t)}$ (★). As shown, any combination of the reflections (—) is a better solution that every combination of the projections (···).

including a new reflection at each iteration, with respect to the (affine) hyperplane defined by the equation $1 = \mathbf{1}^\top \boldsymbol{\alpha}^{(t)}$.

In the second strategy, we propose to operate a normalization after every adaptation, namely by applying successively the following rules at each iteration t:

$$\boldsymbol{\alpha}^{(*)} = \boldsymbol{\alpha}^{(t)} + 2\,\boldsymbol{M}^\top \boldsymbol{\Gamma}(\boldsymbol{x} - \boldsymbol{M}\,\boldsymbol{\alpha}^{(t)}),$$
$$\boldsymbol{\alpha}^{(t+1)} = \frac{\boldsymbol{\alpha}^{(*)}}{\mathbf{1}^\top \boldsymbol{\alpha}^{(*)}}, \qquad [\mathcal{S}_{\text{normalize}}]$$

where $\boldsymbol{\alpha}^{(*)}$ is the intermediate (unnormalized) solution. One can also view the above normalization step as a projection, onto the (affine) hyperplane defined by the equation $1 = \mathbf{1}^\top \boldsymbol{\alpha}^{(t)}$. But this is not an orthogonal projection, and therefore it does no longer satisfy the non-expansivity property. This means that the distances between two projected entries are not guarantied to be less than the distances of the original ones. While this may affect the convergence, the combination of the reflection and projection seems fortunately to work well.

There exists two major differences between these two strategies: On the one hand, the normalization is a non-orthogonal projection onto the hyperplane, as opposed to the orthogonal reflection with the augmented model. On the other hand, normalization guarantees the sum-to-one constraint at each iteration, as opposed to an asymptotic convergence with the previous approach.

3.3. Non-negativity constraint

Next, we derive two strategies to enforce the non-negativity constraint, namely $\alpha_1^{(t+1)}, \alpha_2^{(t+1)}, \ldots, \alpha_K^{(t+1)} \geq 0$ at each iteration t. The first one relaxes the reflections to prevent getting out of the non-negative orthant, while the second one operates projection onto it.

In the first strategy, we revisit the reflection step given in the update rule (4), by including a relaxation with

$$\boldsymbol{\alpha}^{(t)}_{/\ell} = \boldsymbol{\alpha}^{(t)} + 2\,\eta_\ell \frac{x_\ell - \widetilde{\boldsymbol{m}}_\ell^\top\, \boldsymbol{\alpha}^{(t)}}{\|\widetilde{\boldsymbol{m}}_\ell\|^2}\, \widetilde{\boldsymbol{m}}_\ell, \qquad [\mathcal{S}_{\text{relax}}]$$

where the relaxation weight $\eta_\ell \in [0\,;1]$ is chosen in order to impose the non-negativity of entries in $\boldsymbol{\alpha}^{(t)}_{/\ell}$. In practice, it is either set to 1 when the reflection satisfies the constraint; otherwise it is chosen on the boundary, namely when one of the entries of $\boldsymbol{\alpha}^{(t)}_{/\ell}$ goes to zero. In other words,

$$\eta_\ell = \min_k \frac{-\|\widetilde{\boldsymbol{m}}_\ell\|^2}{x_\ell - \widetilde{\boldsymbol{m}}_\ell^\top\, \boldsymbol{\alpha}^{(t)}} \frac{[\boldsymbol{\alpha}^{(t)}]_k}{[\widetilde{\boldsymbol{m}}_\ell]_k}$$

when the right-hand-side is within $[0\,;1]$; otherwise $\eta_\ell = 1$.

The second strategy consists of setting to zero all negative values after each reflect-and-combine. This corresponds to apply the following rules at each iteration t:

$$\boldsymbol{\alpha}^{(*)} = \boldsymbol{\alpha}^{(t)} + 2\,\boldsymbol{M}^\top \boldsymbol{\Gamma}(\boldsymbol{x} - \boldsymbol{M}\,\boldsymbol{\alpha}^{(t)}),$$
$$\boldsymbol{\alpha}^{(t+1)} = [\alpha_1^{(t+1)} \;\cdots\; \alpha_K^{(t+1)}]^\top \text{with } \alpha_k^{(t+1)} = \max\{\alpha_k^{(*)}; 0\} \qquad [\mathcal{S}_{\text{set-to-0}}]$$

where $\boldsymbol{\alpha}^{(*)}$ is the intermediate (unconstrained) solution.

A major difference between these two strategies is that the latter enforces null values to prevent negative ones. The former strategy leads to non-null values, due to the combination process.

3.4. Fully-constrained optimization

In the previous sections, we have proposed 2 strategies to handle the sum-to-one constraint, and 2 strategies to enforce the non-negativity constraint. This leads to 4 variants for the fully-constrained optimization problem. The following table summarizes the properties of these fully-constrained algorithms:

	sum-to-one	non-negativity
$\mathcal{S}_{\text{augment \& relax}}$	$\mathbf{1}^\top \boldsymbol{\alpha} \to 1$	$\boldsymbol{\alpha} > 0$
$\mathcal{S}_{\text{augment \& set-to-0}}$	$\mathbf{1}^\top \boldsymbol{\alpha} \to 1$	$\boldsymbol{\alpha} \geq 0$
$\mathcal{S}_{\text{normalize \& relax}}$	$\mathbf{1}^\top \boldsymbol{\alpha} = 1$	$\boldsymbol{\alpha} > 0$
$\mathcal{S}_{\text{normalize \& set-to-0}}$	$\mathbf{1}^\top \boldsymbol{\alpha} = 1$	$\boldsymbol{\alpha} \geq 0$

This table illustrates that the hard constraints of sum-to-one and non-negativity are replaced by soft constraints, i.e., convergence. The price to pay for fulfilling the hard constraints, as given by the $\mathcal{S}_{\text{normalize \& set-to-0}}$ strategy, is a slower slower convergence rate. This is essentially due to the normalization in $\mathcal{S}_{\text{normalize}}$, which corresponds to a non-orthogonal projection and therefore loosing the non-expansivity property.

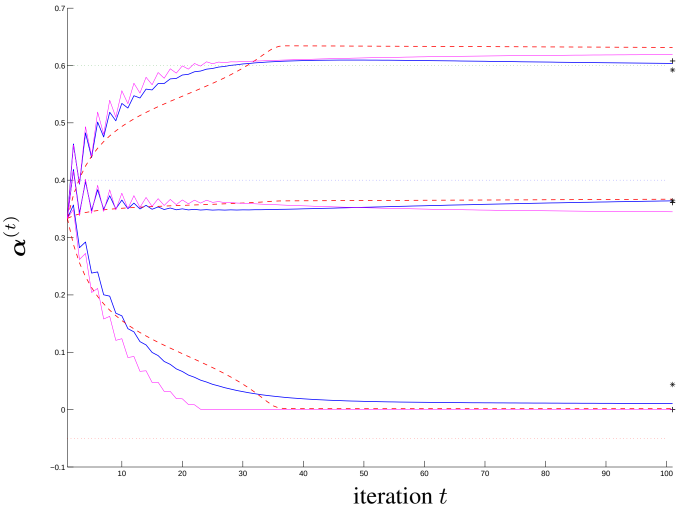

Fig. 2. Convergence of the fully-constrained optimization methods: $\mathcal{S}_{\text{augment \& relax}}$ ("—"), $\mathcal{S}_{\text{augment \& set-to-0}}$ ("—") and $\mathcal{S}_{\text{normalize \& relax}}$ ("- -") strategies. The dotted lines correspond to the real abundances $\alpha = [0.4 \;\; 0.6 \;\; -0.05]^\top$. The estimates using non-negative least-squares ($+$) and the fully-constrained least-squares ($*$) are also given.

In this paper, we consider also the soft constraints. It is worth noting that most of the work on hyperspectral unmixing problem consider relaxing the constraints. See for instance [5].

4. EXPERIMENTATIONS

To avoid the dilemma of the lack of ground truth information for comparing unmixing methods, and due to space limitation, we considered the linear combination of real hyperspectral signatures extracted from the USGC library: grass, cedar and asphalt. These spectra consist of 2151 bands covering wavelengths ranging from 0.35 to 2.5 μm. A spectral mixture was generated by a linear combination of three pure spectral signatures with fractional abundances $\alpha = [0.4 \; 0.6 \; -0.05]^\top$. The data were corrupted by an additive Gaussian noise with a SNR $\approx$ 35 dB.

The convergence of the fully-constrained optimization methods is illustrated in Figure 2 for 100 iterations. As expected, the augmented model as in $\mathcal{S}_{\text{augment}}$ is preferred to the normalization $\mathcal{S}_{\text{normalize}}$. Both the $\mathcal{S}_{\text{augment \& relax}}$ strategy ("—") and the $\mathcal{S}_{\text{augment \& set-to-0}}$ strategy ("—") converge faster than the $\mathcal{S}_{\text{normalize \& relax}}$ strategy ("- -"). The $\mathcal{S}_{\text{normalize \& set-to-0}}$ strategy was discarded due to its poor (and bad) convergence properties. For a comparative study, we also included in the figure the estimates obtained from the non-negative least-squares method ("$+$") and the fully-constrained least-squares method ("$*$"). See for instance [5].

5. CONCLUSION AND PERSPECTIVES

This work shows that conventional Cimmino's reflect-then-combine iterative scheme can be adapted for solving the constrained unmixing problem. Several strategies are studied in order to handle the sum-to-one and non-negativity constraints, and experimental results are analyzed.

As for future work, we are studying the optimal choice of the weights in the combination stage. We seek also to optimize its selection in order to satisfy both sum-to-one and non-negativity constraints. In the latter case for instance, one may allow negative values in reflections as long as there exists a non-negative convex combination. We are also conducting an analysis of the algorithm, in the same spirit as in [2]. It is also interesting to study Kaczmarz's cyclic projections, as well as random projections/reflections. It is clear that this first work opens the way to a new class of algorithms in adaptive filtering and machine learning [9, 8].

6. REFERENCES

[1] 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 J. Sel. Topics Appl. Earth Observations and Remote Sens.*, 5(2):354–379, April 2012.

[2] J. Chen, C. Richard, J. C. M. Bermudez, and P. Honeine. Non-negative least-mean-square algorithm. *IEEE Transactions on Signal Processing*, 59(11):5225–5235, November 2011.

[3] J. Chen, C. Richard, and P. Honeine. Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model. *IEEE Transactions on Signal Processing*, 61(2):480–492, January 15 2013.

[4] G. Cimmino. Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. *La Ricerca Scientifica*, II, 9:326–333, 1938.

[5] D. Heinz and C. Chang. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geoscience and Remote Sensing*, 39(3):529–545, March 2001.

[6] P. Honeine and C. Richard. Geometric unmixing of large hyperspectral images: a barycentric coordinate approach. *IEEE Transactions on Geoscience and Remote Sensing*, 50(6), June 2012.

[7] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres*, 35:355–357, 1937.

[8] S. Theodoridis, K. Slavakis, and I. Yamada. Adaptive learning in a world of projections. *IEEE Signal Processing Magazine*, 28(1):97–123, 2011.

[9] M. Yukawa. Adaptive filtering based on projection method. Lecture Notes, December 6–8, 2010.