CONSTRAINED KACZMARZ'S CYCLIC PROJECTIONS FOR UNMIXING HYPERSPECTRAL DATA

Paul Honeine

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

Henri Lantéri, Cédric Richard

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

ABSTRACT

The estimation of fractional abundances under physical constraints is a fundamental problem in hyperspectral data processing. In this paper, we propose to adapt Kaczmarz's cyclic projections to solve this problem. The main contribution of this work is two-fold: On the one hand, we show that the non-negativity and the sum-to-one constraints can be easily imposed in Kaczmarz's cyclic projections, and on the second hand, we illustrate that these constraints are advantageous in the convergence behavior of the algorithm. To this end, we derive theoretical results on the convergence performance, both in the noiseless case and in the case of noisy data. Experimental results show the relevance of the proposed method.

***Index Terms*—** Constrained optimization, Kaczmarz's cyclic projections, hyperspectral data, unmixing problem

1. INTRODUCTION

Due to low spatial resolution of hyperspectral cameras (i.e., imaging spectrometers), acquired spectra are mixtures of spectra of some pure materials. These pure components are also called endmembers. The unmixing problem consists of breaking down a spectrum into the pure spectra of endmembers and their fractional abundances. A wide range of geometrical and statistical methods has been devised for the unmixing problem. See [2] for an extensive overview. Hyperspectral data processing often suffers from a number of limitations, which includes computation complexity and the ease of implementation, making iterative and parallel algorithms suitable for implementation [7, 1].

Assuming that the endmembers were extracted using any off-the-shelf technique, the problem of estimating the fractional abundances provides new opportunities and challenges to both linear [9] as well as non-linear unmixing problems [4]. In order to have a coherent physical interpretation, two hard constraints need to be satisfied in the estimation problem: the sum-to-one of all the fractional abundances and there non-negativities. In this work, we study the problem of estimating constrained fractional abundances, by suitably adapting a cyclic orthogonal projection scheme.

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

Methods of orthogonal projections onto spaces, and more recently onto convex sets, have been successfully applied for solving many optimization problems, in the field of signal processing in particular. These methods include the celebrated Cimmino's parallel projections [5] and Kaczmarz's cyclic projections [10]. More recently, projections onto spaces have been generalized to projections onto convex sets. These methods provide a unified framework for tackling many signal processing problems, including adaptive filtering and machine learning [14, 13]. To the best of our knowledge, these methods have not yet been exploited in hyperspectral image processing. Very recently, we have successfully adapted Cimmino's parallel projections [5] for the hyperspectral unmixing problem [8].

Coupled with these advances, this paper shows that the spectral unmixing problem can take advantage of these developments. To this end, we propose to adapt Kaczmarz's cyclic projection for estimating the fractional abundances. Appropriate strategies for enforcing the non-negativity and sum-to-one constraints are advanced as a natural extension of the conventional unconstrained Kaczmarz's method: the sum-to-one constraint is enforced by a projection scheme, while the non-negativity is satisfied by relaxing the projections.

2. THE LINEAR UNMIXING 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

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

where $\boldsymbol{M}$ is the L-by-K matrix of the K spectral signatures of a pure material, i.e., endmember, and $\boldsymbol{\epsilon}$ is the vector of fitness error, and $\boldsymbol{\alpha}$ is the vector of K abundances to be determined. In this work, we assume that the endmembers have been properly identified, using a supervised or an unsupervised endmember extraction technique. The paper [2] provides an extensive overview of existing techniques and state-of-the-art methods.

To adopt a physical interpretation of the unmixing problem, the linear combination in model (1) needs to be convex. In other words, the following constraints are required:

- The sum-to-one of the contributions, $\mathbf{1}^\top \boldsymbol{\alpha} = 1$, where $\mathbf{1}$ denotes the unit column-vector of K entries;
- and the non-negativity of the contributions, $\boldsymbol{\alpha} \geq \mathbf{0}$, where the inequality is taken component-wise.

An iterative scheme is required to solve the constrained optimization problem: $\arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{M}\boldsymbol{\alpha} - \boldsymbol{x}\|^2$, under the above constraints. See for instance [6].

3. CONSTRAINED KACZMARZ'S CYCLIC PROJECTIONS

To revisit the unconstrained Kaczmarz's method, we consider the noiseless model, defined by

$$\boldsymbol{M}\boldsymbol{\alpha} = \boldsymbol{x}. \tag{2}$$

In other words, Kaczmarz's method assumes that $\boldsymbol{x}$ belongs to the range of the matrix $\boldsymbol{M}$. The impact of the presence of noise, as given in model (1), is studied in next Section.

3.1. Unconstrained Kaczmarz's cyclic projections

The linear system (2) of L equations and K unknowns gives rise to a set of L affine hyperplanes,

$$\mathcal{H}_\ell = \left\{ \boldsymbol{\alpha} \,\middle|\, \boldsymbol{m}_\ell^\top \boldsymbol{\alpha} = x_\ell \right\}. \tag{3}$$

In this expression, $\boldsymbol{m}_\ell^\top$ denotes the ℓ-th row of $\boldsymbol{M}$, namely the spectral signatures of the endmembers at the ℓ-th wavelength band. The solution of the linear system (2) is the intersection of the L affine hyperplanes, i.e., $\mathcal{H}_1 \cap \mathcal{H}_2 \cap \ldots \cap \mathcal{H}_L$, where the intersection is a unique *point* in the noiseless case. It is worth noting that $\boldsymbol{M}$ is full column rank by construction, since non of the endmembers can be written as a linear combination of other endmembers.

The conventional Kaczmarz's method sweeps through the affine hyperplanes, by projecting orthogonally onto each and taking this as the next iterate. In a mathematical form, at the t-th iteration, the projection of $\boldsymbol{\alpha}_{t-1}$ onto the affine hyperplane $\mathcal{H}_\ell$ reads

$$\boldsymbol{\alpha}_{t-1\perp\ell} = \boldsymbol{\alpha}_{t-1} + \frac{x_\ell - \boldsymbol{m}_\ell^\top \boldsymbol{\alpha}_{t-1}}{\|\boldsymbol{m}_\ell\|^2}\, \boldsymbol{m}_\ell, \tag{4}$$

with $\boldsymbol{\alpha}_0$ some initial guess. This process is iterated until a given convergence criterion is satisfied. In this expression, ℓ is selected at each iteration t, i.e., a slight abuse of notation is made in using ℓ rather than $\ell(t)$.

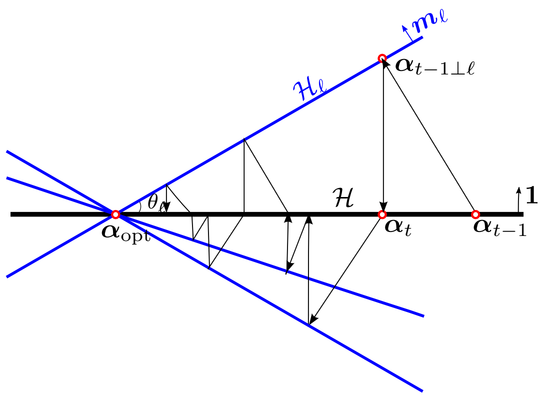

Fig. 1. Illustration in two-dimensions of the constrained Kaczmarz's cyclic projections. The affine hyperplane of the sum-to-one constraint is $\mathcal{H}$ (black line), i.e., the hyperplane defined by the normal vector $\mathbf{1}$.

Several strategies [11, 12, 3] have been proposed to choose the sequence of selected affine hyperplanes, i.e., the choice at iteration t of the ℓ-th row in the matrix $\boldsymbol{M}$. The classical cyclic selection, i.e., $\ell = t \mod L$, is the simplest one; however it could lead to low convergence. Other strategies have been proposed to overcome this drawback, the most known are the maximum error selection and the randomization. The maximum error selection requires the evaluation of the numerator in (4) for all ℓ at each iteration t, which leads to a high computational burden. In randomization ℓ is chosen randomly with probability proportional to the denominator in (4), i.e., $\|\boldsymbol{m}_\ell\|^2$. This strategy has shown very interesting properties [12], but is also criticized in [3].

This paper shows that, in the case of a constrained solution as in the unmixing problem, Kaczmarz's method does not suffer from the problem of appropriate selection of the affine hyperplanes. Moreover, we provide theoretical results on the convergence, for the noiseless and the noisy cases.

3.2. Sum-to-one constraint

To impose the sum-to-one constraint, we consider a two-step strategy at each iteration t: a projection of $\boldsymbol{\alpha}_{t-1}$ onto the affine hyperspace $\mathcal{H}_\ell$, yielding $\boldsymbol{\alpha}_{t-1\perp\ell}$, followed by a normalization to satisfy the constraint, yielding $\boldsymbol{\alpha}_t$. The proposed strategy is illustrated in Figure 1.

Let $\mathcal{H}$ be the affine hyperplane defined by the sum-to-one, namely $\mathcal{H} = \left\{\boldsymbol{\alpha} \mid \mathbf{1}^\top \boldsymbol{\alpha} = 1\right\}$. By analogy with the expression of the projection (4), the normalization of any $\boldsymbol{\alpha}_{t-1\perp\ell}$ is given by

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1\perp\ell} + \tfrac{1}{K}\left(1 - \mathbf{1}^\top \boldsymbol{\alpha}_{t-1\perp\ell}\right)\mathbf{1}, \tag{5}$$

where $\boldsymbol{m}_\ell$ is now substituted with $\mathbf{1}$, the unit column-vector of K ones. By combining this projection (5) with the projection onto $\mathcal{H}_\ell$, as given in (4), at iteration t, we get:

$$\begin{aligned}\boldsymbol{\alpha}_t &= \boldsymbol{\alpha}_{t-1\perp\ell} + \tfrac{1}{K}\left(1 - \mathbf{1}^\top \boldsymbol{\alpha}_{t-1+\perp\ell}\right)\mathbf{1} \\ &= \boldsymbol{\alpha}_{t-1} + \frac{x_\ell - \boldsymbol{m}_\ell^\top \boldsymbol{\alpha}_{t-1}}{\|\boldsymbol{m}_\ell\|^2}\,\boldsymbol{m}_\ell \\ &\quad + \tfrac{1}{K}\left(1 - \mathbf{1}^\top\boldsymbol{\alpha}_{t-1} - \frac{x_\ell - \boldsymbol{m}_\ell^\top \boldsymbol{\alpha}_{t-1}}{\|\boldsymbol{m}_\ell\|^2}\,\mathbf{1}^\top\boldsymbol{m}_\ell\right)\mathbf{1}.\end{aligned}$$

Since we have the identity $\mathbf{1}^\top\boldsymbol{\alpha}_{t-1} = 1$ from the previous iteration, we obtain the following update rule:

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \frac{x_\ell - \boldsymbol{m}_\ell^\top \boldsymbol{\alpha}_{t-1}}{\|\boldsymbol{m}_\ell\|^2}\,\left(\mathbf{I} - \tfrac{1}{K}\mathbf{1}\mathbf{1}^\top\right)\boldsymbol{m}_\ell. \quad (6)$$

Here, $\mathbf{I}$ denotes the matrix identity of size K-by-K, and the matrix $\mathbf{I} - \frac{1}{K}\mathbf{1}\mathbf{1}^\top$ is often called centering matrix in the machine learning literature.

It is worth noting that there is another way to satisfy the sum-to-one constraint, by normalizing using the form:

$$\boldsymbol{\alpha}_t = \frac{\boldsymbol{\alpha}_{t-1\perp\ell}}{\mathbf{1}^\top\boldsymbol{\alpha}_{t-1\perp\ell}}.$$

This normalization can also be viewed as a non-orthogonal projection, and therefore looses the property of nonexpansive transformation. Technically, this means that the distance between two projections may exceed the distance between the two original elements, and therefore compromising the speed of convergence. For these reasons, this normalization is not considered in this paper, as we adopt the (orthogonal) projection given in (5).

3.3. Imposing also the non-negativity constraint

To enforce the non-negativity constraint, namely $\boldsymbol{\alpha}_t \geq 0$ at each iteration, we propose to relax the projection in (6), by considering the following update rule:

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \eta_t\,\frac{x_\ell - \boldsymbol{m}_\ell^\top \boldsymbol{\alpha}_{t-1}}{\|\boldsymbol{m}_\ell\|^2}\,\left(\mathbf{I} - \tfrac{1}{K}\mathbf{1}\mathbf{1}^\top\right)\boldsymbol{m}_\ell. \quad (7)$$

The relaxation parameter $\eta_t \in [0\,;1]$ is chosen in order to impose the non-negativity of entries in $\boldsymbol{\alpha}_t$.

To this end, let

$$\eta_t = \min_{i=1,\ldots,K} \eta_t^{(i)},$$

where $\eta_t^{(i)}$ is a valid stepsize in the i-th direction. The value of the latter depends on one of the two cases:

- If $\frac{x_\ell - \boldsymbol{m}_\ell^\top \boldsymbol{\alpha}_{t-1}}{\|\boldsymbol{m}_\ell\|^2}\left[\left(\mathbf{I} - \frac{1}{K}\mathbf{1}\mathbf{1}^\top\right)\boldsymbol{m}_\ell\right]_i > 0$, there is no constraint on the stepsize, and therefore could be set to $\eta_t^{(i)} = 1$;
- Otherwise, the stepsize should be reduced such that

$$\eta_t^{(i)} \leq \frac{[\boldsymbol{\alpha}_{t-1}]_i}{\left[\left(\mathbf{I} - \frac{1}{K}\mathbf{1}\mathbf{1}^\top\right)\boldsymbol{m}_\ell\right]_i}.$$

In order to account for noise, we consider always a relaxation of the projection. The stepsize is set to a smaller value than one (e.g., 0.1 or below depending on the above relations), as recommended in adaptive filtering literature. Essentially, the update rule (7) is a normalized least mean squares (NLMS) algorithm applied on centered data.

Other strategies to enforce the non-negativity constraint can also be applied, such as a post-processing by replacing every negative entry in $\boldsymbol{\alpha}_t$ with zero. However, this strategy of projection onto the positive half-space may destroy the sum-to-one constraint, and therefore it is not recommended.

4. THEORETICAL RESULTS

In this section, we provide theoretical results, namely on the convergence rate in both the noiseless (2) and the noisy cases (1). To this end, we consider the influence of both sum-to-one and non-negativity constraints by using the update rule (7).

First of all, the impact of the non-negativity constraint is clear. All the results are enforced to be on the non-negative orthant in the K-dimensional space. Furthermore, the studied hyperspectral unmixing problem has the following interesting property: Since $\boldsymbol{m}_\ell$ is a vector of values from spectral signatures at the ℓ-th wavelength band, we have $\boldsymbol{m}_\ell \geq 0$, and therefore all affine hyperplanes defined by (3) have non-negative normal vectors. Moreover, the sum-to-one affine hyperplane $\mathcal{H}$ is the bisector of the orthant, and therefore the angle between it and any $\mathcal{H}_\ell$ cannot exceed $\pi/4$. Let θ_ℓ be the angle between the two affine hyperplanes $\mathcal{H}_\ell$ and $\mathcal{H}$, then $\cos(\theta_\ell) > 1/\sqrt{2}$, for all $\ell = 1, 2, \ldots, L$.

Let $\boldsymbol{\alpha}_{\text{opt}}$ be the (unknown) optimal solution of the noiseless problem (2).

4.1. The noiseless case

The following theorem provides the rate of convergence for the proposed constrained Kaczmarz' method, in the case of the noiseless model (2). To this end, we consider the influence of the sum-to-one constraint on the convergence.

Theorem 1 (Convergence in the noiseless case)**.** *Starting from an initial guess $\boldsymbol{\alpha}_0$, the algorithm converges to the optimal solution $\boldsymbol{\alpha}_{\text{opt}}$ at the rate*

$$\|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{\text{opt}}\| = \|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_{\text{opt}}\| \prod_{\substack{\ell(t')\\ t'\leq t}} \cos^2(\theta_\ell),$$

where θ_ℓ is the angle between $\mathcal{H}_\ell$ and $\mathcal{H}$, and the product is taken over all visited affine hyperplanes until iteration t.

Proof. As in the conventional Kaczmarz's method, it is assumed that the solution of the linear system exists, is unique, and belongs to the set $\mathcal{H}_\ell$. On the one hand, we have from projection (i.e., according to the Pythagorean theorem)

$$\cos(\theta_\ell) = \frac{\|\boldsymbol{\alpha}_{t-1\perp\ell} - \boldsymbol{\alpha}_{\text{opt}}\|}{\|\boldsymbol{\alpha}_{t-1} - \boldsymbol{\alpha}_{\text{opt}}\|},$$

where $\cos(\theta_\ell)$ is positive as discussed above. On the other hand, we also have

$$\cos(\theta_\ell) = \frac{\|\boldsymbol{\alpha}_{t} - \boldsymbol{\alpha}_{\text{opt}}\|}{\|\boldsymbol{\alpha}_{t-1\perp\ell} - \boldsymbol{\alpha}_{\text{opt}}\|}.$$

By combining both expressions, we get

$$\begin{aligned} \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{\text{opt}}\| &= \|\boldsymbol{\alpha}_{t-1\perp\ell} - \boldsymbol{\alpha}_{\text{opt}}\| \cos(\theta_\ell) \\ &= \|\boldsymbol{\alpha}_{t-1} - \boldsymbol{\alpha}_{\text{opt}}\| \cos^2(\theta_\ell) \\ &\vdots \\ &= \|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_{\text{opt}}\| \prod_{\substack{\ell(t') \\ t' \leq t}} \cos^2(\theta_\ell). \end{aligned}$$

This concludes the proof. ☐

This theorem demonstrates the rate of convergence of the proposed algorithm. It is worth noting that $\cos(\theta_\ell)$ can be easily evaluated, since $\cos(\theta_\ell) = \frac{\mathbf{1}^\top \boldsymbol{m}_\ell}{\|\mathbf{1}\|\|\boldsymbol{m}_\ell\|}$, where the numerator is the sum of the entries of $\boldsymbol{m}_\ell$, and the denominator involves the sum of their squares.

Moreover, each iteration in the proposed algorithm is essentially a double projection. This alternating projection can be applied infinitely on a fixed ℓ, i.e., affine hyperplane $\mathcal{H}_\ell$, without the need to sweep through all the affine hyperplanes. In this case, the Theorem 1 becomes for a fixed ℓ: $\|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{\text{opt}}\| = \|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_{\text{opt}}\| \left(\cos(\theta_\ell)\right)^{2t}$. In practice, due to the presence of noise, it is not enough to use a single ℓ, but rather to sweep over all the hyperplanes. The next Section provides a study of the impact of noise on the proposed algorithm.

4.2. The noisy case

Very few works consider the influence of noise in the conventional (unconstrained) Kaczmarz's cyclic projections. The presence of noise is seldom considered; see [11] for the randomized Kaczmarz's method. In this section, we provide a theoretical analysis of the presence of an error in the proposed constrained optimization problem. We assume that the endmember spectra have been well estimated, and the noise is on the investigated spectrum as given in (1), namely $\boldsymbol{M}\boldsymbol{\alpha} = \boldsymbol{x} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = [\epsilon_1 \; \epsilon_2 \; \ldots \; \epsilon_L]^\top$ is the error vector.

The solution of this noisy problem belongs to the intersection of the L affine hyperplanes $\widetilde{\mathcal{H}}_1 \cap \widetilde{\mathcal{H}}_2 \cap \ldots \cap \widetilde{\mathcal{H}}_L$, defined by

$$\widetilde{\mathcal{H}}_\ell = \left\{ \boldsymbol{\alpha} \,\middle|\, \boldsymbol{m}_\ell^\top \boldsymbol{\alpha} = x_\ell + \epsilon_\ell \right\}.$$

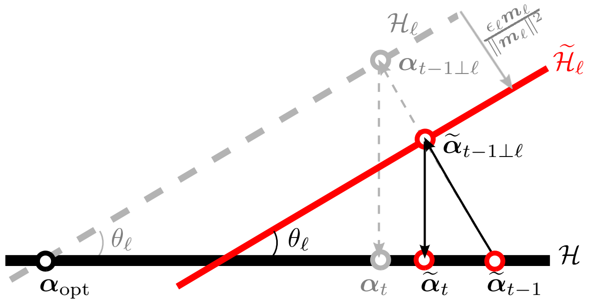

Fig. 2. Illustration in two-dimensions of the presence of noise. In this case, $\mathcal{H}_\ell$ is not available (neither all gray elements), but the affine hyperplane $\widetilde{\mathcal{H}}_\ell$ associated to the noisy data. Therefore, at instant $t-1$, the approximation $\widetilde{\boldsymbol{\alpha}}_{t-1}$ leads to $\widetilde{\boldsymbol{\alpha}}_t$, rather than $\boldsymbol{\alpha}_t$.

It is easy to see that these affine hyperplanes can also be defined from the hyperplanes associated for the noiseless case, with (See for instance [11, Lemma 2.2]):

$$\widetilde{\mathcal{H}}_\ell = \left\{ \boldsymbol{\alpha} + \tfrac{\epsilon_\ell}{\|\boldsymbol{m}_\ell\|^2} \boldsymbol{m}_\ell \,\middle|\, \boldsymbol{\alpha} \in \mathcal{H}_\ell \right\}. \tag{8}$$

Let $\widetilde{\boldsymbol{\alpha}}_{t-1}$ be the estimate obtained at iteration t in presence of noise, and $\widetilde{\boldsymbol{\alpha}}_{t-1\perp\ell}$ its projection onto $\widetilde{\mathcal{H}}_\ell$. The following theorem is a generalization of Theorem 1 to the case of noisy data. See Figure 2 for an illustration.

Theorem 2 (Convergence in the noisy case)**.** *Starting from an initial guess $\widetilde{\boldsymbol{\alpha}}_0$, the algorithm converges at the rate*

$$\begin{aligned} \|\widetilde{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_{\text{opt}}\|^2 \;=\; & \|\widetilde{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_{\text{opt}}\|^2 \prod_{\substack{\ell(t') \\ t' \leq t}} \cos^4(\theta_\ell) \\ & + \sum_{\substack{\ell(t') \\ t' \leq t}} \frac{\epsilon_\ell^2}{\|\boldsymbol{m}_\ell\|^2} \cos^2(\theta_\ell) \prod_{\substack{\ell(t'') \\ t'' < t'}} \cos^4(\theta_{\ell''}), \end{aligned}$$

where θ_ℓ is the angle between $\mathcal{H}$ and $\mathcal{H}_\ell$ (as well as $\widetilde{\mathcal{H}}_\ell$).

Proof. First, we have from projection the following relation: $\cos(\theta_\ell) = \|\widetilde{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_{\text{opt}}\| / \|\widetilde{\boldsymbol{\alpha}}_{t-1\perp\ell} - \boldsymbol{\alpha}_{\text{opt}}\|$. By applying the Pythagorean theorem, the above denominator can be decomposed into

$$\|\widetilde{\boldsymbol{\alpha}}_{t-1\perp\ell} - \boldsymbol{\alpha}_{\text{opt}}\|^2 = \|\boldsymbol{\alpha}_{t-1\perp\ell} - \boldsymbol{\alpha}_{\text{opt}}\|^2 + \|\tfrac{\epsilon_\ell}{\|\boldsymbol{m}_\ell\|^2} \boldsymbol{m}_\ell\|^2,$$

where the orthogonality and the definition (8) are used. The second term in the right-hand-side is simply $\epsilon_\ell^2 / \|\boldsymbol{m}_\ell\|^2$. The first term in the right-hand-side is (see the proof of Theorem 1): $\|\boldsymbol{\alpha}_{t-1\perp\ell} - \boldsymbol{\alpha}_{\text{opt}}\| = \|\widetilde{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_{\text{opt}}\| \cos(\theta_\ell)$. By combining these expressions, we get

$$\begin{aligned}\|\widetilde{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_{\text{opt}}\|^2 =& \|\widetilde{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_{\text{opt}}\|^2 \cos^4(\theta_\ell) + \frac{\epsilon_\ell^2}{\|\boldsymbol{m}_\ell\|^2} \cos^2(\theta_\ell) \\ &\vdots \\ =& \|\widetilde{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_{\text{opt}}\|^2 \prod_{\substack{\ell(t') \\ t' \leq t}} \cos^4(\theta_\ell) \\ &+ \sum_{\substack{\ell(t') \\ t' \leq t}} \frac{\epsilon_\ell^2}{\|\boldsymbol{m}_\ell\|^2} \cos^2(\theta_\ell) \prod_{\substack{\ell(t'') \\ t'' < t'}} \cos^4(\theta_{\ell'}).\end{aligned}$$

This concludes the proof. □

This theorem demonstrates the impact of the selection, i.e., $\ell(t)$ at each iteration t, on the approximation error. In fact, thanks to the geometric progression of the cosines, the hyperplanes considered in the first iterations have their errors highly reduced, as opposed to the last projections with less "cosine" weights. In other words, it is clear that the bandwidths in the hyperspectral data with the largest errors, i.e., large values of $\epsilon_\ell^2/\|\boldsymbol{m}_\ell\|^2$, should be used in the first iterations. Therefore, this theorem provides a new selection criterion for the cyclic projections, optimal for the studied constrained optimization problem.

5. EXPERIMENTATIONS

The proposed learning algorithm was used for hyperspectral data unmixing, generated by a linear combination of three pure materials with abundance $\alpha = [0.4 \;\; 0.65 \;\; -0.05]^\top$, where we injected a sign error on the third abundance. These materials are grass, cedar and asphalt, with spectral signatures extracted from the USGC library. These spectra consist of 2151 bands covering wavelengths ranging from 0.35 to 2.5 μm. The data were corrupted by an additive Gaussian noise. Figure 3 illustrates the convergence of the proposed algorithm, for a SNR of $\approx$ 25 dB and $\approx$ 18 dB, with the largest value of the stepsize fixed to 0.1 and a single sweep over all the bandwidths. For a comparative results, we also show in the same figure the estimates using non-negative least-squares and the fully-constrained least-squares techniques [6].

6. CONCLUSION AND PERSPECTIVES

In this paper, we presented the development of a new class of unmixing methods, by providing constrained counterpart of projection-based algorithms [14, 13]. New theoretical results were derived for the studied constrained optimization problem, including the impact of noise on the analyzed spectrum. Experimentations collaborate these results. As for future work, we seek to study the impact of noise on the endmembers. We are also conducting a statistical analysis on the impact of noise, as well as connections with adaptive filtering.

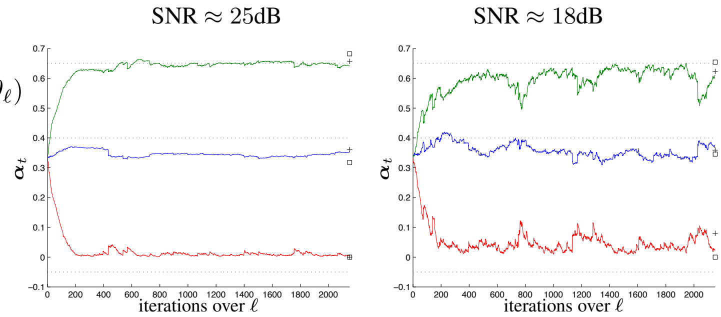

Fig. 3. Convergence of the constrained Kaczmarz's optimization method, for a mixture of two spectra with additive noise. The dotted lines correspond to the real abundances $\boldsymbol{\alpha} = [0.4 \;\; 0.65 \;\; -0.05]^\top$. The estimates using non-negative least-squares (given by "+") and the fully-constrained least-squares (given by "□") are also given.

7. REFERENCES

[1] J. Bioucas-Dias and M. Figueiredo. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In *Proc. 2nd IEEE Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS)*, Reykjavik, Iceland, 2010.

[2] 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.

[3] Y. Censor, G. T. Herman, and M. Jiang. A note on the behavior of the randomized Kaczmarz algorithm of Strohmer and Vershynin. *J. Fourier Anal. Appl.*, 15(4):431–436, 2009.

[4] 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.

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

[6] 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.

[7] R. Heylen, D. Burazerovic, and P. Scheunders. Fully constrained least-squares spectral unmixing by simplex projection. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4112–4122, 2011.

[8] P. Honeine and H. Lantéri. Constrained reflect-then-combine methods for unmixing hyperspectral data. In *Proc. IEEE Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS)*, Gainesville, Florida, USA, 25 - 28 June 2013.

[9] 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.

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

[11] D. Needell. Randomized Kaczmarz solver for noisy linear systems. *BIT Numerical Mathematics*, 50(2):395–403, June 2010.

[12] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15(2):262–278, 2009.

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

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