Iterative Algorithms for Unmixing of Hyperspectral Imagery

Miguel Vélez-Reyes[1a], Angela Puetz[2], P. Hoke[2],
Ronald B. Lockwood[2], and Samuel Rosario[1]
[1]Laboratory for Applied Remote Sensing and Image Processing,
University of Puerto Rico at Mayagüez, P.O. Box 9042, Mayagüez, PR 00681-9042,
[2]Air Force Research Laboratory, 29 Randolph Rd., Hanscom AFB, MA 01731-3010

ABSTRACT

This paper addresses the use of multiplicative iterative algorithms to compute the abundances in unmixing of hyperspectral pixels. The advantage of iterative over direct methods is that they allow incorporation of positivity and sum-to-one constraints of the abundances in an easy fashion while also allowing better regularization of the solution for the ill-conditioned case. The derivation of two iterative algorithms based on minimization of least squares and Kulback-Leibler distances are presented. The resulting algorithms are the same as the ISRA and EMML algorithms presented in the emission tomography literature respectively. We show that the ISRA algorithm and not the EMML algorithm computes the maximum likelihood estimate of the abundances under Gaussian assumptions while the EMML algorithm computes a minimum distance solution based on the Kulback-Leibler generalized distance. In emission tomography, the EMML computes the maximum likelihood estimate of the reconstructed image. We also show that, since the unmixing problem is in general overconstrained and has no solutions, acceleration techniques for the EMML algorithm such as the RBI-EMML will not converge.

Keywords: Unmixing, hyperspectral data, iterative algorithms, abundance estimation

1 INTRODUCTION

Hyperspectral sensors collect hundreds of narrow and contiguously spaced spectral bands of data organized in the so-called hyperspectral cube. Hyperspectral imagery provides fully registered spatial and high-resolution spectral information that is invaluable in discriminating between objects since they have unique spectral signatures that are captured by the data. The spatial resolution of most HSI flown nowadays is larger than the size of the objects being observed; in this case the high spectral sensor resolution can be used to detect and classify subpixel objects by their contribution to the measured spectral signal. The problem of interest is to decompose the measured reflectance (or radiance) into its basic elements. This is the so-called unmixing problem [1] in HSI. Spectral unmixing is the procedure by which the measured spectrum of a pixel is decomposed into a collection of constituent spectra, or endmembers, and a set of corresponding fractions or abundances

This paper addresses the use of multiplicative iterative algorithms to un-mix hyperspectral pixels. The advantage of iterative over direct methods is that they allow incorporation of positivity and sum-to-one constraints in an easy fashion while also allowing better regularization of the solution for the ill-conditioned case. Issues associated with iterative algorithms deal with convergence to the right solution and the actual speed of convergence. For iterative algorithms to be a feasible alternative to process HSI data it is important that they have fast convergence rate with low computational requirements.

The recent paper [2] considered the use of the so-called EMML and RBI-EMML algorithms, developed for the maximum likelihood reconstruction of emission tomography images [3, 4], to solve the unmixing problem. These algorithms have the advantage of producing abundance estimates that meet positivity as well as, with proper scaling, sum-to-one constraints that the abundances need to meet. The main focus of that work was the RBI-EMML [4] algorithm because of its significantly faster convergence rate when compared to the original EMML algorithm [3]. Simulation results presented in [2] showed that the RBI-EMML took about three orders of magnitude fewer iterations (from hundreds of thousand to hundreds) than EMML for the same problem. The derivation of the EMML approach for

[a] E-mail : m.velez@ieee.org, Tel. 787-832-4040 Ext. 2888, FAX 787-832-2485

unmixing presented in [2] is based on a manipulation of Bayes' theorem, which was originally proposed in [5]. Here, we take a different approach by looking at the unmixing problem as a constrained prediction error minimization problem where the error between the measured spectra and its prediction by the endmembers is minimized by minimizing a distance (or generalized distance) metric. We show that the EMML algorithm of [3] is an iterative algorithm to minimize the prediction error using Kullback-Leibler (KL) generalized distance. Furthermore, we also show that maximum likelihood estimation of the abundances leads to a constrained least squares (LS) estimation problem for which algorithms similar to the EMML can be derived. In particular, we present an EMML like iterative estimation algorithm for the abundances called in the emission tomography literature the image space reconstruction algorithm (ISRA). The ISRA algorithm was originally introduced in [6]. It turns out, that both of these algorithms are examples of multiplicative iterative algorithms which are used in many applications such as convex optimization problems [7], solution of symmetric linear complementary problems [8], determining positive solutions to nonlinear equations [9], and signal reconstruction problems with positivity constraints [10]. In the course of this report, we will use the name EMML to refer to the algorithm of [3] and its variants although for the unmixing problem the correct maximum likelihood estimator is the ISRA algorithm (which we will continue to call ISRA).

Section 2 of the paper reviews the linear mixing model for measured spectra in HSI. Section 3 discusses the abundance estimation problem as prediction error and maximum likelihood problems, and derives the basic EMML and ISRA algorithms. Convergence properties of the basic algorithms are discussed in Section 4 and it is shown there that slow convergence is due to the fact that these algorithms are gradient type of algorithms that are known for slow convergence. Acceleration methods are discussed in Section 5. Conclusions are given in Section 6.

2 MIXING MODEL

Analytical models for the mixing of materials provide the basis for the development of techniques to recover estimates of the constituents and their proportions from mixed pixels. In the typical mixing process, the surface is portrayed as a checkerboard mixture. The incident light interacts with the surface and the received light conveys the characteristics of the media in a proportion equal to the area covered by the endmembers and the reflectivity of the media. The typical model for such interaction is given by [1]

$$\mathbf{b} = \sum_{i=1}^{n} x_i \bar{\mathbf{a}}_i + \mathbf{w} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{b} \in \Re_{+}^{m}$ is the pixel of interest, $\bar{\mathbf{a}}_i$ is the spectral signature of the i-th endmember, x_i is the corresponding fractional abundance, $\mathbf{w}$ is the measurement noise, m is the number of spectral channel, and n is the number of endmembers. The matrix $\mathbf{A} \in \Re_{+}^{m \times n}$ is the matrix of endmembers and $\mathbf{x} \in \Re_{+}^{n}$ is the vector of spectral abundances. Notice that all elements of **b**, **A**, and **x** are constrained to be positive and $\sum_{i=1}^{n} x_i = 1$. Typically m>n so we are dealing with an overconstrained linear system of equations. This is called in the literature [1] the linear mixing model (LMM). There are other potential models that are discussed in [1] and references therein.

3 PARAMETER ESTIMATION

For most applications, the measurement noise **w** in (1) is assumed to be independent identically distributed white Gaussian noise. That is

$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

where **I** is the m×m identity matrix and σ^2 is the noise variance. The maximum likelihood estimate of **x** based on **b** is then given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \qquad \text{subject to } \mathbf{x} \geq 0 \text{ and } \sum_{i=1}^{n} x_i = 1 \tag{2}$$

where $\hat{\mathbf{x}}$ is the estimate of **x**, and $\| \ \|_2$ is the Euclidean norm.

Direct solutions to the unconstrained problem and the sum-to-one constrained problem are described in the literature see for instance [11]. For the non-negative least squares (NNLS) problem, only iterative methods can be used. The most used algorithm for the NNLS is an active set strategy described in [11]. Later on we will present multiplicative iterative algorithms to solve this problem.

We could have arrived at LS minimization by plain curve fitting with no statistical considerations. This deterministic formulation allows us to look at the problem of estimating **x** as a problem of minimizing a measure of the difference between the measured pixel and the prediction of the pixel value based on the selected endmembers.

In this context, the abundance estimation can be reformulated as follows

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} D(\mathbf{b}, \mathbf{A}\mathbf{x}) \quad \text{subject to } \mathbf{x} \geq 0 \text{ and } \sum_{i=1}^{n} x_i = 1 \tag{3}$$

where D(•,•) is a distance (or generalized distance) function. Different distance measures lead to different optimization problems with different properties from the point of view of algorithm development and implementation.

A distance function that will lead to the unmixing algorithm presented by [2] is the Kullback-Leibler distance function or Cross-Entropy, derived from Shannon's Entropy using the Bregman function formalism [12]. The reader is referred to [12] for more information on Bregman's functions and distances. The Kullback-Leibler distance between two non-negative vectors is given by

$$KL(\mathbf{b}, \mathbf{c}) = \sum_{i=1}^{m} \left(b_i \log\left(\frac{b_i}{c_i}\right) + c_i - b_i \right)$$

Notice that KL is not a distance in the usual sense since it does not obey the triangular inequality nor is it symmetric. A useful property for KL (and LS as well) is that it is a convex function. Based on this distance we can re-write (3) as follows

$$\hat{\mathbf{x}} = \arg\min_{\boldsymbol{\alpha}} KL(\mathbf{b}, \mathbf{A}\mathbf{x}) \quad \text{subject to } \mathbf{x} \geq 0 \text{ and } \sum_{i=1}^{n} x_i = 1 \tag{4}$$

In the rest of this presentation, we will ignore the sum to one constraint. It turns out that the KL cost function in equation (4) is the negative of the log likelihood function for the Poisson statistical model used in emission tomography (ET) [3].

3.1 The EMML Algorithm

The EMML algorithm was introduced by [3] for image reconstruction in ET. This algorithm is an iterative procedure to minimize (4). In [2], it was suggested to use this algorithm to tackle the spectral unmixing problem. Here, we will re-derive the algorithm as an algorithm for the distance minimization problem (4). Notice that, under the assumption of white Gaussian noise, the "correct" maximum likelihood estimation approach for (1) is given by (2).

Let the Lagrangian function for (4) be given by

$$L(\mathbf{x}, \boldsymbol{\mu}) = KL(\mathbf{b}, \mathbf{A}\mathbf{x}) - \boldsymbol{\mu}^T \mathbf{x}$$

The necessary and sufficient conditions (because of convexity of the KL distance and $\Re_+^n$) for optimality are derived in [13] and given next. If **x** is a minimum of (4), then there exists $\boldsymbol{\mu} \geq 0$ such that

$$\frac{\partial L(\mathbf{x}, \boldsymbol{\mu})}{\partial x_j} = \sum_{i=1}^{m} a_{ij} \left(1 - \frac{b_i}{\mathbf{a}_i^T \mathbf{x}} \right) - \mu_j = 0$$

$$\mu_j x_j = 0$$

where j=1,2, …,n and $\boldsymbol{\mu} \geq 0$. Solving for $\boldsymbol{\mu}$ results in

$$x_j \sum_{i=1}^{m} \left(1 - \frac{b_i}{\mathbf{a}_i^T \mathbf{x}}\right) a_{ij} = 0 \tag{5}$$

From this equation, we can see that if $x_j=0$, $\mu_j \geq 0$ and if $x_j>0$ then $\mu_j=0$. The EMML equation is obtained by solving for x_j in (4)

$$x_j = x_j \frac{\sum_{i=1}^{m} \frac{b_i}{\mathbf{a}_i^T \mathbf{x}} a_{ij}}{\sum_{i=1}^{m} a_{ij}} \tag{6}$$

The iterative EMML algorithm is given by

$$\hat{x}_j^{k+1} = \hat{x}_j^k \frac{\sum_{i=1}^{m} \frac{b_i}{\mathbf{a}_i^T \hat{\mathbf{x}}^k} a_{ij}}{\sum_{i=1}^{m} a_{ij}} \tag{7}$$

where $\hat{x}_j^k$ is the estimate of the abundance of the j-th endmember at the k-th iteration. Notice that this is a multiplicative iterative algorithm [7,10,8,9] since the next iterate is the current iterate multiplied by a scaling factor. Furthermore notice, due to the positiveness of **A** and **b**, that if the initial estimate is positive all iterates remain nonnegative and once a component reaches zero it stays there. So clearly the iterates remain feasible all the time. We will call (7) the basic EMML algorithm.

This is the exact same algorithm derived for ML image reconstruction in emission tomography described in [3] using the EM approach [14]. In the unmixing application, this algorithm does not solve the ML estimation problem but we keep calling it EMML to maintain consistency with [2]. A key distinction between the unmixing and the image reconstruction problems is that unmixing is in general an overconstrained ($m>n$) problem while image reconstruction is underconstrained ($m<n$). This is an important issue when looking at using results on convergence of the EMML algorithm applied to the unmixing case and to the use of block iterative algorithms to accelerate the EMML basic algorithm (7).

Recall that we get the same equation by minimizing a distance measure. A key problem with the basic algorithm (7) is its slow convergence rate. Later, we will review methods to speed up convergence.

3.2 The ISRA Algorithm

The Image Space Reconstruction Algorithm (ISRA) is the multiplicative iterative algorithm to solve (1). It was originally proposed in [6] for image reconstruction in ET based on minimizing the least squares cost function. In its derivation, we follow an approach similar to that used for the EMML algorithm (7).

Let the Lagrangian function for (2) be given by

$$L(\mathbf{x}, \boldsymbol{\mu}) = \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 - \boldsymbol{\mu}^T \mathbf{x}$$

The necessary and sufficient conditions (because of convexity of the LS distance and $\Re_+^n$) for an optimal point of (1) are given by

$$\frac{\partial L(\mathbf{x}, \boldsymbol{\mu})}{\partial x_j} = \sum_{i=1}^{m} a_{ij}\left(\mathbf{a}_i^T \mathbf{x} - b_i\right) - \mu_j = 0 \tag{8}$$

$$\mu_j x_j = 0$$

where $j=1,2, \ldots,n$ and $\boldsymbol{\mu} \geq 0$. Solving for μ_j in (8) results in

$$x_j \sum_{i=1}^{m} \left(\mathbf{a}_i^T \mathbf{x} - b_i\right) a_{ij} = 0 \tag{9}$$

From this equation, we can see that if $x_j=0$, $\mu_j \geq 0$ and if $x_j>0$ then $\mu_j=0$. The ISRA equation is obtained by solving for x_j in (9)

$$x_j = x_j \frac{\sum_{i=1}^{m} b_i a_{ij}}{\sum_{i=1}^{m} a_{ij}\mathbf{a}_i^T \mathbf{x}} \tag{10}$$

The basic iterative ISRA algorithm is given by

$$\hat{x}_j^{k+1} = \hat{x}_j^k \frac{\sum_{i=1}^{m} b_i a_{ij}}{\sum_{i=1}^{m} a_{ij}\mathbf{a}_i^T \hat{\mathbf{x}}^k} \tag{11}$$

This algorithm was originally proposed in [6] although an earlier version appeared in [15]. Similar to (7), (11) remains feasible all the time and once a component becomes equal to zero it stays at zero for the rest of the iterations.

Other algorithms are proposed in the literature to solve the NNLS problem (2). The most commonly used algorithm for NNLS is the active set strategy presented in [11] that is part of the MATLAB™ Optimization Toolbox.

3.3 Comparison with Steepest Descent

In the following sections, we will summarize convergence results for both the EMML and ISRA algorithms. Before continuing with our discussion, it is worth mentioning that both strategies are gradient descent type algorithms. For the distance minimization problems, the gradient descent algorithm has the structure

$$\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^k - \mathbf{D}\left(\hat{\mathbf{x}}^k\right)\frac{\partial \ell\left(\hat{\mathbf{x}}^k\right)}{\partial \mathbf{x}}$$

where $\ell(\mathbf{x})$ represents either the LS or KL distances and $\mathbf{D}(\mathbf{x})$ is a positive semidefinite matrix. For the EMML approach, set $\mathbf{D}(\mathbf{x})=\mathrm{diag}\{\gamma_1 x_1, \gamma_2 x_2, \ldots, \gamma_n x_n\}$, we get the descent algorithm

$$\hat{x}_j^{k+1} = \hat{x}_j^k - \gamma_j \hat{x}_j^k \sum_{i=1}^{m}\left(1 - \frac{b_i}{\mathbf{a}_i^T \mathbf{x}}\right) a_{ij} \tag{12}$$

In particular, if we set

$$\gamma_j = \frac{1}{\sum_{i=1}^{m} a_{ij}} \tag{12a}$$

the basic EMML algorithm (7) results. From (12a) is clear that in the EMML algorithm individual components of the gradient are weighted differently. Notice that (12), like equations (7) and (11), is also a multiplicative type of iteration.

A similar approach can be followed for the ISRA algorithm. The gradient multiplicative update is given by

$$\hat{x}_j^{k+1} = \hat{x}_j^k - \gamma_j \hat{x}_j^k \sum_{i=1}^{m}\left(\mathbf{a}_i^T \hat{\mathbf{x}}^k - b_i\right) a_{ij} \tag{13}$$

By setting

$$\gamma_j = \frac{1}{\sum_{i=1}^{m} a_{ij}\mathbf{a}_i^T \hat{\mathbf{x}}^k}$$

the basic ISRA algorithm (11) is obtained.

Notice that the fixed points of algorithms (12) and (13) are the solutions to (5) and (9) respectively.

4 CONVERGENCE ANALYSIS

4.1 EMML Algorithm

Convergence analysis is shown using an approach similar to that presented in [16]. Convergence results for the basic EMML are presented in [3] and for ISRA in [8]. To show convergence, we need the following [16].

Definition 1: $G(\mathbf{x}, \mathbf{x}')$ is an auxiliary function for $F(\mathbf{x})$ if the conditions

1. $G(\mathbf{x}, \mathbf{x}') \geq F(\mathbf{x})\ \forall\ \mathbf{x}$
2. $G(\mathbf{x}, \mathbf{x}) = F(\mathbf{x})$

are satisfied.

Lemma 1: If G is an auxiliary function then F is non-increasing under the update

$$\hat{\mathbf{x}}^{k+1} = \arg\min_{\mathbf{x}} G(\mathbf{x}, \hat{\mathbf{x}}^k) \tag{14}$$

Proof: $F(\hat{x}^k) = G(\hat{x}^k, \hat{x}^k) \geq G(\hat{x}^{k+1}, \hat{x}^k) \geq F(\hat{x}^{k+1})$.

Proposition 1: The function

$$G(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{m} \left(b_i \log b_i - b_i\right) + \sum_{i=1}^{m} \mathbf{a}_i^T \mathbf{x} - \sum_{i=1}^{m} \sum_{j=1}^{n} b_i \frac{a_{ij} x_l'}{\mathbf{a}_i^T \mathbf{x}'} \left(\log a_{ij} x_l - \log \frac{a_{ij} x_j'}{\mathbf{a}_i^T \mathbf{x}'} \right) \tag{15}$$

is an auxiliary function for $F(\mathbf{x}) = KL(\mathbf{b}, \mathbf{Ax})$.

Proof: Need to show that conditions (1) and (2) in definition 1 hold. Condition (1) of Lemma 1 can be shown by pure substitution. Condition (2) is equivalent to

$$G(\mathbf{x}, \mathbf{x}') - F(\mathbf{x}) \geq 0$$

By direct substitution, we get that

$$G(\mathbf{x}, \mathbf{x}') - F(\mathbf{x}) = \sum_{i=1}^{m} b_i \left(\log \mathbf{a}_i^T \mathbf{x} - \sum_{j=1}^{n} \frac{a_{ij} x_l'}{\mathbf{a}_i^T \mathbf{x}'} \left(\log a_{ij} x_j - \log \frac{a_{ij} x_j'}{\mathbf{a}_i^T \mathbf{x}'} \right) \right) \geq 0$$

which implies that for positivity to hold, we need

$$\log \mathbf{a}_i^T \mathbf{x} - \sum_{l=1}^{n} \frac{a_{ij} x_l'}{\mathbf{a}_i^T \mathbf{x}'} \left(\log a_{ij} x_j - \log \frac{a_{ij} x_j'}{\mathbf{a}_i^T \mathbf{x}'} \right) \geq 0 \tag{16}$$

Define $\alpha_i = \frac{a_{il} x_l'}{\mathbf{a}_i^T x'} \geq 0$ and notice that $\sum_{i=1}^{n} \alpha_i = 1$. Therefore, (16) can be re-written as

$$\log \mathbf{a}_i^T \mathbf{x} \geq \sum_{l=1}^{n} \alpha_i \log \frac{a_{il} x_l}{\alpha_i} \tag{17}$$

which holds by convexity of the log function. Therefore $G(\mathbf{x}, \mathbf{x}') \geq F(\mathbf{x})$ and (15) is an auxiliary function for the KL distance.

Using (15), the optimum for (14) is given by the solution of

$$\frac{\partial G\left(\hat{\mathbf{x}}^{k+1}, \hat{\mathbf{x}}^k\right)}{\partial x_j} = \sum_{i=1}^{m} a_{ij} \hat{x}_j^{k+1} + \sum_{i=1}^{m} \frac{a_{ij} b_i}{\mathbf{a}_i^T \hat{\mathbf{x}}^k} \hat{x}_j^k = 0$$

Solving this equation for $\hat{x}_j^{k+1}$ results in (7), which shows that each iteration of the EMML is equivalent to (4) for the KL auxiliary function (15). Therefore, the sequence $\{\hat{\mathbf{x}}^k\}$ generated by the EMML algorithm results in $KL(\mathbf{b}, \mathbf{A}\hat{\mathbf{x}}^k) = G(\hat{x}^k, \hat{x}^k) \geq G(\hat{x}^{k+1}, \hat{x}^k) \geq KL(\mathbf{b}, \mathbf{A}\hat{\mathbf{x}}^{k+1})$. So we have a non-increasing sequence that is bounded below so it must have an accumulation point. Because of convexity and continuity of KL, (7) converges to a solution of (5).

Notice that $\mathrm{KL}(\mathbf{b}, \mathbf{A}\hat{\mathbf{x}}) = 0$ if the system of equations has a positive solution (i.e. is consistent). If not, it converges to the minimum distance solution based on the KL distance measure.

4.2 Convergence of ISRA Algorithm

The convergence of the ISRA algorithm can be done in a similar manner to that of the EMML shown in the previous section [16]. Here, we just present the auxiliary function and leave to the reader the details of the convergence proof. The auxiliary function for the LS cost [16] is given by

$$\mathrm{G}(\mathbf{x}, \mathbf{x}') = \mathrm{F}(\mathbf{x}') + (\mathbf{x} - \mathbf{x}')^{\mathrm{T}} \nabla \mathrm{F}(\mathbf{x}') + \tfrac{1}{2}(\mathbf{x} - \mathbf{x}')^{\mathrm{T}} \mathbf{K}(\mathbf{x}')(\mathbf{x} - \mathbf{x}')$$

where $\mathrm{F}(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ and $\mathbf{K}(\mathbf{x}')$ is a diagonal matrix with diagonal elements given by

$$[\mathbf{K}(\mathbf{x}')]_{jj} = \frac{x'_j}{\sum_{i=1}^{m} a_{ij} \mathbf{a}_i^{\mathrm{T}} \mathbf{x}'}$$

4.3 Speeding Convergence of the Algorithms

A major complaint about the basic algorithms (7) and (11) is their slow convergence. Examples in [2] show the basic EMML taking several tens to hundreds of thousands iterations to converge to the optimum which is not acceptable for the unmixing application. The slow convergence does not come as a surprise since, however, as shown previously, these algorithms are variations of the gradient descent algorithm which in general exhibit linear convergence behavior [17]. Higher order methods such as Newton’s method exhibit quadratic convergence rates and can in principle be applied to this problem but they require a higher computational burden when compared to the simple EMML and ISRA algorithms.

Here, we present several methods described in the literature used to accelerate the convergence of the EMML algorithm (7). The acceleration methods are grouped in four categories: line search, relaxation, block iterative, and power methods.

4.3.1 Line Search Methods

Speed up in convergence speed of the EMML algorithm has been studied in the literature using several methodologies such as line search and relaxation methods. In the gradient equation (12), the line search parameter was chosen to obtain the EMML equation, an alternative is to choose that parameter in such a way that results in the highest decrease in cost, this will lead to the iterative algorithm

$$\hat{x}_j^{k+1} = \hat{x}_j^k - \gamma^k \hat{x}_j^k \sum_{i=1}^{m} \left(1 - \frac{b_i}{\mathbf{a}_i^{\mathrm{T}} \hat{\mathbf{x}}_k}\right) a_{ij} \tag{18}$$

where γ^k is selected as follows

$$\gamma^k = \arg\min \mathrm{KL}\left(\mathbf{b}, \mathbf{A}\left(\hat{\mathbf{x}}^k - \gamma \mathbf{D}(\hat{\mathbf{x}}^k) \frac{\partial \mathrm{KL}(\mathbf{b}, \mathbf{A}\hat{\mathbf{x}}^k)}{\partial \mathbf{x}}\right)\right) \tag{19}$$

and $\mathbf{D}(\mathbf{x})=\mathrm{diag}\{x_1, x_2, \ldots, x_n\}$. In the optimization literature [13], (19) is called a line search method. Notice that if $\frac{\partial \mathrm{KL}(\mathbf{b}, \mathbf{A}\hat{\mathbf{x}}^k)}{\partial x_j} = \sum_{i=1}^{m} \left(1 - \frac{b_i}{\mathbf{a}_i^{\mathrm{T}} \hat{\mathbf{x}}^k}\right) a_{ij} \le 0$, that particular component of the solution remains feasible for all values of γ. However, care must be exercised when implementing (19) when one or more of the components of the gradient are positive to avoid unfeasible solutions for the next iterate since the KL function is not defined when $\mathbf{a}_i^{\mathrm{T}} \mathbf{x} < 0$ for some i. To find the maximum step size, let us introduce a pseudo componentwise line search parameter defined by

$$\nu_j^k = \begin{cases} \infty & \text{if } \sum_{i=1}^{m}\left(1-\frac{b_i}{\mathbf{a}_i^T\hat{\mathbf{x}}^k}\right)a_{ij} \le 0 \\ 1 \Big/ \sum_{i=1}^{m}\left(1-\frac{b_i}{\mathbf{a}_i^T\hat{\mathbf{x}}^k}\right)a_{ij} & \text{otherwise} \end{cases}$$

From (18), we can see that an unfeasible solution is avoided by choosing the step-size parameter smaller than

$$\gamma^k \le \gamma_{max}^k = \max_j \left\{\nu_j^k\right\}$$

This will ensure that $\hat{x}_j^{k+1} \ge 0 \;\forall\; j = 1,2,3\ldots,n$. The line search approach was used in [18] as a means to accelerate the EMML. The line search approach is attractive as long as the minimization in (19) is easy to carry out. Several methodologies (e.g. Armijo's Rule, Step halving) are proposed in the general optimization literature to reduce the burden of solving line search problems such as (19) see [13, 17].

An interesting observation is that the line search method (19) weights all components of the gradient equally while the EMML gives individual components different weights according to (12a). This concept is further explored in block iterative methods to further increase the speed of convergence of the EMML algorithms for ET applications.

4.3.2 Relaxation Methods

In the relaxation approach, the new iterate is computed as follows

$$\hat{x}_j^{k+1} = \hat{x}_j^k + \omega^k \left(\hat{x}_j^k \frac{\sum_{i=1}^{m} \frac{b_i}{\mathbf{a}_i^T\hat{\mathbf{x}}^k} a_{ij}}{\sum_{i=1}^{m} a_{ij}} - \hat{x}_j^k \right) \tag{20}$$

where ω is called the relaxation parameter [17]. When $\omega \in (0,1)$, this is called an underrelaxation, if $\omega \in (1,\infty)$ this is called an overrelaxation, and when $\omega=1$ we recover the original EMML algorithm. It is pointed out in the literature [17], that using $\omega \in (1,\infty)$ can result in accelerated convergence for iterative algorithms. In our case, a large value of ω may cause an unfeasible iterate. A little algebra applied to (20) using the pseudo step-size parameters ν_j introduced earlier for the gradient approach give us the desired bound

$$\rho^k \le \rho_{max}^k = \max_j \left\{ \sum_{i=1}^{m} a_{ij} \nu_j^k \right\}$$

The optimal value for the relaxation parameter can be found using a line search type of method as follows

$$\rho^k = \arg \min_{\rho \in (0,\rho_{max}^k]} KL\left(b,\; \hat{\mathbf{x}}^k + \rho\left(\bar{\mathbf{x}}^k - \hat{\mathbf{x}}^k\right)\right) \tag{21}$$

which was used in [19]. Here $\bar{\mathbf{x}}^k$ is the basic EMML estimate (7). A step-doubling method was used in [20] that starts with $\omega=1$ and keeps doubling it until no further reduction in the distance function is achieved. These methods suffer from the same problem as the standard line search algorithm in that it requires evaluation of the cost function. In [21], the following heuristic method was suggested to compute the relaxation parameter without function evaluation

$$\omega^k = \min\left\{ \frac{\omega_{max}^k + 1}{2}, 4 \right\}$$

so that

$$1 \le \omega^k \le 4$$

Results presented in [21] show that a significant reduction in the number of iterations is achieved using this approach. However, no convergence analysis for this approach is presented.

4.3.3 Block Iterative Methods

Block implementations of the EMML algorithm [4, 22, 23, 24, 25] have been studied in the literature as a means to accelerate the convergence of the basic EMML iteration (7). Block iterative algorithms were applied to the unmixing problem in [2] and this work is a key motivating factor for our discussion.

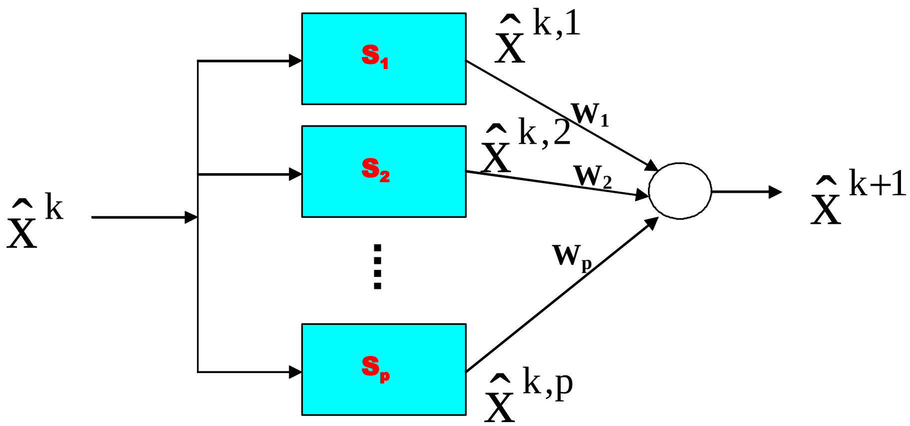

Figure 1: A pictorial representation of block simultaneous algorithms.

To understand the basic block concept, we will derive the block simultaneous approach [12] to the basic EMML as a simpler way to understand the block-processing concept. A general discussion about block-processing methods can be found in [12].

Let us partition the integer interval M={1, 2, 3, …,m} into disjoints subsets S_q such that

$$M = \bigcup_{q=1}^{p} S_q$$

we can rewrite (7) using this partition as follows

$$\hat{x}_j^{k+1} = \sum_{q=1}^{p} \hat{x}_j^k \frac{\sum_{i\in S_q} \frac{b_i}{\mathbf{a}_i^T \hat{\mathbf{x}}^k} a_{ij}}{\sum_{i=1}^{m} a_{ij}} = \sum_{q=1}^{p} \frac{\sum_{i\in S_q} a_{ij}}{\sum_{i=1}^{m} a_{ij}} \hat{x}_j^k \frac{\sum_{i\in S_q} \frac{b_i}{\mathbf{a}_i^T \hat{\mathbf{x}}^k} a_{ij}}{\sum_{i\in S_q} a_{ij}} \tag{22}$$

Notice that what we have done here is to partition the components of the sum in (7) in (p) groups and add the results of the sum for each subgroup. Define

$$w_q = \frac{\sum_{i\in S_q} a_{ij}}{\sum_{i=1}^{m} a_{ij}} \tag{23}$$

$$\hat{x}_j^{k+1,q} = \hat{x}_j^k \frac{\sum_{i\in S_q} \frac{b_i}{\mathbf{a}_i^T \hat{\mathbf{x}}^k} a_{ij}}{\sum_{i\in S_q} a_{ij}} \tag{24}$$

Substituting (23) and (24) in (22), we get the desired expression

$$\hat{x}_j^{k+1} = \sum_{q=1}^{p} w_q \hat{x}_j^{k,q} \tag{25}$$

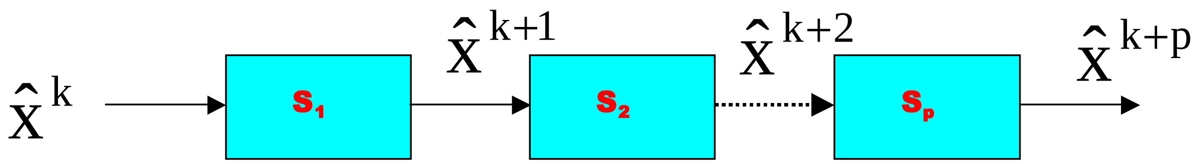

Figure 2: A graphical illustration of block iterative methods.

Notice that $w_q \geq 0$ and $\sum_{q=1}^{p} w_q = 1$. From this expression, we can see that the EMML estimate can be expressed as the convex combination of EMML like estimates (24) based on subsets of rows of **A**. Notice that $\hat{x}_l^{k+1,q}$ is like an EMML estimate based only on the q-th subset of rows of **A**. This type of processing is called in the optimization literature [12] a simultaneous block estimate where each estimate (24) is an estimate of the next iterant based on a subset of rows of **A**. A pictorial representation of (25) is given in Fig. 1. Notice that this expression points to a simple parallel implementation of the EMML algorithm where each processor is assigned a subset of bands and the abundance estimates are computed simultaneously and combined according to (22). Since (22) is a simple manipulation of (7), all the optimality and convergence properties of (7) are inherited by (22). Notice that the block simultaneous algorithm results in (7) for the one block case.

Sequential processing of these blocks (as opposed to simultaneously) results in the sequential block iterative methods presented in [4, 22, 23, 24, 25] for the EMML problem. This is illustrated pictorially in Fig. 2. In the sequential block iterative approach, the next iterate is computed by applying the EMML to a subset of bands at each step. All subsets are visited in a cyclic fashion as follows

$$\hat{x}_j^{k+1} = \hat{x}_j^k \frac{\sum_{i \in S_q} \frac{b_i}{\mathbf{a}_i^T \hat{\mathbf{x}}^k} a_{ij}}{\sum_{i \in S_q} a_{ij}}, \quad q = k \bmod(p) + 1 \tag{26}$$

This is the ordered subset EMML (OS-EMML) algorithm proposed in [22]. This algorithm is shown to converge under subset balance, which is a very restrictive condition, and simulation results show non-convergence when the condition is not met. Further improvements were proposed based on an underrelaxed version of (26) by [4, 23, 24, 25].

$$\hat{x}_j^{k+1} = \hat{x}_j^k + \omega_j^k \left(\hat{x}_j^k \frac{\sum_{i \in S_q} \frac{b_i}{\mathbf{a}_i^T \hat{\mathbf{x}}^k} a_{ij}}{\sum_{i \in S_q} a_{ij}} - \hat{x}_j^k \right), \quad q = k \bmod(p) + 1 \tag{27}$$

Notice that with ω=1, we recover the OS-EMML algorithm. The main difference between those approaches is how to select ω_j^k. For the block iterative EMML (BI-EMML) algorithm of [24], we have that

$$\omega_j^k = \frac{\sum_{i \in S_q} a_{ij}}{\sum_{i=1}^{m} a_{ij}} \tag{28}$$

For the rescaled BI-EMML (RBI-EMML), we get [4, 25]

$$\omega_j^k = \frac{\sum_{i \in S_q} a_{ij}}{\overline{\omega}_q \sum_{i=1}^{m} a_{ij}}, \text{ where } \overline{\omega}_q = \max_j \frac{\sum_{i \in S_q} a_{ij}}{\sum_{i=1}^{m} a_{ij}} \tag{29}$$

Finally, for the row-action maximum likelihood algorithm (RAMLA) of [23], we get

$$\rho_j^k = \lambda_k \sum_{i \in S_q} a_{ij}, \quad \text{where } \lambda_k \to 0 \text{ as } k \to \infty \text{ and} \sum_{k=1}^{\infty} \lambda_k = \infty \tag{30}$$

If we consider the one block case, the block iterative methods become an underrelaxed iteration for the basic EMML (7). The RBI-EMML method was applied to the unmixing problem in [2].

Convergence is shown for BI-EMML, RBI-EMML, and RAMLA in the consistent case (when Ax=**b** has a solution). In the inconsistent case, results in [25] show that BI-EMML and RBI-EMML converge to a limit cycle that depends on system dimension and noise while results in [23] show convergence for the inconsistent case but does not give clear guidelines on how to select the relaxation parameter. Notice that in the case of the basic EMML (7) and its block simultaneous version (25) converges to a solution even for the inconsistent case. This difference comes from the fact that algorithms (26)-(29) do not solve the distance minimization problem (4) but a related convex feasibility problem (CFP) [12], which can be described as finding an element at the intersection of a set of convex sets. The solution to a consistent linear system of equations can be formulated as a CFP, see [12]. The basic algorithm for this problem starts with a random guess and projects sequentially onto each convex set sequentially until convergence is achieved. In our case, $\hat{\mathbf{x}}^{k+1}$ is the generalized projection of $\hat{\mathbf{x}}^{k}$ in the convex set

$$C_q = \left\{ \mathbf{x} \in \Re_+^n \middle| \mathbf{a}_i^T \mathbf{x} = b_i, \text{for } i \in S_q \right\} \; q = k \bmod(p) + 1$$

using the KL distance [12].

4.3.4 Power Methods

A different approach to accelerate the basic algorithms (7) and (11) was proposed in [26, 27] as follows

$$\hat{x}_l^{k+1} = \hat{x}_l^k \left(\frac{\sum_{i=1}^{m} \frac{b_i}{\mathbf{a}_i^T \hat{\mathbf{x}}^k} a_{il}}{\sum_{i=1}^{m} a_{il}} \right)^n \tag{31}$$

$$\hat{x}_l^{k+1} = \hat{x}_l^k \left(\frac{\sum_{i=1}^{m} b_i a_{il}}{\sum_{i=1}^{m} a_{il} \mathbf{a}_i^T \hat{\mathbf{x}}^k} \right)^n \tag{32}$$

where $n \le 2$. Experimental results show that the number of iterations is reduced by a factor of n. In [10], it is shown that this algorithm can be derived from the basic distance minimization problem (3) by substituting the positivity constraint $x_j \ge 0$ by $(x_j)^{1/n} \ge 0$.

5 CONCLUSIONS

This paper tries to summarize our work on analyzing the theoretical aspects of the EMML and ISRA algorithms and their potential application to the spectral unmixing problem. This analysis has served to clarify what can and cannot be done with the EMML approach in the spectral unmixing application and its optimality properties.

The first important result is the fact that ISRA solves the ML for the unmixing application while EMML searches for the best approximation in the KL sense. The ML estimate, from a statistical estimation point of view, is superior to the minimimum KL distance estimate. Comparisons between the EMML and the ISRA algorithms that are presented in the tomography literature [28] focus on the features in the reconstructed image and their common mathematical framework [29]. From a computational point of view, the EMML (7) and ISRA algorithms (11) require the same computational effort so from that point of view there is no clear advantage of one algorithm over the other. Outside the statistical advantage of ISRA over EMML, it is necessary, to compare these algorithms with metrics of great relevance to the spectral unmixing problem such as speed of convergence, and noise robustness. The final goal is a fast, reliable and robust algorithm even if it only gives suboptimal results from an estimation point of view.

ACKNOWLEDGEMENTS

This work was sponsored primarily by the Air Force Summer Faculty Fellowship Program administered by the National Research Council. The research performed here used facilities of the Center for Subsurface Sensing and Imaging Systems at the University of Puerto Rico-Mayaguez sponsored by the Engineering Research Centers Program of the US National Science Foundation under grant EEC-9986821.

REFERENCES

[1] N. Keshava and J.F. Mustard, "Spectral unmixing." In *IEEE Signal Processing Magazine*, pp. 44-57, January 2002.

[2] E. Meidunas, A. Puetz and M. Hoke, "RBI-EMML Separation for Imaging Techniques." In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*, Proceedings of SPIE, Volume 4725, 2002.

[3] L.A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography." In *IEEE Transactions on Medical Imaging*, Vol. MI-1, No. 2, October 1982.

[4] C.L. Byrnes, "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods." In *IEEE Transactions on Image Processing*, Vol. 7, No. 1, January 1998.

[5] M.L. Hoke and E.C. Meidunas, "Bayes' theorem applied to linear regression." AFRL Technical Report, AFRL-VS-TR-2001-1617, May 2002.

[6] M.E. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorithm suitable for volume ECT." In *IEEE Transactions on Medical Imaging*, Vol. MI-5, No. 2, June 1986.

[7] P.P.B. Eggermont, "Multiplicative iterative algorithms for convex programming." In *Linear Algebra and its Applications*, 130:25-42, 1990.

[8] A.R. De Pierro, "Nonlinear relaxation methods for solving symmetric linear complementarity problems." In *Journal of Optimization Theory and Applications*, Vol. 64, No. 1, January 1990.

[9] Y.S. Popkov, "Multiplicative algorithms for determining positive solutions of nonlinear equations." In *Journal of Computational and Applied Mathematics*, Vol. 69, 1996.

[10] H. Lanteri, M. Roche, O. Cuevas, and C. Aime, "A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints." In *Signal Processing*, Vol. 81, 2001.

[11] C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, 1974.

[12] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications*, Oxford University Press, 1997.

[13] D.G. Luenberger, *Linear and Nonlinear Programming*, 2nd Edition, Prentice-Hall, 1984.

[14] T.K. Moon, "The Expectation-Maximization Algorithm." In *IEEE Signal Processing Magazine*, November 1996.

[15] M.T. Chahine, "Inverse problems in radiative transfer: determination of atmospheric parameters." In *Journal of Atmospheric Sciences*, Vol. 27, 1970.

[16] D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization." In *Advances in Neural Information Processing*, p. 556-562, 2001.

[17] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1970.

[18] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography." In *IEEE Transactions on Medical Imaging*, Vol. MI-6, No. 1, March 1987.

[19] Y. Vardi, L.A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography." In *Journal of the American Statistical Association*, Vol. 80, March 1985.

[20] T.J. Holmes and Y.H. Liu, "Acceleration of maximum likelihood image restoration for fluorescence microscopy and other noncoherent imagery." In *Journal of the Optical Society of America: Part A*, Vol. 8, No. 6, June 1991.

[21] R.M. Lewitt and G. Muehllehner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation." In *IEEE Transactions on Medical Imaging*, Vol. MI-5, No. 1, March 1986.

[22] H.M. Hudson and R.S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data." In *IEEE Transactions on Medical Imaging*, Vol. 13, No. 4, December 1994.

[23] J. Browne and A.R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography." In *IEEE Transactions on Medical Imaging*, Vol. 15, No. 5, October 1996.

[24] C.L. Byrne, "Block-iterative methods for image reconstruction from projections." In *IEEE Transactions on Image Processing*, Vol. 5, No.5, May 1996.

[25] C.L. Byrne, "Convergent block-iterative algorithms for image reconstruction from inconsistent data." In *IEEE Transactions on Image Processing*, Vol. 6, No. 9, September 1997.

[26] T.S. Zaccheo and R.A. Gonsalves, "Iterative maximum-likelihood estimators for positively constrained objects." In *Journal of the Optical Society of America: Part A*, Vol. 13, No. 2, February 1996.

[27] D.S. Biggs and M. Andrews, "Accelerration of iterative image restoration algorithms." In *Applied Optics*, Vol. 36, No. 8, March 1997.

[28] H. Lanteri, R. Soummer, and C. Aime, "Comparison between ISRA and RLA algorithms: use of a Wiener filter based stopping criterion." In *Astronomy & Astrophysics Suplement Series*, Vol. 140, 1999.

[29] A.R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography." In *IEEE Transactions on Medical Imaging*, Vol. 12, No. 2, June 1993.