LINEAR UNMIXING OF HYPERSPECTRAL IMAGES USING A SCALED GRADIENT METHOD

Céline Theys[(1)]*, Nicolas Dobigeon*[(2)]*, Jean-Yves Tourneret*[(2)] *and Henri Lantéri*[(1)]

[(1)] Laboratoire Fizeau, UMR 6525, University of Nice Sophia-Antipolis, 06108 Nice cedex 2, France
[(2)] University of Toulouse, ENSEEIHT-IRIT-TéSA, 31071 Toulouse cedex 7, France
{Celine.Theys,Henri.Lanteri}@unice.fr, {Nicolas.Dobigeon,Jean-Yves.Tourneret}@enseeiht.fr

ABSTRACT

This paper addresses the problem of linear unmixing for hyperspectral imagery. This problem can be formulated as a linear regression problem whose regression coefficients (abundances) satisfy sum-to-one and positivity constraints. Two scaled gradient iterative methods are proposed for estimating the abundances of the linear mixing model. The first method is obtained by including a normalization step in the scaled gradient method. The second method inspired by the fully constrained least squares algorithm includes the sum-to-one constraint in the observation model with an appropriate weighting parameter. Simulations on synthetic data illustrate the performance of these algorithms.

Index Terms— Hyperspectral imagery, unmixing, optimization.

1. INTRODUCTION

Hyperspectral and multispectral imagery have received considerable attention in the literature (see for instance [1], [2] and references therein). Hyperspectral and multispectral data are collected in many spectral bands. As a consequence, each pixel of the image is characterized by a spectral signature which contains the measurements associated to the different frequencies contained in these spectral bands. Moreover, each pixel of the image can be modeled accurately as a linear mixture of different pure spectral components usually referred to as *endmembers*. Important problems result from this linear mixture model (LMM). First, it is important to estimate the spectral signatures associated to the endmembers. A very popular endmember extraction algorithm (EEA) is the N-finder (N-FINDR) algorithm proposed by Winter [3]. Another important problem is the linear unmixing of each pixel of the image. Linear unmixing consists of estimating the proportions of each endmember contained in a given pixel [4]. These proportions are referred to as *abundances* in hyperspectral imagery.

The abundances have to satisfy sum-to-one and positivity constraints since they correspond to the fractions of the materials involved in the mixture. There are mainly two kinds of approaches which can be used to estimate the abundances. The first approach defines appropriate prior distributions for the abundances (satisfying the sum-to-one and positivity constraints) following the principles of Bayesian inference [5]. However, the complexity of the parameter posterior distribution (due essentially to the constraints inherent to abundances) requires to develop sophisticated sampling algorithms to compute the Bayesian estimators. These algorithms include the Gibbs sampler or the Metropolis-within-Gibbs algorithm [6]. This approach was for instance advocated in [5] and provided interesting results. The price to pay with Bayesian unmixing algorithms is the high computational complexity resulting from the sampling strategy. The second approach consists of estimating the abundances by minimizing an appropriate cost function such as the least squares criterion under sum-to-one and positivity constraints. As explained in [7], there is no analytical solution for this optimization problem because of the linear inequalities resulting from the positivity constraints. However, an efficient iterative algorithm referred to as fully constrained least square (FCLS) algorithm was proposed in [7]. The FCLS was applied successfully to the unmixing of hyperspectral images. One of its main advantages over the Bayesian estimators is its computational cost since few iterations are required to ensure convergence to a local minimum of the least-squares criterion. However, the convergence of this algorithm to a global minimum of the LS criterion is not guarantied which is its main drawback.

This paper studies new constrained scaled gradient methods (SGM) for estimating the abundances of the LMM. It is organized as follows. Section 2 presents the standard LMM. Section 3 derived the SGM for linear mixing models satisfying positivity constraints. Two modified SGM allowing us to handle sum-to-one constraints are presented in Section 4. Simulation results are presented in Section 5. Conclusions are reported in Section 6.

2. LMM FOR HYPERSPECTRAL IMAGERY

The LMM assumes that a mixed pixel resulting from the observation in L spectral bands can be written as

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_L)^T$ is the observation vector, $\boldsymbol{M}$ is a $L \times R$ matrix formed by the R endmember spectra identified by an EEA (and thus assumed to be known), R is the number of endmembers involved in the mixture (with $R \ll L$), $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_R)^T$ is the abundance vector and $\boldsymbol{n} = (n_1, \ldots, n_L)^T$ is the noise vector.

The linear unmixing problem considered in this paper consists of estimating $\boldsymbol{\alpha}$ under positivity and sum-to-one constraints

$$\alpha_r \geq 0 \text{ and } \sum_{r=1}^{R} \alpha_r = 1 \quad \forall r = 1, \ldots, R.$$

A standard assumption related to the LMM defined in (1) is that the noise vector is distributed according to a Gaussian distribution with zero-mean and covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_L$, where $\boldsymbol{I}_L$ is the $L \times L$ identity matrix, which will be denoted $\boldsymbol{n} \sim \mathcal{N}(0_L, \sigma^2 I_L)$. Note that this statistical model assumes that the noise variances are the same in all bands. This assumption has been used extensively in the literature (see for instance [8] or [9]). Since the variance σ^2 can be easily estimated from the observation vector, it is assumed to be known in this paper. After removing the additive and multiplicative

constants, the resulting negative log-likelihood function reduces to the following cost function

$$J(\boldsymbol{\alpha}) = (\boldsymbol{y} - \boldsymbol{M}\boldsymbol{\alpha})^T(\boldsymbol{y} - \boldsymbol{M}\boldsymbol{\alpha}). \quad (2)$$

3. SGM FOR POSITIVITY CONSTRAINTS

This section studies an iterative method referred to as SGM for estimating the abundances under positivity constraints. In other words, we first do not consider the sum-to-one constraint which will be handled in the next section. Minimizing a convex cost function under inequality constraints is classically achieved by introducing an appropriate Lagrange function. The Lagrange function $\mathcal{L}$ associated to the linear unmixing problem with positivity constraints is

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = J(\boldsymbol{\alpha}) - \boldsymbol{\lambda}^T \mathbf{g}(\boldsymbol{\alpha})$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_R)^T$ contains the Lagrange multipliers, $\mathbf{g}(\boldsymbol{\alpha}) = (g(\alpha_1), \ldots, g(\alpha_R))^T$. The function g has to be chosen in order to express the positivity constraints. More precisely, g is an increasing function that must be positive for inactive constraints ($\alpha_r > 0$), and zero for active constraints ($\alpha_r = 0$). The Kuhn-Tucker conditions at the optimum $\boldsymbol{\alpha}^*, \boldsymbol{\lambda}^*$ express as follows

$$\nabla\mathcal{L}(\boldsymbol{\alpha}^*, \boldsymbol{\lambda}^*) = 0 \quad (3)$$
$$g(\alpha_r^*) \geq 0, \qquad \forall r \quad (4)$$
$$\lambda_r^* \geq 0, \qquad \forall r \quad (5)$$
$$\lambda_r^*\, g(\alpha_r^*) = 0, \qquad \forall r \quad (6)$$

where $\nabla\mathcal{L}$ is the gradient vector of $\mathcal{L}$. The gradient vector of the cost function (2) can be easily computed

$$\nabla J(\boldsymbol{\alpha}) = \boldsymbol{M}^T\boldsymbol{M}\boldsymbol{\alpha} - \boldsymbol{M}^T\boldsymbol{y}. \quad (7)$$

As a consequence, (3) leads to

$$\lambda_r^* \left[\frac{\partial g(\boldsymbol{\alpha}^*)}{\partial \boldsymbol{\alpha}}\right]_r = [\nabla J(\boldsymbol{\alpha}^*)]_r$$

where the notation $[\cdot]_r$ is used for the rth component of a vector. Equivalently, the rth Lagrange multiplier can be expressed as

$$\lambda_r^* = \frac{[\nabla J(\boldsymbol{\alpha}^*)]_r}{[\frac{\partial g(\boldsymbol{\alpha}^*)}{\partial \boldsymbol{\alpha}}]_r} \quad (8)$$

for $r = 1, \ldots, R$. Equations (6) and (8) lead to

$$[\nabla J(\boldsymbol{\alpha}^*)]_r g_r(\boldsymbol{\alpha}^*) = 0, \qquad r = 1, \ldots, R.$$

This last equation allows us to derive an iterative algorithm to estimate $\boldsymbol{\alpha}$ under positivity contraints. A simple choice for g is $g(\alpha) = \alpha$. However, more generally the r^{th} component of the descent direction can be chosen as follows

$$f_r(\alpha)\hat{\alpha}_r[-\nabla J(\hat{\boldsymbol{\alpha}})]_r \quad (9)$$

with $f(\alpha)$ a positive function scaling the gradient, leading to the SGM. Thus the SGM is defined as

$$\alpha_r^{(k+1)} = \alpha_r^{(k)} + \gamma_r^{(k)} f_r(\alpha^{(k)})\alpha_r^k[-\nabla J(\boldsymbol{\alpha}^{(k)})]_r \quad (10)$$

where $\gamma_r^{(k)}$ is the descent step-size that must be adjusted to ensure convergence of the algorithm.

An interesting choice for the scaling function $f_r(.)$ initially proposed in [10] is recalled below. It is interesting to note here that this particular choice with step-size $\gamma_r^{(k)} = 1$ ensures the algorithm convergence. The negative gradient of any convex cost function $J(\boldsymbol{\alpha})$ can be expressed as the difference between two positive vectors (i.e., vectors whose components are all positive)

$$-\nabla J(\boldsymbol{\alpha}) = U(\boldsymbol{\alpha}) - V(\boldsymbol{\alpha}).$$

For instance, when all elements of $\boldsymbol{y}$ are positive, the two positive vectors associated to (7) are[1]

$$U(\boldsymbol{\alpha}) = \boldsymbol{M}^T\boldsymbol{y} \quad \text{and} \quad V(\boldsymbol{\alpha}) = \boldsymbol{M}^T\boldsymbol{M}\boldsymbol{\alpha}. \quad (11)$$

Choosing the scaling function as

$$f_r(\boldsymbol{\alpha}) = \frac{1}{[V(\boldsymbol{\alpha})]_r} \quad (12)$$

and $\gamma_r^{(k)} = 1, \forall r = 1, \ldots, R$, we obtain the classical image space reconstruction algorithm (ISRA) initially studied in [11] whose convergence has been proved in [12]. After replacing the expression of $V(\boldsymbol{\alpha}^{(k)})$ in (12), the following iterative algorithm is obtained

$$\alpha_r^{(k+1)} = \alpha_r^{(k)}[\boldsymbol{M}^T\boldsymbol{y}]_r[\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{\alpha}^{(k)}]_r^{-1}$$

whose convergence is ensured provided $\hat{\alpha}_r^0 > 0,\ \forall r$.

4. ALGORITHMS FOR POSITIVITY AND SUM-TO-ONE CONSTRAINTS

4.1. Normalized SGM

In the case where abundances are subjected to positivity and sum-to-one constraints, we propose a normalized SGM defined as follows:

$$\alpha_r^{(k+1)} = \alpha_r^{(k)}[\boldsymbol{M}^T\boldsymbol{y}]_r[\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{\alpha}^{(k)}]_r^{-1} \quad (13)$$
$$\alpha_r^{(k+1)} = [P(\boldsymbol{\alpha}^{(k+1)})]_r \quad (14)$$

where $P(.)$ is the projection operator associated to the sum-to-one constraint. In other words the abundances are normalized at each iteration similarly to the normalized nonnegativity constrained least squares algorithm [7], except the scaled gradient descent is used instead of the standard descent scheme.

4.2. Fully constrained SGM

An alternative to the normalized SGM is to follow the strategy advocated in [7] yielding the FCLS algorithm. The FLCS algorithm considers the sum-to-one constraint as an additional noiseless measurement equation, i.e., the abundances are estimated using the following set of equations

$$\delta\boldsymbol{y} = \delta\boldsymbol{M}\boldsymbol{\alpha} + \delta\boldsymbol{n}$$
$$1 = \mathbf{1}_R^T\boldsymbol{\alpha}$$

where $\mathbf{1}_R = (1, \ldots, 1)^T$ is a vector made of R ones. By defining an extended observation vector $\boldsymbol{s} = (\delta\boldsymbol{y}, 1)^T$ and the mixing matrix

$$\boldsymbol{N} = \begin{bmatrix} \delta\boldsymbol{M} \\ \mathbf{1}_{\boldsymbol{R}}{}^T \end{bmatrix}$$

we obtain a new linear regression problem which was solved in [7] using a nonnegativity constrained least squares algorithm. Note that

[1]For any negative component $[\boldsymbol{M}^T\boldsymbol{y}]_r$, $[U(\boldsymbol{\alpha})]_r$ and $[V(\boldsymbol{\alpha})]_r$ can be defined as $[U(\boldsymbol{\alpha})]_r = 0$ and $[V(\boldsymbol{\alpha})]_r = [\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{M}^T\boldsymbol{y}]_r$.

the factor δ controls the impact of the abundance sum-to-one constraint with respect to the positivity constraint. By replacing the non-negativity constrained least squares algorithm by the SGM proposed in this paper, we obtain a so-called fully constrained SGM (similar to FCLS). This algorithm referred to as FC-SGM consists of replacing the cost function defined in (2) by

$$J_\delta(\boldsymbol{\alpha}) = \delta^2(\boldsymbol{y} - \boldsymbol{M}\boldsymbol{\alpha})^T(\boldsymbol{y} - \boldsymbol{M}\boldsymbol{\alpha}) + (1 - \mathbf{1}_R^T\boldsymbol{\alpha})^2. \quad (15)$$

The corresponding gradient can be easily computed

$$\nabla J_\delta(\boldsymbol{\alpha}) = \delta^2\left(\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{\alpha} - \boldsymbol{M}^T\boldsymbol{y}\right) + \left(\sum_{r=1}^{R}\alpha_r - 1\right)\mathbf{1}_R.$$

The opposite gradient can be decomposed as the difference between two positive functions given by

$$U(\boldsymbol{\alpha}) = \delta^2\boldsymbol{M}^T\boldsymbol{y} + \mathbf{1}_R$$

$$V(\boldsymbol{\alpha}) = \delta^2\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{\alpha} + \left(\sum_{r=1}^{R}\alpha_r\right)\mathbf{1}_R.$$

Choosing the scaling function as in (12) and $\gamma_r^{(k)} = 1, \forall r$ leads to the following recursion for the abundance estimates

$$\hat{\alpha}_r^{(k+1)} = \hat{\alpha}_r^{(k)}\frac{[\boldsymbol{M}^T\boldsymbol{y} + \frac{1}{\delta^2}\mathbf{1}_R]_r}{[\boldsymbol{M}^T\boldsymbol{M}\hat{\boldsymbol{\alpha}}^{(k)} + \frac{1}{\delta^2}(\sum_{r=1}^{R}\hat{\alpha}_r^k)\mathbf{1}_R]_r}.$$

Note that the projection operation of (14) is not applied here since the sum-to-one constraint has been considered as a penalization in the least-squares criterion (15).

5. SIMULATION RESULTS

5.1. Synthetic data

Many simulations have been conducted to validate the previous algorithm. The first experiment corresponds to a linear mixture of $R = 3$ endmembers with $\boldsymbol{\alpha} = [0.3, 0.6, 0.1]^T$. The three endmembers used in this example have been extracted from the ENVI library [13] and correspond to the spectra of the construction concrete, green grass, and micaceous loam. The normalized SGM defined by (13) and (14) has been applied on these simulated data with a signal-to-noise ratio SNR $= 15$dB. Figure 1 shows an example of abundance estimates as a function of the number of iterations. The algorithm clearly converges after few iterations. Note that the initial values for the different abundances have been generated uniformly in the domain defined by the constraints. The mean and standard deviations of the estimated abundances averaged over 500 Monte Carlo runs are depicted in Fig. 2 as a function of the SNR. These results have to be compared with the results obtained with the FCLS and the Bayesian algorithms. It can be observed that the Bayesian algorithm shows a small bias but provides smaller confidence intervals. FCLS and normalized SGM perform very similarly.

The next simulations illustrate the advantage of the fully constraint SGM when compared to the normalized SGM. An experiment similar to the example 2 of [7] is considered. The mixture is composed of three endmembers (construction concrete, green grass, and micaceous loam). However, we are looking for two additional signatures (green paint and galvanized steel Metal) which are not present in the mixture. The abundances for this example are summarized in $\boldsymbol{\alpha} = [0.3, 0.6, 0.1, 0, 0]^T$. Figure 3 shows that the fully constrained SGM has to be preferred when the model contains endmembers which are absent in the observed pixel. Note that a similar observation was made in [7] for the normalized least-squares and fully constrained least-squares methods.

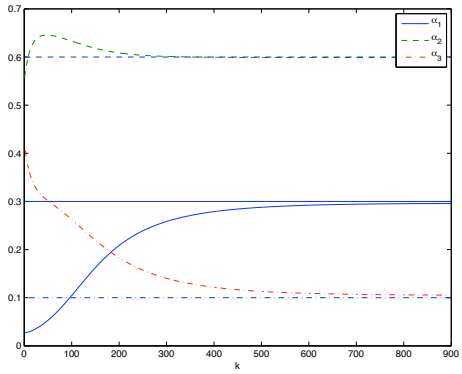

Fig. 1. Normalized SGM estimates of the abundances versus the iteration number k for $SNR = 15$dB.

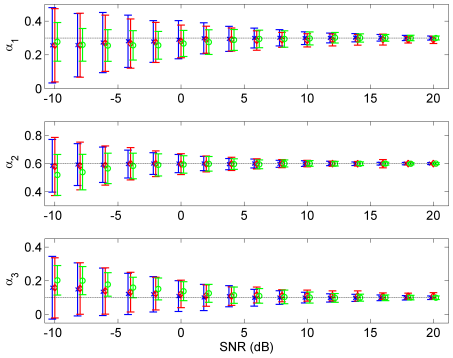

Fig. 2. Estimates (symbols) and standard deviations (vertical bars) versus SNR for the abundances estimated by different methods: normalized SGM (red diamond), FCLS (blue cross) and Bayesian unmixing (green circles).

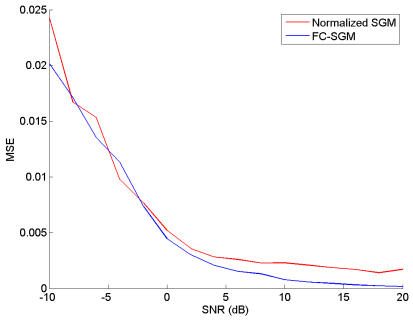

Fig. 3. MSEs for the estimated abundances: normalized SGM (red), fully constrained SGM (blue).

5.2. Real data

The proposed algorithms have also been applied on a small real image that has been widely used in the remote sensing literature [5]. It

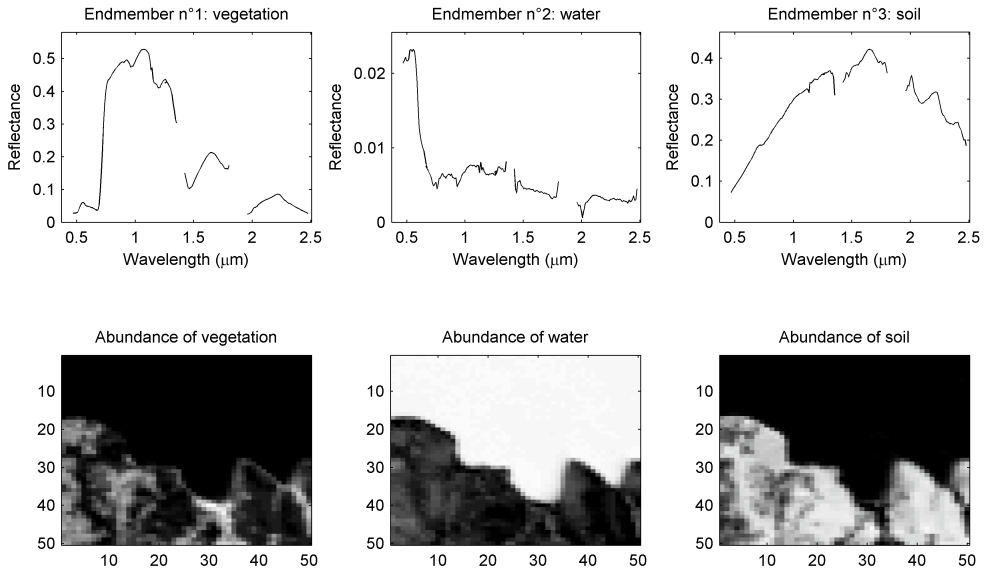

Fig. 4. Top: $R = 3$ endmember spectra recovered by N-FINDR. Bottom: abundance maps estimated by FC-SGM.

consists of a lake shore area whose coastal part is mainly covered by soil and vegetation. The $R = 3$ corresponding endmember spectra (vegetation, water and soil) have been identified by the N-FINDR algorithm [3] and are represented in Figure 4 (top) for $L = 413$. The results obtained by the FC-SGM algorithm are summarized by the abundance maps depicted in the bottom figures. In these maps, white (resp. black) pixels denote the presence (resp. absence) of the corresponding endmembers. Figure 4 shows that the lake is accurately recovered by the the algorithm. The estimations of the soil and vegetation abundances are also in agreement with results previously obtained in the literature.

6. CONCLUSIONS

Constrained scaled gradient methods were initially derived for linear models subjected to positivity constraints. This paper studied two modified constrained scaled gradient methods referred to as "normalized constraint gradient" and "fully constrained gradient" for estimating the abundances of a linear mixing model under positivity and sum-to-one constraints. The performance of these methods obtained on synthetic hyperspectral images and real data is promising. Future investigations include the analysis of algorithms such as those developed in [14–16] for estimating jointly the endmembers and abundances.

7. REFERENCES

[1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York: Wiley, 2003.

[2] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York: Plenum Publishing Co., 2003.

[3] M. E. Winter, "Fast autonomous spectral end-member determination in hyperspectral data," in *Proc. 13th Int. Conf. on Applied Geologic Remote Sensing*, vol. 2, Vancouver, April 1999, pp. 337–344.

[4] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[5] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2684–2695, July 2008.

[6] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. New York: Springer-Verlag, 2005.

[7] D. C. Heinz and C.-I Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 39, no. 3, pp. 529–545, March 2001.

[8] J. Wang and C.-I Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sensing*, vol. 44, no. 9, pp. 2601–2616, Sept. 2006.

[9] ——, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.

[10] H. Lanteri, M. Roche, and C. Aime, "Penalized maximum likelihood image restoration with positivity constraints - multiplicative algorithms," *Inverse problems*, vol. 18, no. 5, pp. 1397–1419, May 2002.

[11] M. E. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorithm suitable for volume ECT," *IEEE Trans. Medical Imaging*, vol. 5, no. 2, pp. 61–66, June 1986.

[12] A. R. DePierro, "On the convergence of the image space reconstruction algorithm for volume RCT," *IEEE Trans. Medical Imaging*, vol. 6, no. 2, pp. 328–333, June 1987.

[13] RSI (Research Systems Inc.), *ENVI User's guide Version 4.0*, Boulder, CO 80301 USA, Sept. 2003.

[14] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[15] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural computation*, vol. 19, pp. 2756–2779, 2007.

[16] J. Igual and R. Llinares, "Nonnegative matrix factorization of laboratory astrophysical ice mixtures," *IEEE J. Sel. Top. Signal Process.*, vol. 2, no. 5, pp. 697–706, Oct. 2008.