Dimensionality Reduction, Classification, and Spectral Mixture Analysis using Nonnegative Underapproximation

Nicolas Gillis[a] and Robert J. Plemmons[b]

[a]Université catholique de Louvain, Department of Mathematical Engineering and Center for Operations Research and Econometrics, B-1348 Louvain-la-Neuve, Belgium;
[b]Departments of Mathematics and Computer Science, Wake Forest University, Winston-Salem, NC 27106.

ABSTRACT

Nonnegative Matrix Factorization (NMF) and its variants have recently been successfully used as dimensionality reduction techniques for identification of the materials present in hyperspectral images. In this paper, we present a new variant of NMF called Nonnegative Matrix Underapproximation (NMU): it is based on the introduction of underapproximation constraints which enables one to extract features in a recursive way, like PCA, but preserving nonnegativity. Moreover, we explain why these additional constraints make NMU particularly well-suited to achieve a parts-based and sparse representation of the data, enabling it to recover the constitutive elements in hyperspectral data. We experimentally show the efficiency of this new strategy on hyperspectral images associated with space object material identification, and on HYDICE and related remote sensing images.

Keywords: Hyperspectral Images, Nonnegative Matrix Factorization, Underapproximation, Dimensionality Reduction, Classification, Spectral Mixture Analysis

1. INTRODUCTION

A crucial aspect of hyperspectral image analysis is the identification of materials present in an object or scene being imaged. Dimensionality reduction techniques such as PCA are widely used as a preprocessing step in order to reducing the computational cost while keeping the pertinent information. In this context, it is often preferable to take advantage of the intrinsic properties of hyperspectral data: each image corresponds to a wavelength and the spectral signature of each pixel results from the additive combination of the nonnegative spectral signatures of its constitutive materials. Taking these nonnegativity constraints into account enhances interpretability of the extracted factors. This can be done using the Nonnegative Matrix Factorization[1] (NMF) technique, generally formulated as the following optimization problem: given a $m \times n$ real nonnegative matrix M and a positive integer $r < \min(m, n)$, find two real nonnegative matrices U and V of dimensions $m \times r$ and $n \times r$ in order to minimize the sum of the squared entries of $M - UV^T$:

$$\min_{U,V} \ ||M - UV^T||_F^2 \quad \text{such that} \quad U \geq 0 \text{ and } V \geq 0. \tag{NMF}$$

NMF has been successfully used in many other applications, e.g., images processing, text mining, air emission control, microarray data analysis, clustering, etc.[2]

Assuming that the matrix M is constructed as follows: each 2D image corresponding to a wavelength is vectorized and is a column m_j of M and each row m^i of M corresponds to the spectral signature of a pixel; the above decomposition can be interpreted as follows

$$m^i \ \approx \ \sum_k u_k^i \ v_k^T \quad \forall i,$$

E-mail addresses: nicolas.gillis@uclouvain.be (corresponding author), plemmons@wfu.edu. Nicolas Gillis is a research fellow of the Fonds de la Recherche Scientifique (F.R.S.-FNRS). Research by Robert Plemmons is supported in part by the U.S. Air Force Office of Scientific Research (AFOSR), with award number FA9550-08-1-0151. The scientific responsibility is assumed by the authors.

Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI, edited by Sylvia S. Shen, Paul E. Lewis, Proc. of SPIE Vol. 7695, 76951A · © 2010 SPIE
CCC code: 0277-786X/10/$18 · doi: 10.1117/12.849345

i.e., the spectral signature of each pixel (m^i, a row of M) is approximated with a nonnegative linear combination (with weights u_k^i, representing abundances) of end-members signatures (v_k, columns of V) which hopefully correspond to the signatures of the constituent materials of the hyperspectral image.

(NMF) is an additive linear model for nonnegative data and has been observed to be particularly well-suited to achieve a parts-based and sparse representation, enhancing interpretability of the decomposition. This model* has been successfully applied for identification of the materials and the spectral unmixture in hyperspectral images.[5,6] However, NMF features some drawbacks. In particular,

1. NMF is a NP-hard nonlinear optimization problem with many local minimizers;[7] In practice, NMF is solved using iterative schemes based on nonlinear optimization techniques, see[8] and references therein.
2. The optimal solution is in general non-unique† which makes the problem ill-posed;[9] Additional constraints are often added to reduce the degrees of freedom, e.g., smoothness,[5] sparsity,[10] orthogonality,[11,12] minimum-volume,[13] sum-to-one constraint of the rows of U,[14] etc.
3. One needs to recompute a solution from scratch when the rank of the approximation is modified.

In this paper, we use Nonnegative Matrix Underapproximation, a new variant of NMF which overcomes some of its drawbacks‡ (2. and 3. above), as a dimensionality reduction technique to analyze hyperspectral data. In Section 2, we formulate NMU as an optimization problem using ℓ_2-norm minimization, and present an algorithm to solve it. We then give some theoretical evidences that NMU is in fact able to detect materials in hyperspectral data and illustrate this on a simple example. In Section 3, we explain why ℓ_1-norm based minimization is theoretically more appealing since it is potentially able to extract the materials in hyperspectral data in a more efficient and robust way. An algorithm is proposed with the same computational complexity as the one presented in Section 2. Finally, in Section 4, we experimentally show the efficiency of these new strategies on hyperspectral images associated with space object material identification, and on HYDICE remote sensing images.

NOTATION 1. *$\mathbb{R}^{m\times n}$ is the set of real matrices of dimension m by n; for $A \in \mathbb{R}^{m\times n}$, we note a_i the i^{th} column of A, a^j the j^{th} row of A, and a_i^j the entry at position (i,j); for $b \in \mathbb{R}^{m\times 1} = \mathbb{R}^m$, we note b_i the i^{th} entry of b. $\mathbb{R}_+^{m\times n}$ is the set $\mathbb{R}^{m\times n}$ with component-wise nonnegative entries.* $\mathrm{supp}(x)$ *denotes the support of x, i.e., the set on nonzero entries of x; $||.||_0$ is the ℓ_0-'norm' where $||x||_0$ is the cardinality of* $\mathrm{supp}(x)$. *A^T is the transpose of A. $||.||_2$ is the ℓ_2-norm with $||b||_2^2 = b^T b$; $||.||_F$ is the related matrix norm called Frobenius norm with $||A||_F^2 = \sum_{i,j}(a_i^j)^2$ and $\langle A, B\rangle = \sum_{i,j} a_i^j b_i^j$ the corresponding scalar product. $||.||_1$ is the ℓ_1-norm with $||A||_1 = \sum_{i,j}|a_i^j|$.*

2. NONNEGATIVE MATRIX UNDERAPPROXIMATION

Combining the Perron-Frobenius and Eckart-Young theorems,[16] it is easy to find an optimal nonnegative rank-one approximation of a nonnegative matrix. Therefore, the rank-one NMF problem can be solved in polynomial time (e.g., taking the absolute value of the first rank-one factor generated by the singular value decomposition). One would then be tempted to use this result to compute a NMF one rank-one factor at a time. However, when the first rank-one approximation is subtracted from the original matrix, we obtain a residual which contains negative entries which makes the recursive approach unpractical. Adding underapproximation constraints makes this idea possible; solving at each step

$$\min_{x\in\mathbb{R}_+^m,\, y\in\mathbb{R}_+^n} \quad ||M - xy^T||_F^2 \quad \text{such that} \quad xy^T \leq M, \tag{NMU}$$

*(NMF) is closely related to an older approach based on the geometric interpretation of the distribution of spectral signatures: they are located inside a low-dimensional simplex which vertices are the pure pixel signatures (i.e., the signatures of each individual material).[3,4]

†Any invertible matrix D such that $VD \geq 0$ and $D^{-1}W \geq 0$ generates an equivalent solution.

‡Unless P = NP, drawback 1. can not be 'resolved' since the underlying problem of spectral unmixing is of combinatorial nature[15] and can be shown to be equivalent to a NP-hard problem.[7]

and ending up with $R = M - xy^T \geq 0$ which can be underapproximated as well, etc. This problem is referred to as Nonnegative Matrix Underapproximation and has been introduced in.[17,18] It has been shown to achieve better part-based representation of nonnegative data because the underapproximations constraints require the extracted part to really be common features of the original data. We will see how this property enables NMU to extract constitutive materials in hyperspectral images.

Since only a rank-one matrix is computed at each step, (NMU) is in general well-posed in the sense that the optimal solution is unique (up to a scaling factor)[§]. In fact, for any rank-one nonnegative matrix A, there exists one and only one $(u, v) \geq 0$ such that $||u||_2 = 1$ and $A = uv^T$. In our experiments, we observed that NMU is much less sensitive to initialization and that, in general, when we allow several restarts of the algorithm with different initializations, it ends up with similar solutions (hopefully, close to the optimum). This is (in general) not the case with the standard NMF formulation because of non-uniqueness.[9]

2.1 Algorithm for Rank-One Nonnegative Underapproximation

(NMU) is convex in x and y separately, and the corresponding optimal solutions can actually be trivially computed:

$$x^* = \text{argmin}_{x \geq 0, xy^T \leq M} ||M - xy^T||_F, \quad x_i^* = \min_{\{j \,|\, y_j \neq 0\}} \left\{ \frac{m_j^i}{y_j} \right\} \forall i, \tag{2.1}$$

and,

$$y^* = \text{argmin}_{y \geq 0, xy^T \leq M} ||M - xy^T||_F, \quad y_j^* = \min_{\{i \,|\, x_i \neq 0\}} \left\{ \frac{m_j^i}{x_i} \right\} \forall j, \tag{2.2}$$

and correspond to the stationary conditions of (NMU). However, alternating optimization does not generate satisfactory results: the algorithm will stop after one or two updates of x and y and it is not able to locate good solutions; potentially far away from the initialization. The reason is that feasibility is imposed at each step, and that solutions are rapidly blocked on the boundary of the feasible domain.

A Lagrangian relaxation scheme has been proposed[18] to solve (NMU). It works as follows: let $\Lambda \in \mathbb{R}_+^{m \times n}$ be the Lagrangian multipliers associated with the underapproximation constraints and write the corresponding Lagrangian dual problem as:

$$\sup_{\Lambda \in \mathbb{R}_+^{m \times n}} \quad \min_{\substack{x \in \mathbb{R}_+^m \\ y \in \mathbb{R}_+^n}} \quad ||M - xy^T||_F^2 + 2 \left\langle \Lambda, M - xy^T \right\rangle = ||(M - \Lambda) - xy^T||_F^2 - ||\Lambda||_F^2. \tag{2.3}$$

A possible way to solve (2.3) is to alternate optimization[¶] over x, y and Λ: the optimal solution for x and y can be written in closed-form (cf. steps 6 and 7 of Algorithm 1) while Λ is updated with a subgradient type update (step 10). If Λ is too large, it might happen that x and/or y are set to zero leading to a trivial stationary point. We propose to reduce the value of Λ if that happens and to set x and y to their old values (step 12). (x, y) are initialized with the optimal rank-one solution of the unconstrained problem (i.e., the optimal rank-one approximation of the residual, step 2, corresponding to $\Lambda = 0$); Λ is initialized with the nonnegative part of the residual matrix (step 4). Since the algorithm is not guaranteed to generate a feasible solution[‖], only the nonnegative part of the residual is considered (step 15). Note that the updates of x and y share some similarities with the power method (applied to $M - \Lambda$, with projection on the nonnegative orthant) which computes the maximum singular value and its corresponding left and right singular vectors.

[§]Note that (NMF) with $r = 1$ is also well-posed; in fact, the optimal solution is unique if the maximum singular value of M ($\sigma_1(M)$) is strictly greater than the second biggest singular value ($\sigma_2(M) < \sigma_1(M)$), cf. singular value decomposition.[16]

[¶]The problem of optimizing both x and y reduces to a rank-one nonnegative factorization problem (same problem as (NMF), where M might have negative entries; in this case $M - \Lambda$) which is NP-hard.[19]

[‖]This feature is actually an advantage for practical applications. In fact, this gives the algorithm some flexibility when dealing with noisy data. However, one can obtain a feasible stationary point by using updates (2.1) and (2.2) as a post-processing step.

Algorithm 1 Lagrangian NMU (L-NMU)

Require: $M \in \mathbb{R}_+^{m\times n}$, $r > 0$, maxiter.
Ensure: $(U, V) \in \mathbb{R}_+^{m\times r} \times \mathbb{R}_+^{n\times r}$ s.t. $UV^T \lesssim M$.

for $k = 1 :$ r **do**
  $[x, y] =$ optimal rank-one approximation(M);
  $u_k \leftarrow x$; $v_k \leftarrow y$;
  $\Lambda \leftarrow \max(0, -(M - xy^T))$;
  for $p = 1 :$ maxiter **do**
    $x \leftarrow \max\left(0, \frac{(M-\Lambda)y}{||y||_2^2}\right)$;
    $y \leftarrow \max\left(0, \frac{(M-\Lambda)^T x}{||x||_2^2}\right)$;
    if $x \neq 0$ and $y \neq 0$ **then**
      $u_k \leftarrow x$; $v_k \leftarrow y$;
      $\Lambda \leftarrow \max(0, \Lambda - \frac{1}{p}(M - xy^T))$;
    else
      $\Lambda \leftarrow \frac{\Lambda}{2}$; $x \leftarrow u_k$; $y \leftarrow v_k$;
    end if
  end for
  $M = \max(0, M - u_k v_k^T)$;
end for

2.2 Hyperspectral Data Analysis in the Ideal Case

If we assume that each pixel contains only one material, the corresponding matrix has the following form:

ASSUMPTION 1. *$M \in \mathbb{R}_+^{m\times n}$ with $M = UV^T$ where*

1. *$U \in \{0, 1\}^{m\times r}$ is a binary matrix of dimension m by r, with $r \leq \min(m, n)$, and its columns are orthogonal:*

$$u_i^T u_j = 0, \forall i \neq j \quad \text{and} \quad u_i^T u_i \neq 0, \forall i,$$

   i.e., there is one and only one nonzero element in each row of U and

$$u_k^i = 1 \iff \text{pixel } i \text{ contains material } k.$$

2. *$V \in \mathbb{R}_+^{n\times r}$ is full-rank.*

Of course, recovering U and V in these settings is trivial and, in practice, because of blurring effects, limited resolution and mixed materials, the spectral signature of each pixel will be a mixture of spectral signatures of several materials (in particular, pixels located at the boundary of materials) plus noise. However, classifying each pixel in a single category amounts to approximating M with a matrix satisfying Assumption 1. This problem is referred to as orthogonal NMF (oNMF) and is equivalent to k-means clustering.[11]

We now show that the underapproximation technique is able to retrieve the underlying structure in the ideal case, when each pixel corresponds to only one material. It will shed some light on the behavior of the above recursive algorithm based on underapproximations and justify its efficiency when dealing with non-ideal hyperspectral images.

2.2.1 First Rank-One Factor

As for PCA, the first rank-one factor of NMU will reduce the error the most and will already be a fairly good approximation of the hyperspectral data.

LEMMA 2.1. *Let (x, y) be a nontrivial stationary point of* (NMU) *(i.e., $x \neq 0$ and $y \neq 0$), then the residual $R = M - xy^T$ has at least one zero by row and by column.*

Proof. It follows directly from Equations (2.1) and (2.2). □

LEMMA 2.2. *Let (x, y) be a nontrivial stationary point of* (NMU) *for $M = UV^T$ satisfying Assumption 1, then the residual $R = M - xy^T$ can be written as $R = UV'^T$ for some $V' \geq 0$.*

Proof. Because columns of U are binary and orthogonal, each row of M is equal to a column of V. Therefore, the entries of x corresponding to the rows of M equal to each other must take the same value, i.e., $\forall i \in \{1, 2, \ldots, r\}, \forall k, l \in \text{supp}(v_i) : \; x_k = x_l$. In fact, one can check that for $y \neq 0$, solution of Equation (2.1) is unique. It follows that $x = Ud$, for some $d \in \mathbb{R}^r_+$, and then $R = UV^T - Udy^T = U[V - yd^T]^T$. The facts that R is nonnegative and that U is orthogonal implies that $V' = [V - yd^T]^T \geq 0$. □

COROLLARY 1. *Let (x, y) be a nontrivial stationary point of* (NMU) *and $M > 0$, then $x > 0$ and $y > 0$. Moreover, the residual $R = M - xy^T$ can be written as $R = UV'^T$ for some $V' \geq 0$ with at least one zero by row and by column in V'.*

Proof. Positivity of x and y follows directly from Equations (2.1) and (2.2) while structure of the residual matrix R is a consequence of Lemma 2.1 and 2.2. □

Typically, it is very unlikely for two columns of V' to have the same support. In fact, the number of zeros in each column of V' is at least in $O(\frac{n}{k})$ since there is a zero in each row of V' and, for two columns to have the same sparsity pattern I, one needs

$$v_i(I) = \alpha \, v_j(I), \text{ since } v_i(I) - d_i y(I) = v_j(I) - d_j y(I) = 0, \tag{2.4}$$

for some constant $\alpha > 0$. If $|I| > 1$ (i.e., v'_i and v'_j have more that one zero element which is likely since they are expected to have $O(\frac{n}{k})$ zeros, $n \gg r$), it is not likely for Equation (2.4) to happen because v_i and v_j are linearly independent. For example, if v_i and v_j were randomly generated, the probability would be equal to zero.

Conclusion. After the first NMU reduction, the residual R can be written in the same form as $M = UV^T$ (cf. Assumption 1) with $R = UV'^T$, and it is highly probable that the columns of V' will have different sparsity patterns.

2.2.2 Next Rank-One Factors

Assuming that the columns of V in Assumption 1 have different sparsity patterns, we show that the recursion outlined above will eventually locate each material individually.

THEOREM 2.3. *Let (x, y) be a nontrivial stationary point of* (NMU) *for $M = UV^T$ satisfying Assumption 1 and the columns of V have different sparsity patterns. Then $R = M - xy^T = UV'$, with $x = Ud$ for some $d \in \mathbb{R}^r_+$ so that $V' = V - yd^T \geq 0$. Moreover,*

$$\text{supp}(x) = \cup_{i \in \Omega} \text{supp}_i(u_i), \quad \textit{for some } \Omega \subset \{1, 2, \ldots, r\},$$

and

$$|\Omega| = 1 \iff \textit{one column of } V' \textit{ is equal to zero} \iff y = \alpha \, v_i \textit{ for some } i, \alpha > 0.$$

Proof. The first part is a consequence of Corollary 1.

Clearly, v'_i is equal to zero for some i if and only if $y = \alpha v_i$ for some positive constant α. Moreover, $y = \alpha v_i$ implies that $|\Omega| = 1$ because of the underapproximation constraints and since the columns of V have different sparsity patterns, cf. Equation (2.1). Finally, it is clear that for $|\Omega| = 1$, the solution obtained with Equation (2.2) is $y = \alpha v_i$. □

Theroem 2.3 implies that, at each step of the NMU recursion, a set of materials are extracted together. Moreover, since the recursive approach outlined above will eventually end up with a zero matrix (say, after r_u steps), we will have that

$$M = \sum_{i=1}^{r_u} x_i y_i^T,$$

and, under the different sparsity pattern assumption, $\forall 1 \leq i \leq r, \exists 1 \leq j \leq r_u \quad \text{s.t.} \quad \text{supp}(x_j) = \text{supp}(u_i)$. In fact, for the residual $R = UV'$ to be equal to zero, all the columns of V' must be identically zero. This feature of the NMU recursion will be experimentally verified in Section 4.

REMARK 1. *The different sparsity pattern assumption is a sufficient but not a necessary condition for exact recovery. In fact, if two columns of V' have the same sparsity pattern and they are extracted together, it is likely that the corresponding optimal solution will not be exactly equal to one of these two columns (because there are linearly independent) and therefore, at the next step, they will have a different sparsity pattern.*

2.3 Illustration of Basis Recovery with NMU

Let us construct the following synthetic data: 4 binary orthogonal images of 5×5 pixels (which are the columns of U, $U \in \{0,1\}^{25\times 4}$, see left image of Figure 2) are randomly mixed ($V \in \mathbb{R}^{25\times 4}$ is randomly generated) to generate a 25×25 matrix $M = UV^T$ satisfying Assumption 1. Figure 1 displays a sample of the 25 images contained in the columns of M, which then result from the nonnegative linear combination of the columns of U. Figure 2 displays the original images and the basis elements obtained with NMF and NMU. We observe that

Figure 1. Sample of images of the data matrix M: clean (left) and with mixed pixels (right).

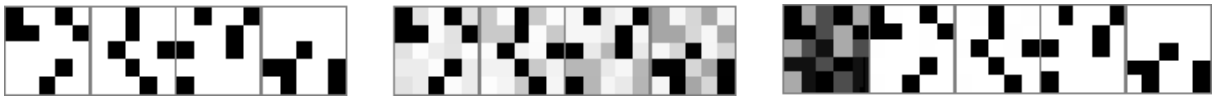

Figure 2. From left to right: 4 original images (i.e., columns of U), basis elements obtained with NMF and with NMU.

NMF is not able to extract perfectly the 4 original basis elements (even though the objective function is equal to zero; the reason is the non-uniqueness of the solution: NMF retrieves a mixture of the basis elements) while NMU is able to do the extraction.

2.4 Hyperspectral Data Analysis in the Non-Ideal Case

As we already mentioned, practical problems don't have the nice structure mentioned in Assumption 1 and the spectral signature of most pixels results from an additive linear combination of several materials. What can we expect of NMU in that case? Since the data matrix is positive, the first rank-one factor will still be a mixture of all materials (cf. Lemma 2.1). It seems more difficult to provide theoretical guarantees for the next factors in more general settings and this will be a topic for further research. However, extracting a single constitutive material would allow one to approximate all the pixels containing it (removing from their spectral signature this component) and, since NMU aims at extracting components explaining the data as closely as possible in order to reducing the error the most, this indicates that NMU is also incited to extract constitutive materials in non-ideal cases.

For example, let us add to the matrix U in the illustration of the previous paragraph a randomly generated matrix, uniformly distributed between 0 and 0.5. It means that each pixel is now a mixture of several materials but one material is still predominant. Figure 3 displays the visual results: NMF performs even worse, while NMU is still able to extract the original parts fairly well. It actually gives a soft clustering for each pixel, as it will also be shown in Section 4.

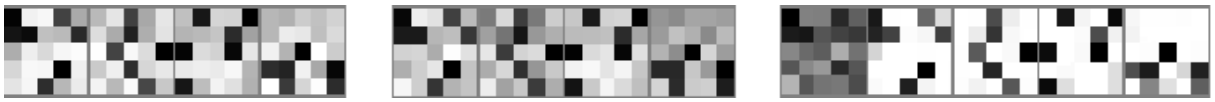

Figure 3. From left to right: 4 original images (i.e., columns of U), basis elements obtained with NMF and with NMU.

3. ℓ_0-PSEUDO-NORM MINIMIZATION AND ℓ_1-NORM RELAXATION

Ideally, each basis element extracted with the recursive approach outlined above should correspond to a different material present in the hyperspectral image: we would like that each extracted rank-one factor corresponds to only one material, i.e., that only a submatrix of M (a set of rows of M) corresponding to pixels containing the same material is approximated at each step. Unfortunately, the ℓ_2-norm is not appropriate for this purpose: it is very sensitive to 'outliers', i.e., it cannot neglect some entries of the matrix M and set only a subset of the entries of the residual error to zero. It is more likely that it will try to approximate several materials at the same time in order to avoid large entries in the residual error. For this reason, we will see that the ℓ_2-norm based algorithm (Algorithm 1) first extracts (in general) several materials together.

If the ℓ_0-'norm' is used instead, i.e., if the number of zero entries in the residual is maximized, one can check that for a matrix satisfying Assumption 1, this would lead to an exact recovery in r steps; because extracting one material (i.e., taking $y = v_i$ for some i at each step) will lead to the highest number of zeros in the residual $R = M - xy^T$ (rows corresponding to the extracted material are identically zero; plus one zero by row and by column for the other ones). Unfortunately, ℓ_0-'norm' minimization is nonconvex, even when one factor is fixed (i.e., $||M - xy^T||_0$ for x or y fixed). Moreover, in practice, because of noise and blur, ℓ_0-'norm' would not be appropriate since rows of M representing the same material cannot be approximated exactly. However, its convex relaxation, the ℓ_1-norm, is known to be less sensitive to outliers, i.e., it is disposed to let some entries of the error large, in order to approximate better other entries. We will experimentally observe in Section 4 that using ℓ_1-norm allows to extract materials individually in a more efficient manner, i.e., using a smaller number of recursive steps.

3.1 Algorithm for ℓ_1-Norm Minimization

Using the idea of Lagrangian duality presented in Secton 2.1, we propose to solve**

$$\max_{\Lambda \in \mathbb{R}_+^{m \times n}} \quad \min_{x \in \mathbb{R}_+^m,\, y \in \mathbb{R}_+^n} \quad ||(M - \Lambda) - xy^T||_1. \tag{3.1}$$

Fixing y and Λ and noting $A = M - \Lambda$, x can be optimized by solving the following m independent problems

$$\min_{x_i \geq 0} ||a^i - x_i y||_1 = \sum_j |a^i_j - x_i y_j| = \sum_{j \,\in\, \mathrm{supp}(y)} y_j \Big| \frac{a^i_j}{y_j} - x_i \Big| + \sum_{j \,\notin\, \mathrm{supp}(y)} |a^i_j|, \tag{3.2}$$

which can be solved by computing the weighted median of z with $z_j = (a^i_j / y_j)\, \forall j$ with weights y_j. The same can be done for y by symmetry, and we propose to replace updates of x and y in Algorithm 1 by

$$x_i = \max\Big(0, \text{weighted-median}\Big(\frac{[(M - \Lambda)^i(J)]}{[y(J)]}, y(J)\Big)\Big)\ \forall i, \quad J = \mathrm{supp}(y),$$

and

$$y_j = \max\Big(0, \text{weighted-median}\Big(\frac{[(M - \Lambda)_j(I)]}{[y(I)]}, y(I)\Big)\Big)\ \forall j, \quad I = \mathrm{supp}(x).$$

The weighted median of a n dimensional vector can be computed in $O(n)$ operations, cf.[20] and references therein, so that the algorithm can be implemented in $O(mn)$ operations per iteration. The ℓ_2-norm and ℓ_1-norm then have the same computational complexity even though in practice the ℓ_1-norm based algorithm will be slower, but only up to a constant factor.

**Note that Λ does not correspond to the Lagrangian dual variables of $\min_{x \geq 0, y \geq 0, xy^T \leq M} ||M - xy^T||_1$. However, this formulation is closely related to the Lagrangian relaxation and allows us to use the same derivations as in Algorithm 1.

4. APPLICATION TO HYPERSPECTRAL DATA

In this section, Algorithm 1 (ℓ_2-norm based minimization) and its modification for ℓ_1-norm proposed in Section 3 are used as dimensionality reduction techniques for hyperspectral data in order to achieve *classification* (selecting from the basis elements the different clusters), and *spectral unmixing* (using nonnegative least squares). We analyze the Urban HYDICE images and the Hubble space telescope images. We perform up to 100 iterations of both algorithms (i.e., maxiter = 100) at each step of the recursion.

4.1 Classification and Spectral Unmixing

NMU can be used as a standard dimensionality reduction technique and any type of post-processing procedure can be used to extract the constitutive parts of spectral data, e.g., k-means, nearest neighbor, etc. However, we have shown why NMU is potentially able to extract these parts automatically. Therefore, the simplest approach would be to visually select each cluster from the generated basis elements. We stick to this approach and select, from the basis elements, each individual cluster: from the U matrix obtained with NMU, we only keep a subset of the columns, each corresponding to an individual material.

The second post-processing step is to normalize U. In fact, as NMF, NMU is invariant to the scaling of the columns of U ($\forall k\ u_k v_k^T = (\alpha u_k)(\alpha^{-1} v_k)^T\ \forall \alpha > 0$). Moreover, in the context of hyperspectral image analysis, rows of U have a physical interpretation: u_j^i is the abundance of material j in pixel i. Therefore, $u_j^i \leq 1\ \forall i, j$ and the columns of U are normalized with

$$u_j = \frac{u_j}{\max_i(u_j^i)} \quad \forall j.$$

This means that, for each rank-one factor extracted with the NMU procedure, the maximum abundance of each pixel for the corresponding spectral signature v_k is at most 1. Moreover, since rows of U correspond to abundances, $\sum_j u_j^i = 1\ \forall i$ and we can scale the rows of U as follows

$$u^i \leftarrow \frac{u^i}{||u^i||_1 + \epsilon}, \quad \epsilon \lll 1,$$

so that they sum to one (except if they are identically zeros). This allows us to equilibrate the relative importance of each pixel in each basis element. With this procedure, we end up with a soft clustering: each pixel i is constituted of several materials v_j's with the corresponding abundances given by u^i.

Once the pixels have been classified, one might be interested in recovering the spectral signatures of each individual material (corresponding to the columns of matrix V in the decomposition $M \approx UV^T$), called the end-members. A standard approach is to solve a nonnegative least squares problem of the form

$$\min_{V \geq 0} ||M - UV^T||_F^2, \tag{NNLS}$$

where U represents the basis vectors, with dedicated algorithms.[21]

4.2 Urban HYDICE Image

We consider first the Urban hyperspectral image[††] taken with HYper-spectral Digital Imagery Collection Experiment (HYDICE) air-borne sensors. We analyze the data where the noisy bands have been removed (162 bands left, originally 210). Figures 4 and 5 gives some basis elements of the NMU decomposition based on ℓ_2- and ℓ_1-norms. The Urban data is mainly constituted of 6 types of materials: road, dirt, trees, roofs, grass and metal.[22] Table 1 gives the index of the NMU basis elements corresponding to a single material. Figure 6 displays the results of the spectral unimxing procedure for both NMU algorithms (ℓ_2 and ℓ_1) which are compared to 6 end-members obtained in.[22]

In this example, NMU performs relatively well and is able to detect all the materials individually, which can then be used to classify the pixels and finally recover the end-member signatures. We also note that, as announced, the ℓ_2-norm based algorithm needs more recursion than ℓ_1-norm to extract all materials (23 vs. 17).

[††]Available at *http://www.agc.army.mil/hypercube/*.

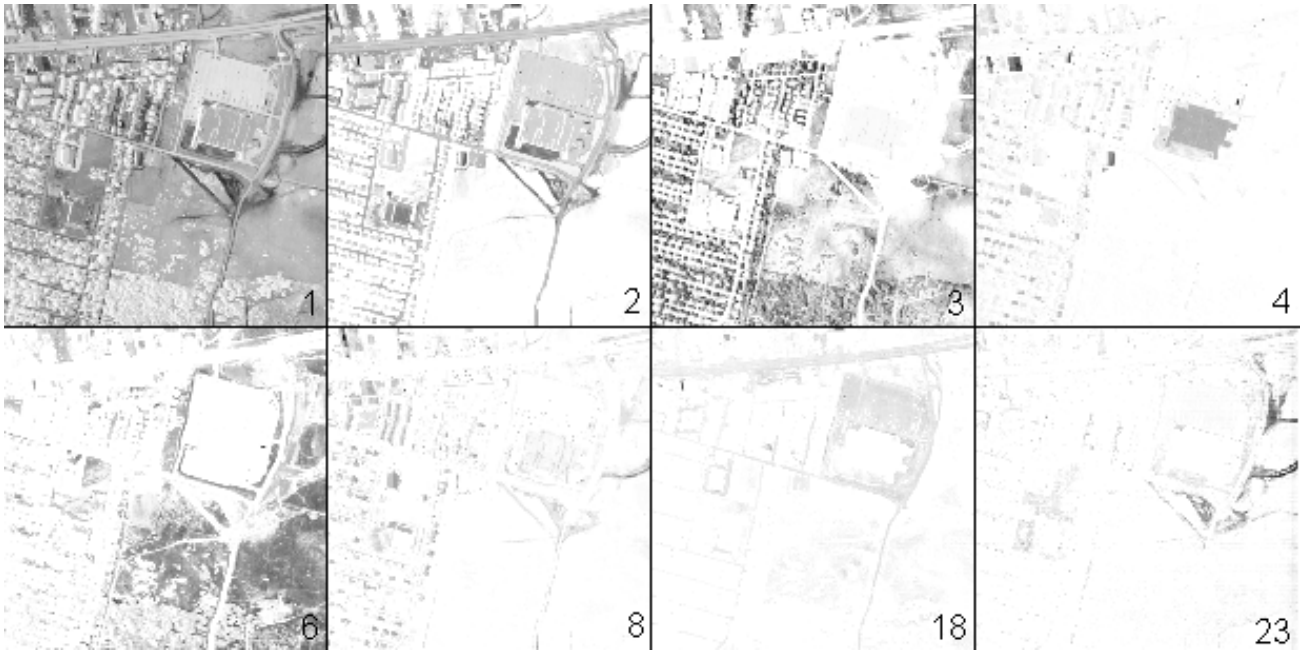

Figure 4. Some basis elements (columns of matrix U) for the Urban dataset extracted with the ℓ_2-norm based NMU; dark tones indicate a high value of an entry (0 is white, 1 black); numbers indicate the position of the factor in the NMU decomposition.

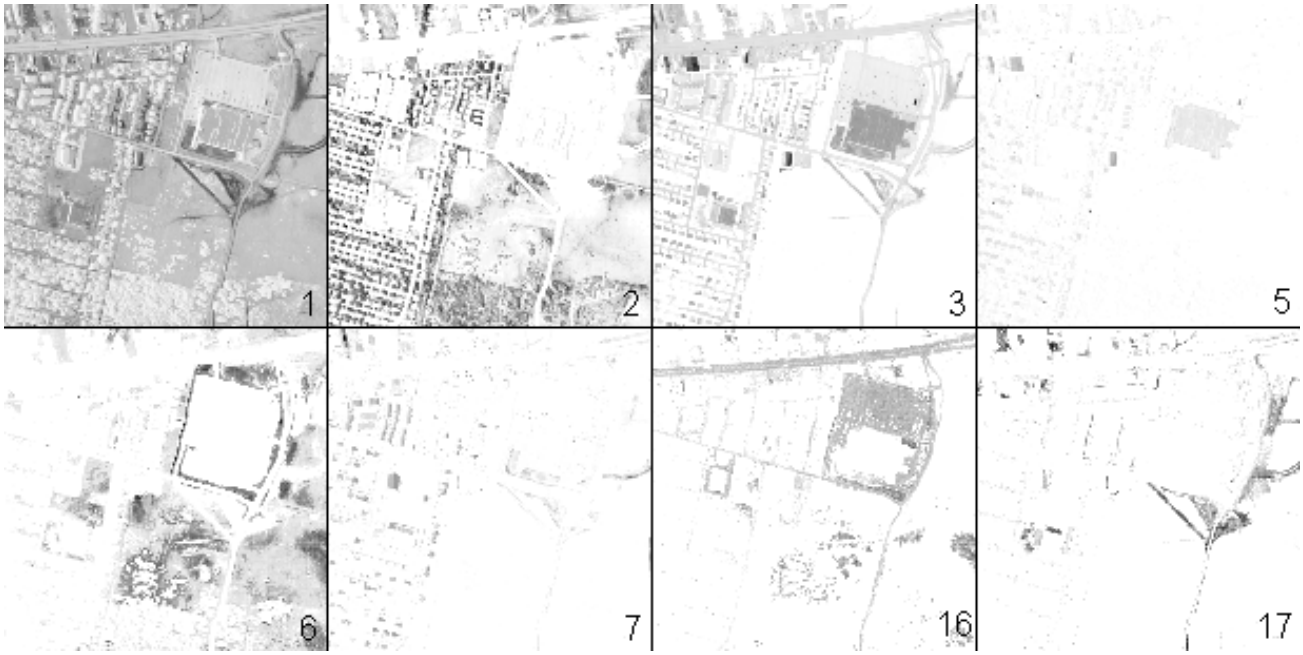

Figure 5. Some basis elements of NMU for the Urban dataset extracted with the ℓ_1-norm based NMU, identified as in Figure 4.

4.3 Simulated Hubble Space Telescope

Figure 7 displays some sample images of the Hubble database which consists of 100 spectral images of the Hubble telescope, 128×128 pixels each, with added blur and noise[‡‡].[5] It is constituted of 8 materials, see Figure 8.

[‡‡]Point spread function on 5 by 5 pixels and with standard deviation of 1 and white Gaussian noise $\sigma = 1\%$ of the values of M and Poisson noise $\sigma = 1\%$ of the mean value of M.

Clusters	Road	Dirt	Trees	Roofs	Grass	Metal
ℓ_2 basis #	18	23	3	4	6	8
ℓ_1 basis #	16	17	2	5	6	7

Table 1. Basis elements obtained: cluster selection.

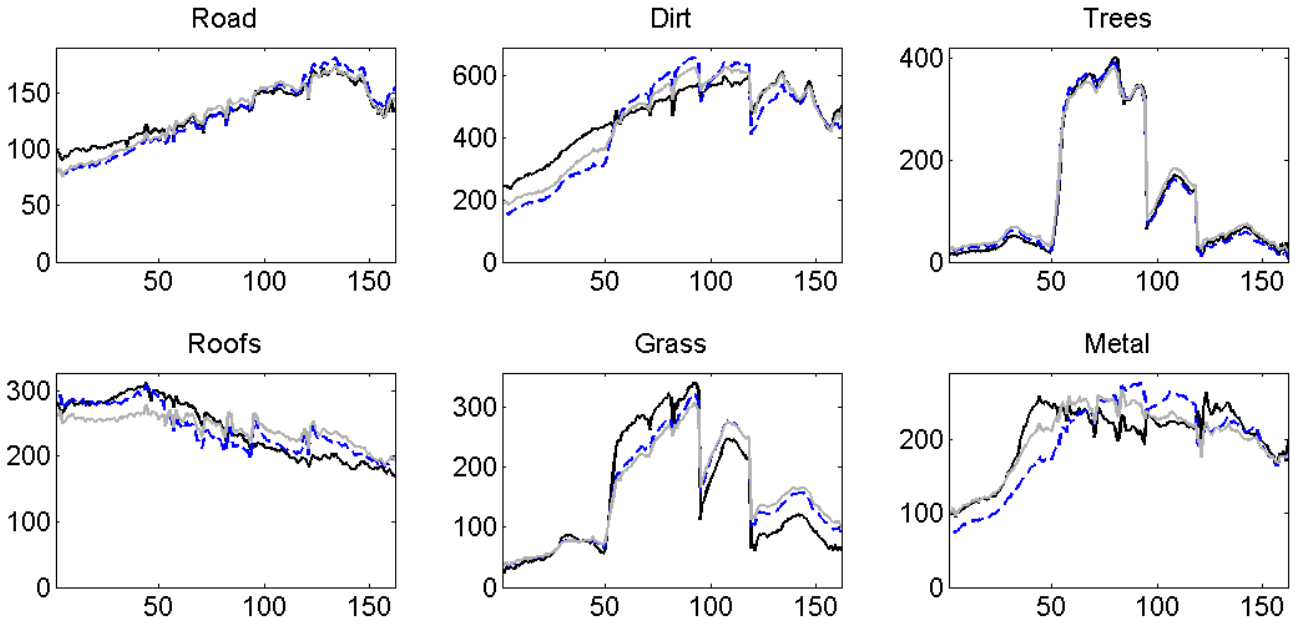

Figure 6. End-members extraction: NMU with ℓ_2 (gray solid) and NMU with ℓ_1 (dashed) vs. 6 end-members from the image using N-FINDR5[23] plus manual adjustment (dark solid) from.[22] The x-axis gives the wavelength bands while y-axis gives the reflectance values (intensities of reflected light).

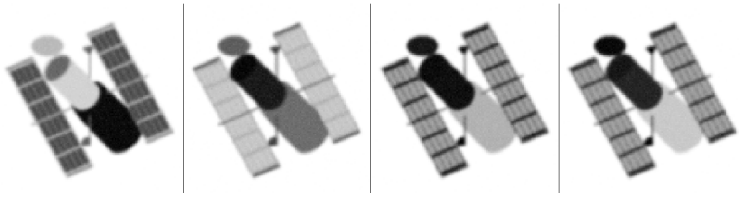

Figure 7. Sample of images in the Hubble tensor with blur and noise.

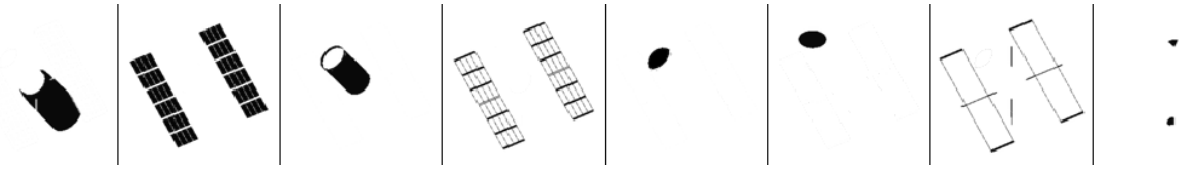

Figure 8. 8 Materials of the Hubble teleschope. From left to right: Aluminum, Solar Cell, Green Glue, Copper Stripping, Honeycomb Side, Honeycomb Top, Black Rubber Edge and Bolts.

Figure 9 shows the basis elements obtained with NMU, Table 2 gives the classification of the basis elements and Figure 10 shows the end-members extraction: original vs. noisy and blurred. Spectral signatures of black rubber edge and bolts are not recovered very accurately (or not at all in the case of the ℓ_2-norm). The reason is that they are the smallest and thinest parts: they get mixed with surrounding materials which make them difficult to extract. Moreover, for the bolts, its spectral signature is very similar to the one of copper stripping

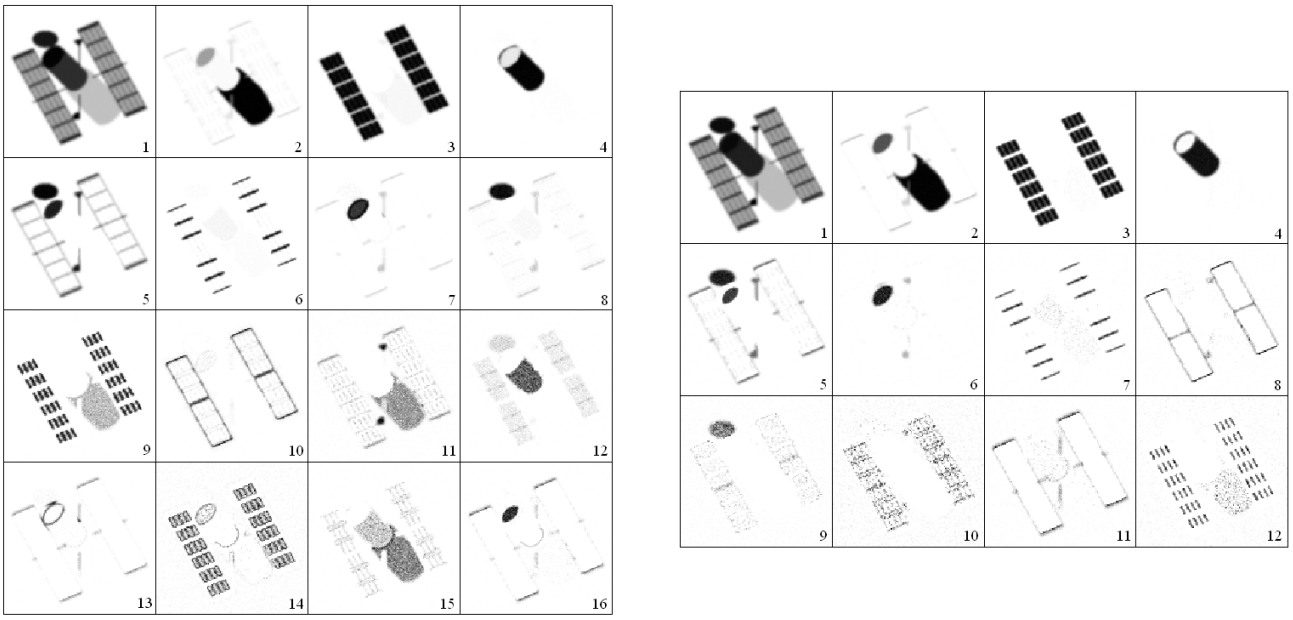

Figure 9. Basis elements of NMU for Hubble telescope with ℓ_2-norm (left) and ℓ_1-norm (right) with added blur and noise.

Clusters	Aluminum	Solar Cell	Glue	Copper	H. Side	H. Top	Edge	Bolts
ℓ_2 basis #	2	3	4	6	7	8	13	11
ℓ_1 basis #	2	3	4	7	6	9	11	8

Table 2. Basis elements obtained: cluster selection for the Hubble telescope database with noise and blur, see Figure 9.

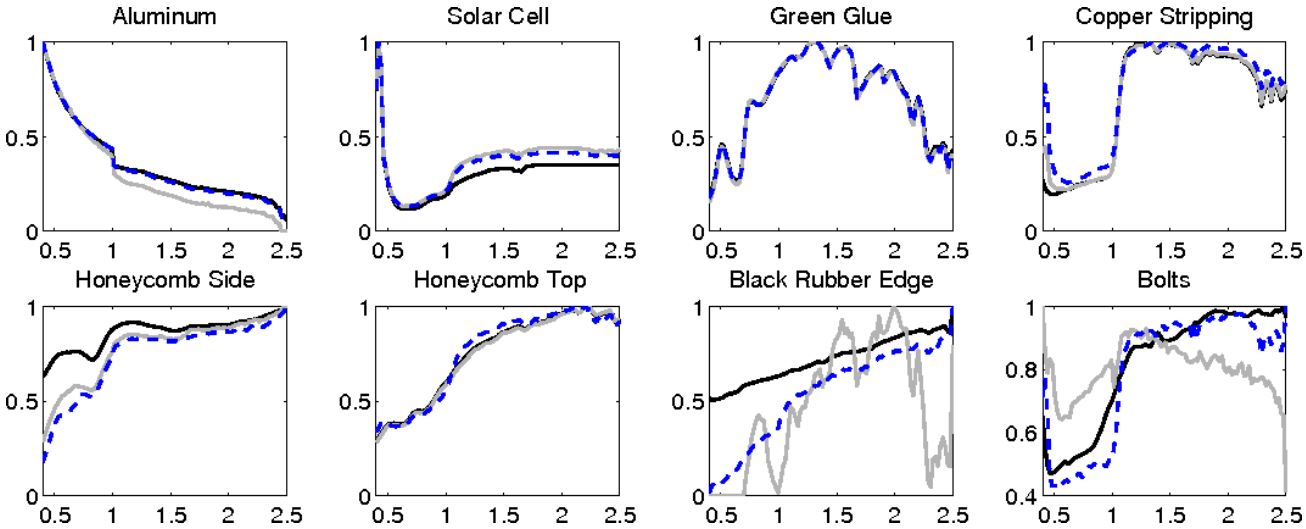

Figure 10. End-Members with noise and blur. NMU with ℓ_2 (gray solid) and NMU with ℓ_1 (dashed) vs. 8 true end-members (black solid).

and therefore, when noise and blur are added, they are extracted together (basis elements 11 for ℓ_2-norm and 8 for ℓ_1-norm).

As for the Urban dataset, the ℓ_2-norm extracts more mixed materials and therefore needs more recursions to get all the parts separated than the ℓ_1-norm, which seems to do a better job (especially for the black rubber edge).

5. SUMMARY AND FURTHER WORK

In this paper, we have presented the approximate nonnegative matrix factorization problem with underapproximation constraints, called Nonnegative Matrix Underapproximation (NMU), which can be used to design a recursive procedure to approximate nonnegative data. We then gave theoretical and experimental evidences showing that NMU is able to perform soft-clustering of hyperspectral data. A main advantage of NMU is that no sparsity parameters have to be tuned and parts-based representation is naturally achieved.

It would be interesting to compare NMU with other dimensionality reduction techniques such as PCA, NMF, ICA, etc. Another direction of research would be to design an automatic classification algorithm, based on the properties of NMU, to classify the pixels. It would be particularly interesting to study these properties in more depth and see if it possible to obtain stronger theoretical guarantees for the factors generated by NMU. In future work, comparisons of our NMU method will be made with the recent development of variational iterative methods for deblurring, denoising, and segmentation by Li, Ng, and Plemmons.[24]

REFERENCES

[1] Lee, D. and Seung, H., "Learning the parts of objects by nonnegative matrix factorization," *Nature* **401**, 788–791 (1999).

[2] Berry, M., Browne, M., Langville, A., Pauca, P., and Plemmons, R., "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics and Data Analysis* **52**, 155–173 (2007).

[3] Boardman, J., "Geometric mixture analysis of imaging spectrometry data," in [*Proc. IGARSS 4, Pasadena, Calif., pp. 2369–2371*], (1994).

[4] Craig, M., "Minimum-volume tranforms for remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing* **32 (3)**, 542–552 (1994).

[5] Pauca, P., Piper, J., and Plemmons, R., "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra and its Applications* **406 (1)**, 29–47 (2006).

[6] Zhang, Q., Wang, H., Plemmons, R., and Pauca, P., "Tensor methods for hyperspectral data analysis: a space object material identification study," *J. Optical Soc. Amer. A* **25 (12)**, 3001–3012 (2008).

[7] Vavasis, S. A., "On the complexity of nonnegative matrix factorization," *SIAM Journal on Optimization* **20**(3), 1364–1377 (2009).

[8] Cichocki, A., Amari, S., Zdunek, R., and Phan, A., [*Non-negative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*], Wiley-Blackwell (2009).

[9] Laurberg, H., Christensen, M., Plumbley, M., Hansen, L., and Jensen, S., "Theorems on positive data: On the uniqueness of NMF," *Computational Intelligence and Neuroscience* **2008**. Article ID 764206.

[10] Hoyer, P., "Nonnegative matrix factorization with sparseness constraints," *J. Machine Learning Research* **5**, 1457–1469 (2004).

[11] Ding, C., Li, T., Peng, W., and Park, H., "Orthogonal nonnegative matrix tri-factorizations for clustering," in [*Proc. SIGKDD, Int. Conf. on Knowledge Discovery and Data Mining*], 126–135 (2006).

[12] Li, H., Adal, C., Wang, W., Emge, D., and Cichocki, A., "Non-negative matrix factorization with orthogonality constraints and its application to raman spectroscopy," *The Journal of VLSI Signal Processing* **48**, 83–97 (2007).

[13] Miao, L. and Qi, H., "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing* **45 (3)**, 765–777 (2007).

[14] Masalmah, Y. and Veléz-Reyes, M., "A full algorithm to compute the constrained positive matrix factorization and its application in unsupervised unmixing of hyperspectral imagery," in [*Proc. SPIE, Vol. 6966; doi:10.1117/12.779444*], (2008).

[15] Ifarraguerri, A. and Chang, C.-I., "Multispectral and hyperspectral image analysis with convex cones," *IEEE Transactions on Geoscience and Remote Sensing* **37(2)**, 756–770 (1999).

[16] Golub, G. and Van Loan, C., [*Matrix Computation, 3rd Edition*], The Johns Hopkins University Press Baltimore (1996).

[17] Gillis, N., *Approximation et sous-approximation de matrices par factorisation positive: algorithmes, complexité et applications*, Master's thesis, Université catholique de Louvain (2007). In French.

[18] Gillis, N. and Glineur, F., "Using underapproximations for sparse nonnegative matrix factorization," *Pattern Recognition* **43(4)**, 1676–1687 (2010).

[19] Gillis, N. and Glineur, F., "Nonnegative factorization and the maximum edge biclique problem," CORE Discussion paper 2008/64 (2008).

[20] Gurwitz, C., "Weighted median algorithms for l_1 approximation," *BIT Numerical Mathematics* **30 (2)**, 301–3106 (1990).

[21] Chen, D. and Plemmons, R., "Nonnegativity Constraints in Numerical Analysis," in [*A. Bultheel and R. Cools (Eds.), Symposium on the Birth of Numerical Analysis, World Scientific Press.*], (2009).

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

[23] Winter, M., "N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in [*Proc. SPIE Conference on Imaging Spectrometry V*], (1999).

[24] Li, F., Ng, M., and Plemmons, R., "Coupled segmentation and denoising/deblurring models for hyperspectral material identification," Available at `http://www.wfu.edu/~plemmons` (2009).