Dimensionality reduction, classification, and spectral mixture analysis using non-negative underapproximation

Nicolas Gillis
Université catholique de Louvain
Center for Operations Research and Econometrics
Department of Mathematical Engineering
Voie du Roman Pays, 34
B-1348 Louvain-La-Neuve, Belgium

Robert J. Plemmons
Wake Forest University
Departments of Mathematics and Computer
    Science
Manchester Hall 341
Winston-Salem, North Carolina 27109
E-mail: plemmons@wfu.edu

Abstract. Non-negative matrix factorization (NMF) and its variants have recently been successfully used as dimensionality reduction techniques for identification of the materials present in hyperspectral images. We study a recently introduced variant of NMF called non-negative matrix underapproximation (NMU): it is based on the introduction of underapproximation constraints, which enables one to extract features in a recursive way, such as principal component analysis, but preserving non-negativity. 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 images. Both ℓ2-norm and ℓ1-norm–based minimization of the energy functional are considered. 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. © *2011 Society of Photo-Optical Instrumentation Engineers (SPIE).* [DOI: 10.1117/1.3533025]

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 principal component analysis (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 non-negative spectral signatures of its constitutive materials. Taking these non-negativity constraints into account enhances interpretability of the extracted factors. This can be done using the non-negative matrix factorization[1] (NMF) technique, generally formulated as the following optimization problem: given a $m \times n$ real non-negative matrix M and a positive integer $r < \min(m, n)$, find two real non-negative matrices U and V of dimensions $m \times r$ and $n \times r$ in order to minimize the sum of the squared entries of $M - UV^T$,

$$\min_{U,V} \|M - UV^T\|_F^2 \quad \text{such that} \quad U \geq 0 \text{ and } V \geq 0. \quad (1)$$

NMF has been successfully used in many applications (e.g., see the survey[2]), including general clustering, general images processing, text mining, air emission control, microarray data analysis, food quality and safety analysis, face recognition, as well as multispectral and hyperspectral data analysis.

Assuming that the matrix M is constructed as follows: each 2-D 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:

$$m^i \approx \sum_k u_k^i \upsilon_k^T \quad \forall i, \quad (2)$$

i.e., the spectral signature of each pixel (m^i, a row of M) is approximated with a non-negative linear combination (with weights u_k^i, representing abundances) of end-members' signatures (υ_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 non-negative data and has been observed to be particularly well suited to achieve a parts-based and sparse representation, enhancing interpretability of the decomposition. [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]] This model has been successfully applied for identification of the materials and the spectral unmixing in hyperspectral images.[5,6] However, NMF features some drawbacks, including, in particular, the following:

1. Equation (1) is a NP-hard nonlinear optimization problem with many local minimizers.[7] In practice, Eq. (1) is solved using iterative schemes based on nonlinear optimization techniques (see Ref. 8, and references therein), and only convergence to stationary points of Eq. (1) is typically guaranteed.

2. The optimal solution is, in general, nonunique, which makes the problem ill-posed.[9] (Any invertible matrix D such that $UD \geq 0$ and $(VD^{-T}) \geq 0$
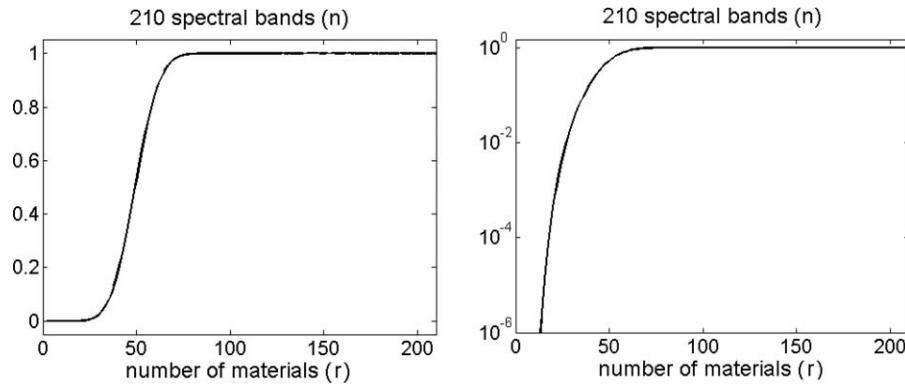
Fig. 1 Probability for at least one column of V' to have one zero element, while another column has a zero at the same position for $n = 210$, assuming that (*i*) V is randomly generated and (*ii*) V' contains n zeros which are randomly distributed with uniform distribution (with at least one zero by column).

generates an equivalent solution.) 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 non-negative matrix underapproximation (NMU), a new variant of NMF that overcomes some of its drawbacks (above-listed drawbacks 2 and 3), as a dimensionality reduction technique to analyze hyperspectral data. (Unless $P = $ NP, drawback 1 cannot be "resolved" because the underlying problem of spectral unmixing is of combinatorial nature[15] and can be shown to be equivalent to a NP-hard problem.[7]) In Sec. 2, we study NMU as an optimization problem using ℓ_2-norm minimization: we implement an algorithm proposed in Ref. 16 based on Lagrangian relaxation and propose a more robust version using a proper initialization and a safety procedure (see algorithm ℓ_2-NMU). We then give some theoretical evidence that NMU is in fact able to detect materials in hyperspectral data and illustrate this with a simple example. In Sec. 3, we explain why ℓ_1-norm–based minimization is theoretically more appealing because 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 Sec. 2 for ℓ_2-norm. Finally, in Sec. 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. Finally, a rather different type of example, related to biometric identification, is provided to illustrate the diversity of NMU applications in spectral imaging. (A preliminary and abbreviated conference version of this paper was published in Ref. 17.)

1.1 Notation

$\mathbb{R}^{m \times n}$ is the set of real matrices of dimension m by n; for $A \in \mathbb{R}^{m \times n}$, we denote a_j as the j'th column of A, a^i the i'th row of A, and a_j^i the entry at position (i, j); for $b \in \mathbb{R}^{m \times 1} = \mathbb{R}^m$, we denote b_i as the i'th entry of b. $\mathbb{R}_+^{m \times n}$ is the set $\mathbb{R}^{m \times n}$ with componentwise non-negative entries. The cardinality of a set S is denoted $|S|$. supp(x) denotes the support of x [i.e., the set on nonzero entries of x]; and $\overline{\text{supp}}(x)$ its complement (i.e.,

the sparsity pattern of x). $\|.\|_0$ is the ℓ_0-norm, where $\|x\|_0$ is the cardinality of 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 $\|A\|_F^2 = \Sigma_{i,j}(a_j^i)^2$ and $\langle A, B \rangle = \Sigma_{i,j} a_j^i b_j^i$ is the corresponding scalar product. $\|.\|_1$ is the ℓ_1-norm with $\|A\|_1 = \Sigma_{i,j} |a_j^i|$.

2 NMU

Combining the Perron–Frobenius and Eckart–Young theorems,[18] it is easy to find an optimal non-negative rank-one approximation of a non-negative 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 an 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 that contains negative entries, and this makes the recursive approach unpractical. Adding underapproximation constraints makes this idea possible; solving at each step

$$\min_{x \in \mathbb{R}_+^m, y \in \mathbb{R}_+^n} \|M - xy^T\|_F^2 \quad \text{such that} \quad xy^T \leq M \quad (3)$$

and ending up with $R = M - xy^T \geq 0$, which can be underapproximated as well, etc. This problem is referred to as an NMU and was introduced in Refs. 16 and 19. It has been shown to achieve better part-based representation of non-negative 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.

Because only a rank-one matrix is computed at each step, [Eq. (3)] is, in general, well posed in the sense that the optimal solution is unique (up to a scaling factor). [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.[18]] In fact, for any rank-one non-negative 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
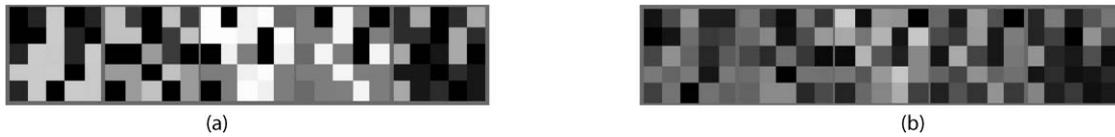
Fig. 2 Sample of images of the data matrix M: (a) clean, (b) with mixed pixels.

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 Non-Negative Underapproximation

Equation (3) is convex in x and y separately, and the corresponding optimal solutions can actually be trivially computed when $x \geq 0$ and $y \geq 0$,

$$x^* = \operatorname{argmin}_{x \geq 0, xy^T \leq M} \|M - xy^T\|_F,$$

$$x_i^* = \min_{\{j \mid y_j \neq 0\}} \left\{ \frac{m_j^i}{y_j} \right\} \forall i, \tag{4}$$

and

$$y^* = \operatorname{argmin}_{y \geq 0, xy^T \leq M} \|M - xy^T\|_F,$$

$$y_j^* = \min_{\{i \mid x_i \neq 0\}} \left\{ \frac{m_j^i}{x_i} \right\} \forall j, \tag{5}$$

and this corresponds to the stationary conditions of Eq. (3). Note that these conditions are the same for other objective functions, such as the ℓ_1-norm of the error, which will be analyzed in Sec. 3.

Alternating optimization (i.e., updating x and y alternatively, which is also referred to as block-coordinate descent) does not generate satisfactory results: the algorithm will stop after one or two updates of x and y and is then unable to locate good solutions, potentially far away from the initialization. The reason is that feasibility is imposed at each step and solutions are rapidly blocked on the boundary of the feasible domain.

A Lagrangian relaxation scheme has been proposed[16] to solve Eq. (3). 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

Fig. 3 Ideal Case. From top to bottom: four original images (i.e., columns of U), basis elements obtained with NMF, and with NMU.

problem as

$$\sup_{\Lambda \in \mathbb{R}_+^{m \times n}} \min_{\substack{x \in \mathbb{R}_+^m \\ y \in \mathbb{R}_+^n}} \|M - xy^T\|_F^2 - 2\langle \Lambda, M - xy^T \rangle$$

$$= \|(M - \Lambda) - xy^T\|_F^2 - \|\Lambda\|_F^2. \tag{6}$$

A possible way to solve Eq. (6) 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 ℓ_2-NMU) while Λ is updated with a subgradient type update (step 10). Note that the problem of optimizing both x and y reduces to a rank-one non-negative factorization problem [same problem as Eq. (1)], where M might have negative entries; in this case $M - \Lambda$) which is NP-hard.[20]

We propose the following two improvements, which make the algorithm more robust:

1. (x, y) are initialized with the optimal rank-one solution of the problem without the underapproximation constraints (i.e., the optimal non-negative rank-one approximation of the residual, see step 2, corresponding to $\Lambda = 0$, which we compute with the singular value decomposition); Λ is initialized with the non-negative part of the residual matrix (step 4).

2. If Λ is too large, then 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).

Because the algorithm is not guaranteed to generate a feasible solution, only the non-negative part of the residual is considered (step 15). [Note that 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 Eqs. (4) and (5) as a postprocessing step.] Also note that the updates of x and y share some similarities with the power method (applied to $M - \Lambda$, with projection on the non-negative orthant), which computes the maximum singular value and its corresponding left and right singular vectors.[18] It seems that Algorithm ℓ_2-NMU behaves similarly as the power method in the sense that it converges in general relatively fast. Extensive experiments on a host of data and applications allow us to conclude 100 iterations at each step of the recursion is sufficient (i.e., maxiter = 100, which will be used for the numerical experiments; see Sec. 4).

Algorithm ℓ_2-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. $UV^T \lesssim M$.

1.   **for** $k = 1: r$ **do**

2. $[x, y]$ = optimal rank-one approximation (M);

3. $u_k \leftarrow x$; $v_k \leftarrow y$;

4. $\Lambda \leftarrow \max \left[0, -(M - xy^T) \right];$

5.   **for** $p = 1 :$ maxiter **do**

6. $x \leftarrow \max \left[0, \dfrac{(M - \Lambda)y}{\|y\|_2^2} \right];$

7. $y \leftarrow \max \left[0, \dfrac{(M - \Lambda)^T x}{\|x\|_2^2} \right];$

8.   **if** $x \neq 0$ and $y \neq 0$ **then**

9. $u_k \leftarrow x;\ \upsilon_k \leftarrow y;$

10. $\Lambda \leftarrow \max \left[0, \Lambda - (1/p)(M - xy^T) \right];$

11.   **else**

12. $\Lambda \leftarrow (\Lambda/2);\ x \leftarrow u_k;\ y \leftarrow \upsilon_k;$

13.   **end if**

14.   **end for**

15. $M = \max \left(0, M - u_k \upsilon_k^T \right);$

16.   **end for**

2.2 Hyperspectral Data Analysis in the Ideal Case

If we assume that each pixel contains only one material, then 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 $r \leq min(m, n)$, its columns are orthogonal:

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

and there is one and only one nonzero element in each row of U

$$u_k^i = 1 \quad \Leftrightarrow \quad pixel\ i\ contains\ material\ k.$$

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

Of course, recovering U and V in these settings is trivial and, in practice, because of blurring and other mixing 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. This will shed some light on the behavior of the above recursive algorithm based on underapproximations and justify its efficiency when dealing with nonideal 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 we have the following results:

Lemma 1. *Let* (x, y) *be a nontrivial stationary point of Eq. (3) (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. The proof follows directly from Eqs. (4) and (5).

Lemma 2. *Let* (x, y) *be a nontrivial stationary point of Eq. (3) for* $M = UV^T$ *satisfying Assumption 1, then the residual* $R = M - xy^T$ *can be written as* $R = UV'^T$ *for some* $V' \geq 0$.
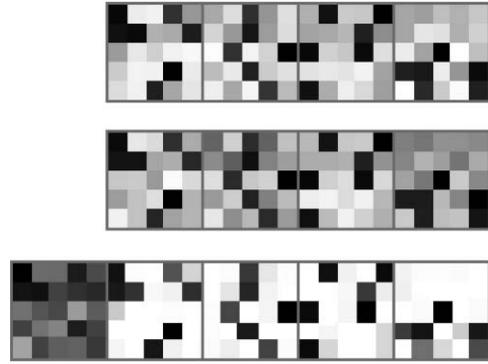

Fig. 4 Nonideal Case. From top to bottom: four original images, basis elements obtained with NMF, and with NMU.

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., $\forall i \in \{1, 2, \ldots, r\}, \forall k, l \in supp(u_i): x_k = x_l.$] In fact, one can check that for $y \neq 0$, the solution of Eq. (4) is unique. It follows that $x = Ud$, for some $d \in \mathbb{R}_+^m$, and then $R = UV^T - Udy^T = U(V - yd^T)^T$. The facts that R is non-negative and that U is binary and orthogonal implies that $V' = V - yd^T \geq 0$.

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 Eqs. (4) and (5) whereas structure of the residual matrix R is a consequence of Lemma 1 and 2.

Let us use the notations of Corollary 1. We observe that it is typically very unlikely for the sparsity pattern of a column of V' to be contained in the sparsity pattern of another column, i.e., that

$$I = \overline{supp}(\upsilon_i') \subset \overline{supp}(\upsilon_j'), \quad \text{for some } i \neq j, \quad (7)$$

for some nonempty set $I \subset \{1, 2, \ldots, n\}$. There are two basic reasons for this fact

1. We know there is at least one zero by row and by column in V' (Corollary 1). Clearly,

   $$\upsilon_i'(I) = \upsilon_j'(I) = 0 \Leftrightarrow \upsilon_i(I) = \alpha \, \upsilon_j(I),$$

   for some constant $\alpha > 0$. In fact, recall that $\upsilon_i' = \upsilon_i - d_i y$ so that $\upsilon_i'(I) = \upsilon_j'(I) = 0$ if and only if $\upsilon_i(I) - d_i y(I) = \upsilon_j(I) - d_j y(I) = 0$ [i.e., $\upsilon_i(I) = (d_i/d_j)\upsilon_j(I)$]. If $|I| \geq 2$ and if we assume that V is generated randomly, then the probability of having $\upsilon_i(I) = \alpha \upsilon_j(I)$ is zero (randomly generated vectors in two dimensions or more are multiple of each other with probability zero). If $|I| = 1$, then it means that at least one column of V' has only one zero element. We know that there are at least n zeros in V' (one by row) and at least one zero in each of the r columns of V'. There are still at least $(n - r)$ zeros to be placed in the r columns of V'. Assuming there are only $(n - r)$ zeros (typically, there are much more zeros in the residual) and that they are distributed randomly with uniform distribution among the columns of V', then we can compute the probability of having one column of V' with only one
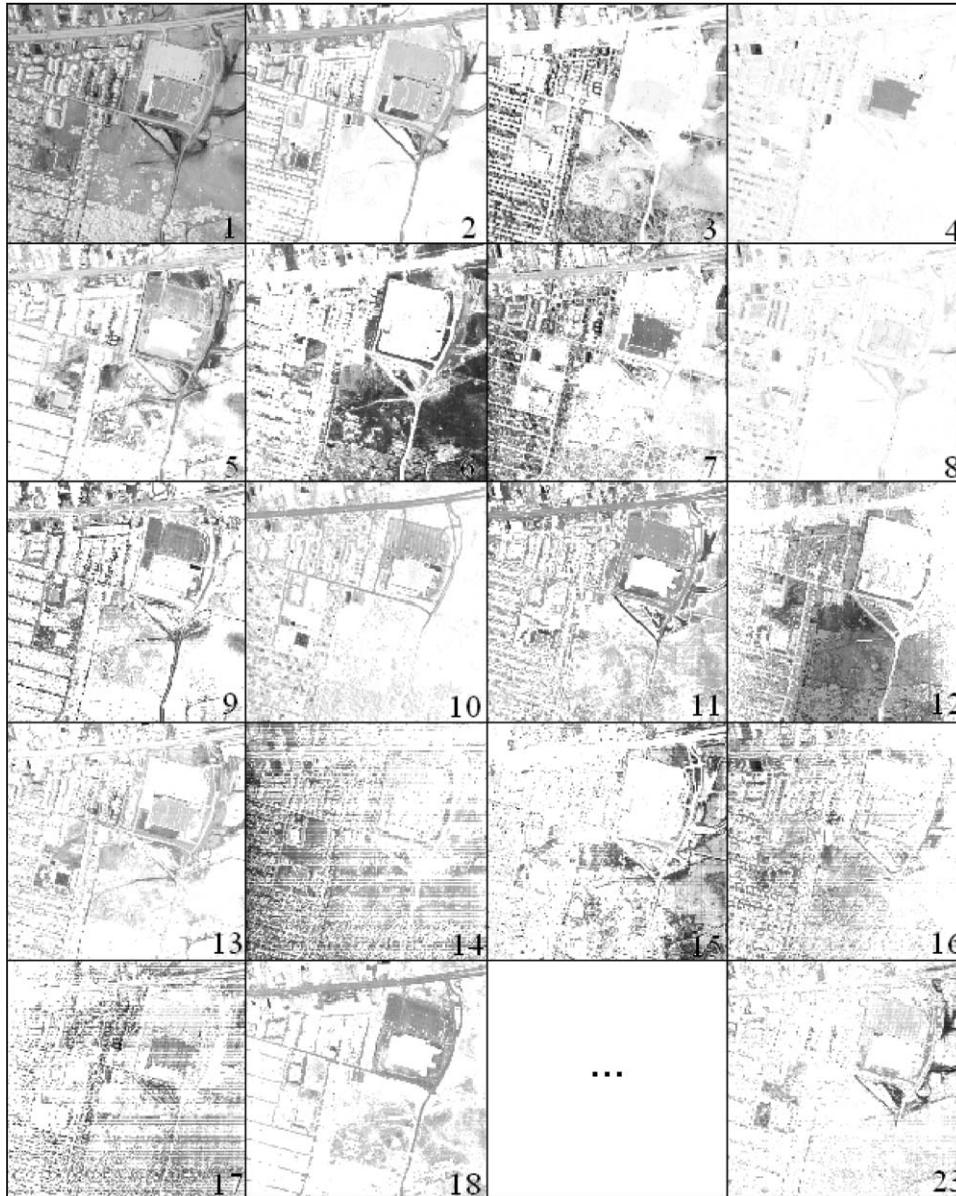
Fig. 5 Basis elements (columns of matrix U) for the Urban dataset extracted with ℓ_2-NMU. Dark tones indicate a strong presence of a pixel in a cluster (0 is white, 1 black), and numbers indicate the position of the factor in the NMU decomposition.

zero element, and with its sparsity pattern contained in the support of another column [i.e., case $|I| = 1$ in Eq. (7)]. [Note: We added up the probabilities to have i columns with only one zero element multiplied by the probability for at least one of these zeros to be located at the same position of another one (either in one of these i columns with one zero, or in the remaining $r - i$).] Figure 1 displays this probability for $n = 210$ (which is the number of spectral bands for some HYDICE images we consider, cf. Sec. 4) with respect to r (number of materials). For example, for $r \leq 25$ (i.e., <25 materials present in the image), the probability for at least one column of V' to have only one zero entry, and for another column of V' to have a zero at the same position is $<10^{-2}$.

2. In practice, it is observed that the zeros are not distributed randomly among the columns of V'. Typically, the columns of V' have the same number of zeros ($\sim n/r$), located at different positions.

Using another objective function, the sum of the logarithms of the ratios between the entries of xy^T and $M > 0$, it can be proved[19] that the problem is equivalent to an assignment problem. In fact, using a logarithmic change of variables, we have the problem

$$\min_{x>0, y>0} \sum_{ij} \log\left(\frac{m_j^i}{x_i y_j}\right) \text{ s.t. } x_i y_j \leq m_j^i$$

which is equivalent to

$$\max_{x>0, y>0} n \sum_i \log(x_i) + m \sum_j \log(y_j)$$

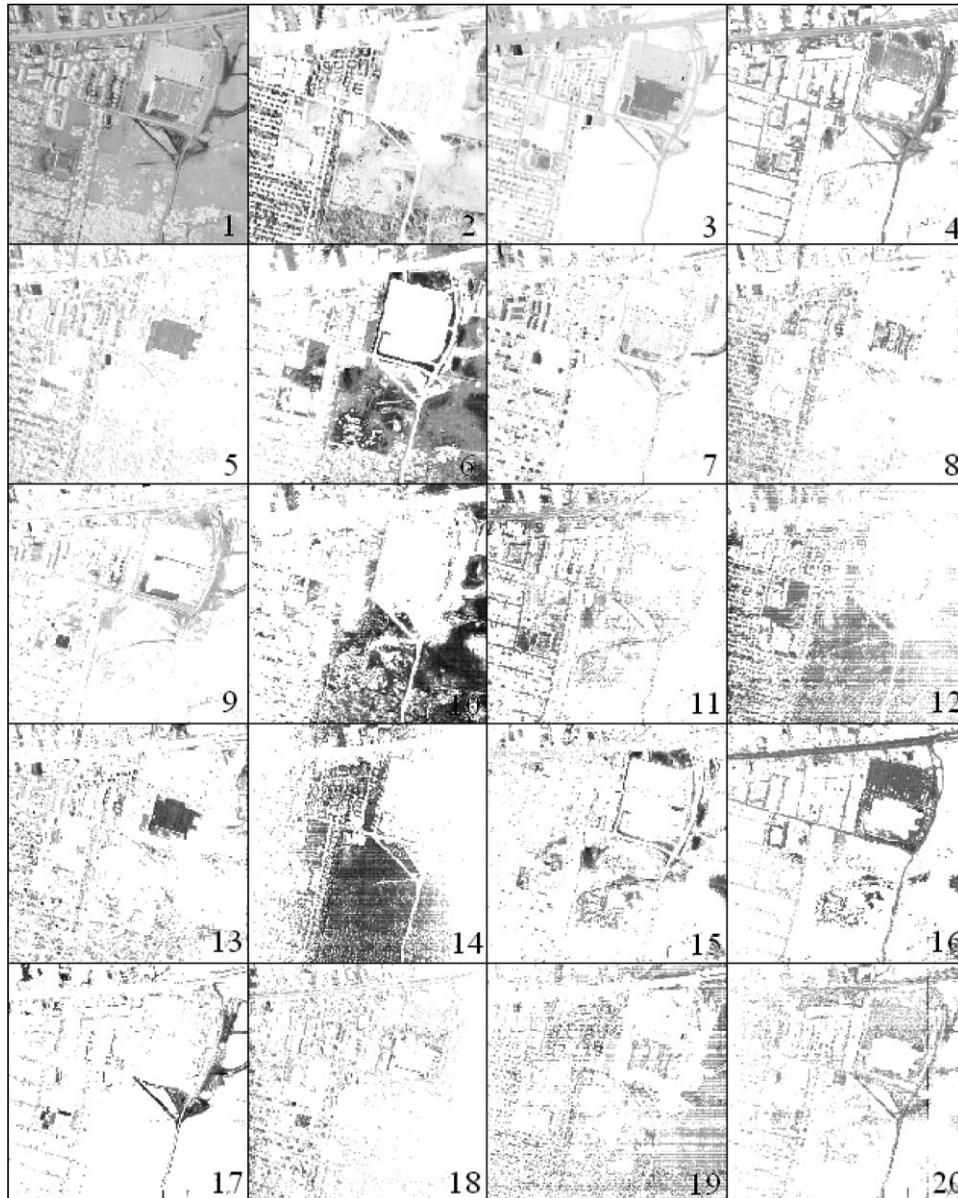$$\text{s.t. } \log(x_i) + \log(y_j) \leq \log(m_j^i). \tag{8}$$

Fig. 6 Basis elements of ℓ_1-NMU for the Urban dataset extracted, identified as in Fig. 5.

If M is a square matrix ($m = n$) and (x, y) an optimal solution of Eq. (8), then the zeros of $R = M - xy^T = UV'$ will be located on its diagonal (up to a permutation) and no columns will share common zeros. [Note: This is related to the way assignment problems are solved.[21] One can show that a solution (x, y) of Eq. (8) is optimal if and only if the matrix D with $d_{ij} = \log(m_j^i) - \log(x_i) - \log(y_j)$ has, up to a permutation, zeros on its diagonal (this is how the Hungarian method used to solve assignment problems has been designed).] When using the Frobenius norm as an objective function, it seems that the zeros follow the same sparsity pattern, even though we do not have a proof for this fact. However, note that Eq. (8) shares the same stationarity conditions as Eq. (3) [see Eqs. (4) and (5)], so that optimal solutions of Eq. (8) are stationary points of Eq. (3).

Conclusion. After the first NMU recursion, the residual R can be written in the same form as $M = UV^T$ (cf. Assumption 1) with $R = U V'^T$, and it is highly probable that the columns of V' will have disjoint sparsity patterns. This will be experimentally confirmed in Sec. 4.

2.2.2 Next rank-one factors

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

Theorem 1. *Let (x, y) be a nontrivial stationary point of (NMU) for $M = UV^T$ satisfying Assumption 1 and the columns of V have disjoint sparsity patterns, i.e., $\overline{\mathrm{supp}}(v_i) \not\subseteq \overline{\mathrm{supp}}(v_j) \forall i \neq j$. Then $R = M - xy^T = U V'$, with $x = Ud$ for some $d \in_+^r$ so that $V' = V - yd^T \geq 0$. Moreover,*

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

and

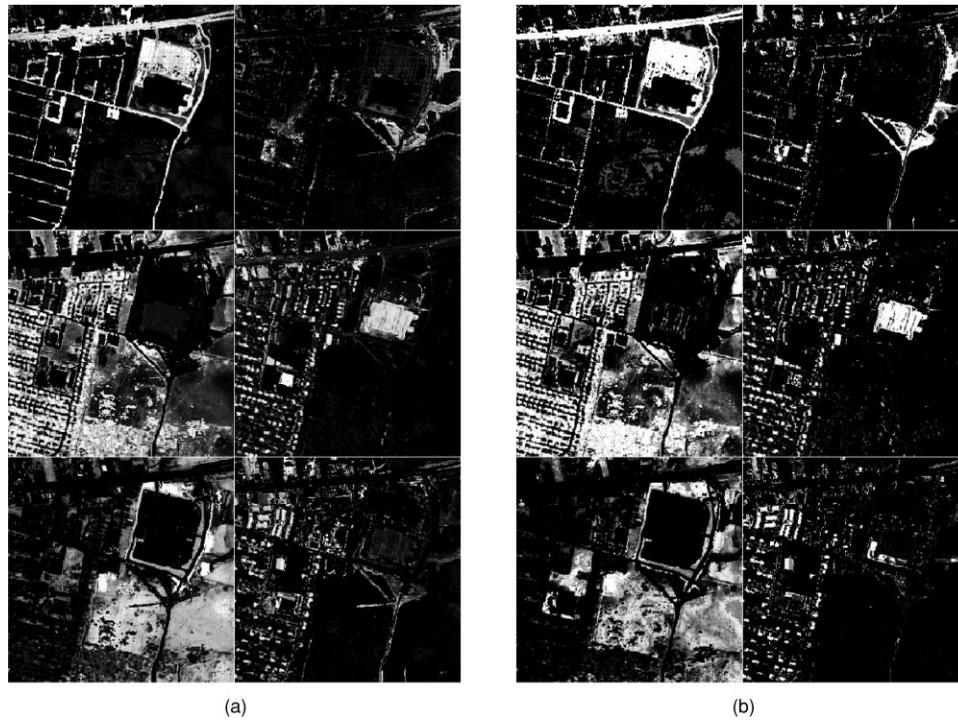$$\Omega = \{i\} \Leftrightarrow v_i' = 0 \Leftrightarrow d_i y = v_i, \quad 1 \leq i \leq r. \tag{9}$$

Fig. 7 Postprocessed basis elements of NMU for Urban dataset with (a) ℓ_2-norm and (b) ℓ_1-norm. Light tones represent high degree of membership.

Proof. The first part is a consequence of Corollary 1. It remains to show that Eq. (9) holds. The second equivalence is trivial since $\upsilon_i' = \upsilon_i - d_i y$ is equal to zero for some i if and only if $d_i y = \upsilon_i$. For the first equivalence, observe that $d_i y = \upsilon_i$ implies that $\Omega = \{i\}$ because of the underapproximation constraints and because the columns of V have disjoint sparsity patterns [cf. Eq. (4)]. In fact, because y has the same support as υ_i, we have $\forall j \neq i, \exists k$ s.t. $\upsilon_j(k) = 0$ and $y(k) >$

0 implying $d_j = 0$. Finally, it is clear that for $\Omega = \{i\}$, the solution obtained with Eq. (5) is $y = (1/d_i)\upsilon_i$.

Theorem 1 implies that, at each step of the NMU recursion, a set of materials are extracted together. Moreover, a material is extracted alone if and only if the corresponding column of V' is set to zero. Because the recursive approach outlined above will eventually end up with a zero matrix (say, after r_u steps), we will have
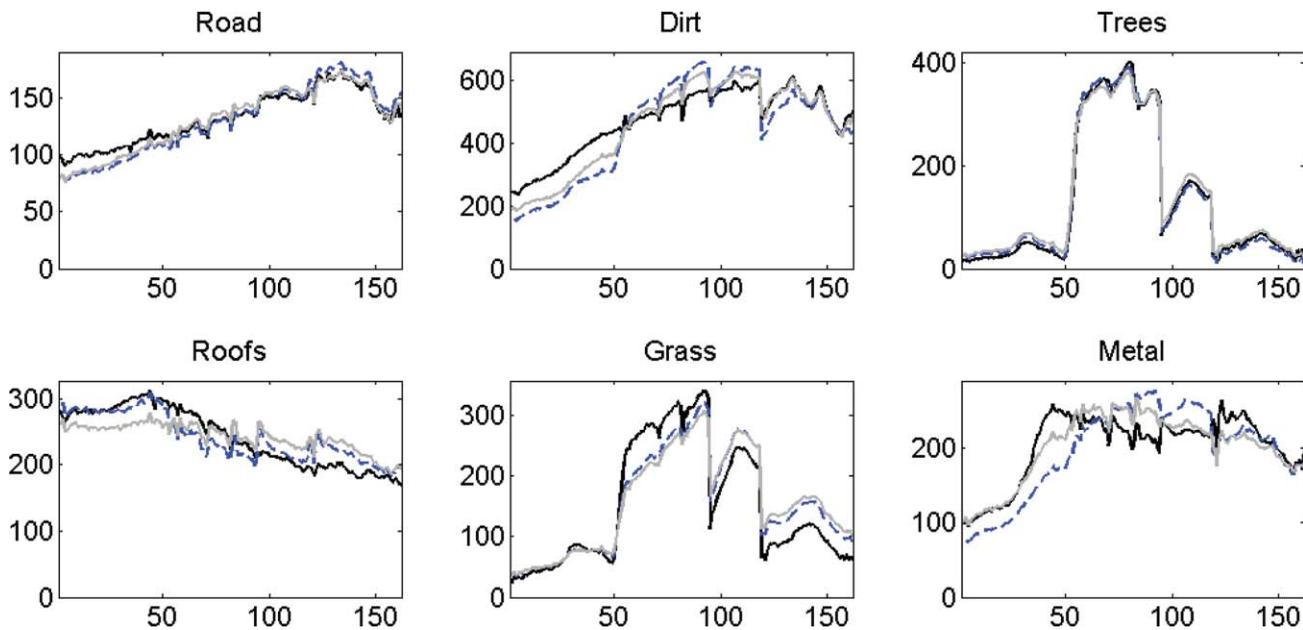

Fig. 8 End-member extraction for the Urban dataset: ℓ_2-NMU (gray solid) and ℓ_1-NMU (dashed) versus six end members from the image using N-FINDR5,[33] plus manual adjustment (dark solid) from Ref. 25. The *x*-axis gives the wavelength bands while *y*-axis gives the reflectance values (intensities of reflected light).
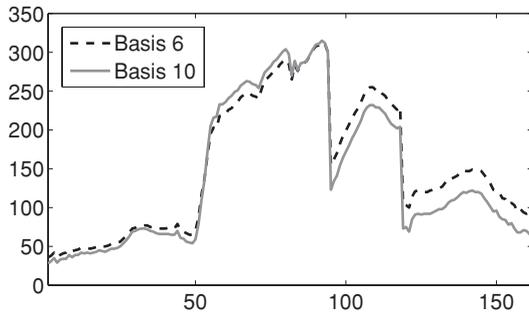
Fig. 9 Spectral signatures of grass: weighted average of the spectral signatures of the pixels present in basis elements 6 and 10 of the ℓ_1-norm solution.

$$M = \sum_{i=1}^{r_u} x_i y_i^T, \tag{10}$$

and under the disjoint sparsity patterns assumption (at each step of the recursion),

$$\forall 1 \leq i \leq r, \exists 1 \leq j \leq r_u \quad \text{s.t.} \quad \text{supp}(x_j) = \text{supp}(u_i). \tag{11}$$

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 Sec. 4.

Remark 1. *The disjoint sparsity patterns assumption is a sufficient but not a necessary condition for exact recovery. In fact, if two columns 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 it is likely that they will have disjoint sparsity patterns.*

2.3 Illustration of Basis Recovery with NMU versus NMF

Let us construct the following synthetic data: four binary orthogonal images of 5×5 pixels (which are the columns of U, $U \in \{0, 1\}^{25 \times 4}$, see top image of Fig. 3) are randomly mixed ($V \in \mathbb{R}^{25 \times 4}$ is randomly generated with uniform distribution between 0 and 1) to generate a 25×25 matrix $M = UV^T$ satisfying Assumption 1. [Note: We used the function rand(4,25) of MATLAB®.] Figure 2 displays a sample of the 25 images contained in the columns of M, which then result from the non-negative linear combination of the columns of U.

Figure 3 displays the original images and the basis elements obtained with NMF and NMU. As was mentioned in Sec. 2.2.1, the first rank-one factor of NMU will reduce the error the most and we include it and list a total of five for NMU (because each end-member has been extracted individually after five steps, the residual error is equal to zero, i.e., the approximation is exact; see Theorem 1).

We observe that NMF is not able to perfectly extract the four original basis elements (even though the objective function is equal to zero; the reason is the nonuniqueness of the solution: NMF retrieves a mixture of the basis elements, while NMU is able to do the extraction. [Note that for $n = 25$ and $k = 4$, the probability for two columns of V' to have nondisjoint sparsity patterns is $<10^{-8}$, if we assume that the zeros are uniformly distributed and that V is randomly generated (which is the case here); see Sec. 2.2.1.]

2.4 Hyperspectral Data Analysis in the Nonideal Case

As we have already mentioned, practical problems do not have the nice structure mentioned in Assumption 1 and the spectral signature of most pixels results from a combination of several materials. What can we expect of NMU
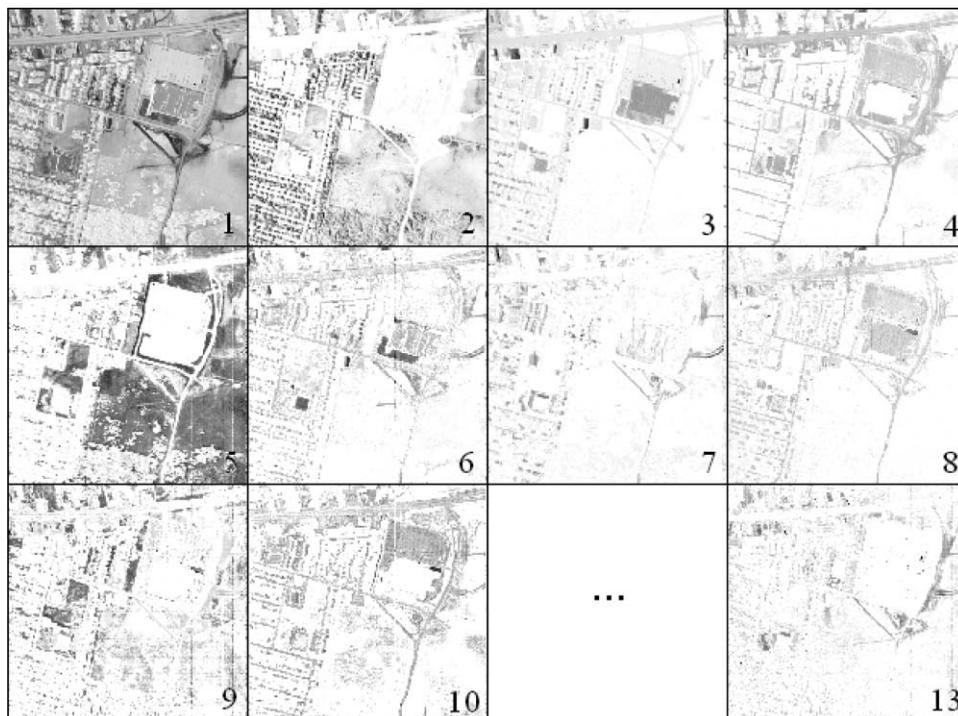

Fig. 10 Basis elements of ℓ_2-NMU for the Urban dataset using only nine bands (bands 1, 20, 40, . . . , 160).
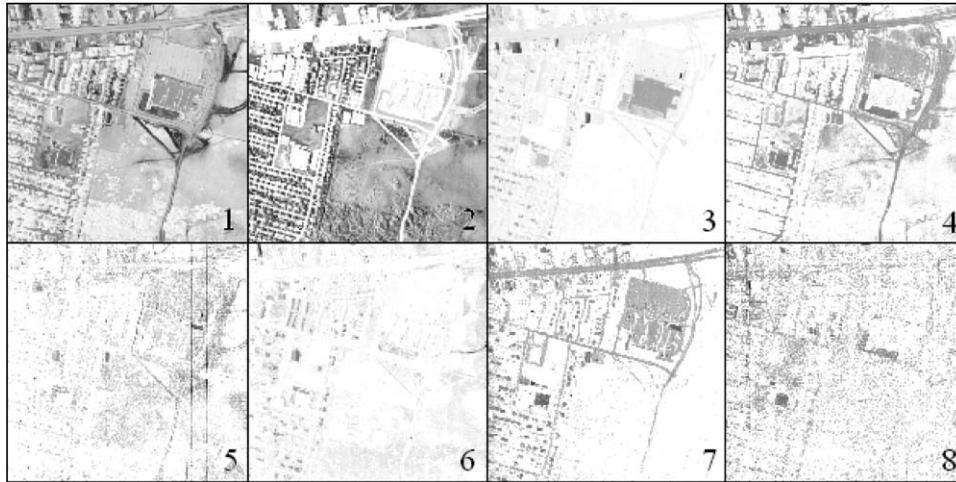
Fig. 11 Basis elements of ℓ_2-NMU for the Urban dataset using only five bands (1, 41, 81, 121, 161).

in that case? When the data matrix is positive, the first rank-one factor will still be a mixture of all materials (cf. Lemma 1). It seems more difficult to provide theoretical guarantees for the next factors in more general settings; 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 this component from their spectral signature) and, because 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 incited to extract constitutive materials in nonideal 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. This means that each pixel is now a mixture of several materials, but one material is still predominant. Figure 4 displays the visual results: NMF performs even worse, whereas NMU is still able to extract the original parts fairly well. It actually provides a soft clustering for each pixel, which will also be shown in Sec. 4. [Note: Soft clustering means that a single element of the dataset can be assigned to different clusters; a (nonnegative) weight being attached to each cluster (e.g., corresponding to the probability of the pixel to belong to the cluster). In a hyperspectral image, it makes sense since the pixels can be composed of different materials (the clusters) with different abundances (the weights, summing to one).]

3 ℓ_0-Pseudo-Norm Minimization and ℓ_1-Norm Relaxation

Ideally, each basis element extracted with the recursive approach outlined previously should correspond to a different material present in the hyperspectral image: we would like each extracted rank-one factor to correspond 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 ℓ_2-NMU) 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, then this will 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 very difficult to work with (nondifferentiable, 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, the ℓ_0-'norm' would not be appropriate because 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 and is then disposed to let some entries of the error be large in order to approximate better other entries. We will experimentally observe, in Sec. 4, that using ℓ_1-norm allows
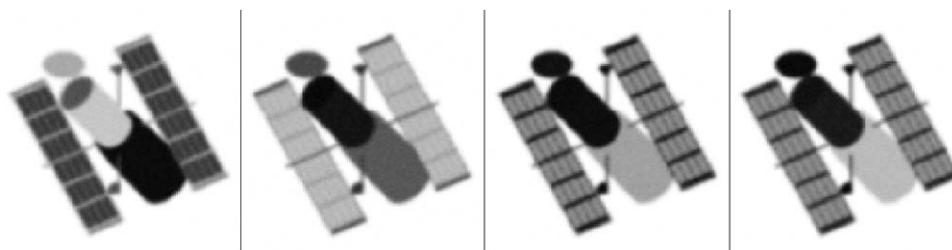

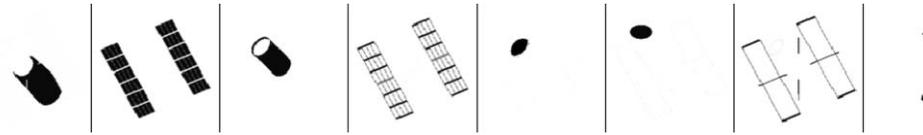**Fig. 12** Sample of images in the Hubble tensor with blur and noise.

Fig. 13 The eight materials for the Hubble telescope data provided to us by NASA. From left to right: Aluminum, Solar Cell, Green Glue, Copper Stripping, Honeycomb Side, Honeycomb Top, Black Rubber Edge, and Bolts.

us 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 Sec. 2.1, we propose to solve

$$\max_{\Lambda \in \mathbb{R}_+^{m \times n}} \quad \min_{x \in \mathbb{R}_+^m, \, y \in \mathbb{R}_+^n} \quad \|(M - \Lambda) - xy^T\|_1. \tag{12}$$

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 for Algorithm ℓ_2-NMU.

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

$$\min_{x_i \geq 0} \|a^i - x_i y\|_1 = \sum_j |a_j^i - x_i y_j|$$

$$= \sum_{j \in \text{supp}(y)} y_j \left| \frac{a_j^i}{y_j} - x_i \right| + \sum_{j \notin \text{supp}(y)} |a_j^i|, \tag{13}$$

which can be solved by computing the weighted median of z with $z_j = (a_j^i / y_j) \, \forall j$ 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 ℓ_2-NMU (steps 6 and 7) by

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

$$J = \text{supp}(y), \tag{14}$$

and

$$y_j = \max \left\{ 0, \text{weighted-median} \left[\frac{(M - \Lambda)_j(I)}{x(I)}, x(I) \right] \right\} \forall j,$$

$$I = \text{supp}(x) \tag{15}$$

The weighted median of an n dimensional vector can be computed in $O(n)$ operations (cf. Ref. 22, and references therein) so that the algorithm can be implemented in $O(mn)$ operations per iteration when the data matrix M has dimension $m \times n$. We will refer to this algorithm as ℓ_1-NMU. The ℓ_2-NMU and ℓ_1-NMU algorithms then have the same computational complexity even though in practice ℓ_1-NMU will be slower, but only up to a constant factor. [Note: Implementation of both algorithms is available at http://www.core.ucl.ac.be/~ngillis/.]

4 Applications to Spectral Data

In this section, Algorithm ℓ_2-NMU and its modification for ℓ_1-norm minimization proposed in Sec. 3 (ℓ_1-NMU) are used as dimensionality reduction techniques for hyperspectral and multispectral data in order to achieve *classification* (selecting from the basis elements the different clusters), and *spectral unmixing* (using non-negative least squares). In the first part, we carefully analyze the Urban HYDICE image (Sec. 4.2) and a Hubble space telescope simulated image (Sec. 4.3) developed in Ref. 6. In the second part, we provide some visual results for aerial images of a desert region and of the San Diego airport (Secs. 4.4.1 and 4.4.2), and for an eye image with only four spectral bands, useful in biometric identification systems (Sec. 4.4.3).
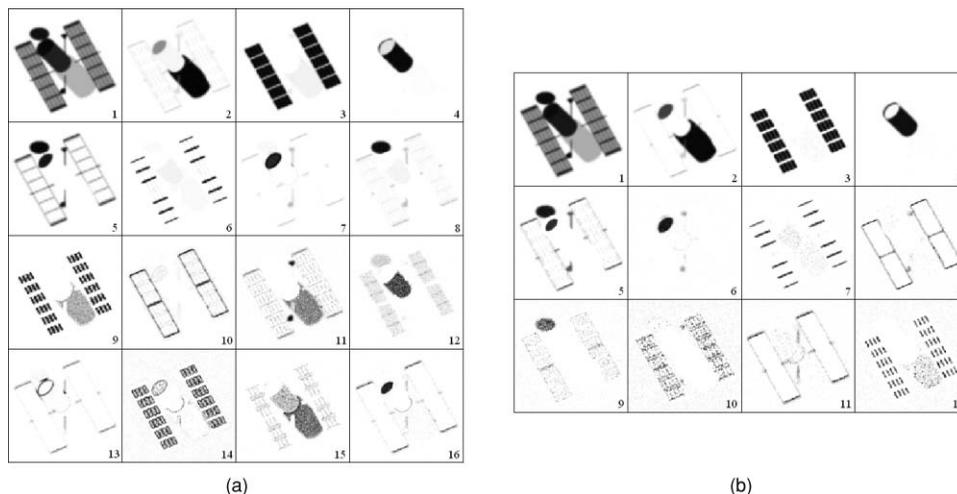

Fig. 14 Basis elements of NMU for the Hubble telescope dataset with added blur and noise: (a) ℓ_2-norm and (b) ℓ_1-norm.

Table 1 Basis element obtained: cluster selection for the Urban dataset.

Clusters	Road	Dirt	Trees	Roofs	Grass	Metal
ℓ_2-basis No.	18	23	3	4	6	8
ℓ_1-basis No.	16	17	2	5	6	7

4.1 Classification and Spectral Unmixing

NMU can be used as a standard dimensionality reduction technique, and any type of postprocessing 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 (i.e., manually selected the basis elements representing a single material). 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 postprocessing step is to normalize U. In fact, as for 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

$$u_j = \frac{u_j}{\max_i(u_j^i)} \quad \forall j. \tag{16}$$

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, because rows of U correspond to abundances, $\Sigma_j u_j^i = 1 \; \forall i$ and we can scale the rows of U as follows:

$$u^i \leftarrow \frac{u^i}{\|u^i\|_1 + \varepsilon}, \quad \varepsilon \lll 1 \tag{17}$$

so that they sum to 1 (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 composed of several materials with the corresponding abundances given by u^i.

Once the materials have been identified (by selecting a subset of the columns of U) and the pixels have been classified (by scaling properly the selected columns of U), one can recover the spectral signatures of each individual material, called the end members. A standard approach is to solve a non-negative least-squares problem of the form

$$\min_{V \geq 0} \|M - U V^T\|_F^2, \tag{18}$$

where the columns of U are the postprocessed basis vectors, with dedicated algorithms (cf. Refs. 23 and 24). [Note: The columns of V in the NMU solution do not correspond directly to the spectral signatures of the end members because several materials are sometimes extracted together, in particular at the first steps of the algorithm. In order to get the spectral signature of an end member from the NMU solution, one has to sum up all the columns of V corresponding to a column of U containing the desired end-member. For example, for $M > 0$, the first column of V always has to be used to obtain a spectral signature because all materials are present in the first basis element with $u_1 > 0$; see Corollary 1. It is therefore more convenient and more accurate to compute the spectral signatures by solving Eq. (18).]

4.2 Urban HYDICE Image

We consider first the Urban hyperspectral image (available at http://www.agc.army.mil/hypercube/) taken with hyperspectral digital imagery collection experiment (HYDICE) airborne sensors. We analyze the data where the noisy bands have been removed (162 bands left, originally 210), and the data cube has dimension $307 \times 307 \times 162$. Figures 5 and 6 give the basis elements of the ℓ_2- and ℓ_1-NMU decompositions.

The Urban data are mainly composed of six types of materials: road, dirt, trees, roofs, grass, and metal, as reported in Ref. 25. Table 1 gives the index of the NMU basis elements corresponding to single materials. Figure 7 shows the classification obtained from the basis elements obtained with NMU (cf. Figs. 5 and 6, and Table 1). Figure 8 displays the results of the spectral unimxing procedure for both NMU algorithms (ℓ_2 and ℓ_1) which are compared to 6 end-members, as listed as the "true" end members for this data in Ref. 25.

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 predicted, the ℓ_2-NMU needs more recursion than ℓ_1-NMU to extract all materials individually (23 versus 17). For example, it is interesting to observe that ℓ_1-NMU actually extracts the grass as two separate basis elements (6 and 10, cf. Fig. 6). The reason is that the spectral signatures of the pixels in these two basis elements differ (especially after the 100th band in the hyperspectral image): there are two types of grass with similar spectral signatures, and they are different enough to be assigned to different clusters (see Fig. 9). Because ℓ_1-NMU is able to extract parts separately in a more efficient way (cf. Sec. 3), it is able to detect this anomaly, whereas ℓ_2-NMU is not. This fact also provides an explanation of the differences in the spectral signatures of the grass in Fig. 8.

4.2.1 Number of spectral bands

In Sec. 2.2, a sufficient condition was presented to recover each material individually: the number of spectral bands (n)

Table 2 Basis elements obtained: Cluster selection for the Hubble telescope database with noise and blur (see Fig. 14).

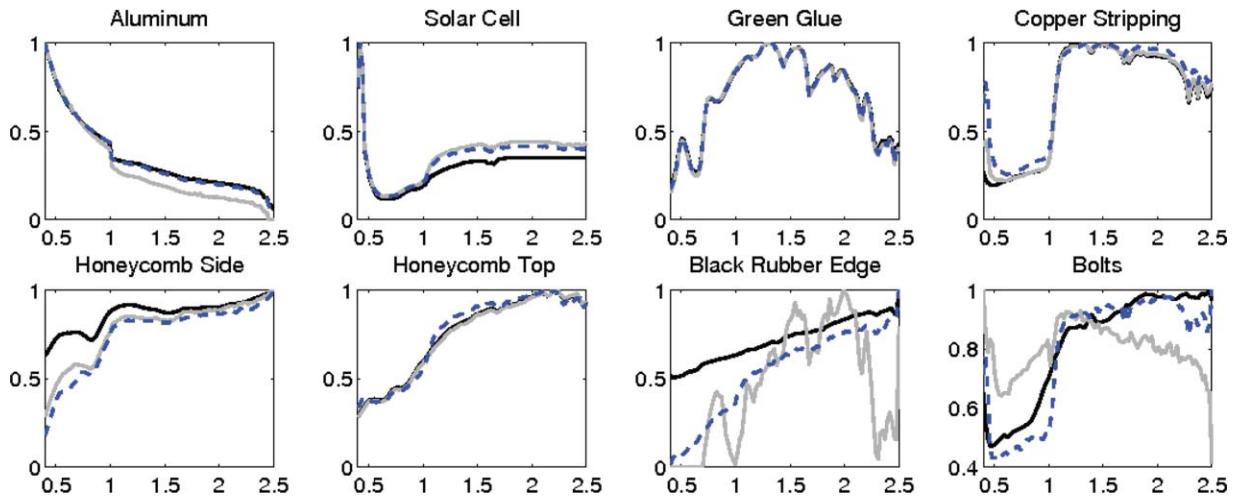| Clusters	Alum.	S. Cell	Glue	Copper	H. Side	H. Top	Edge	Bolts
ℓ_2-basis No.	2	3	4	6	7	8	13	11
ℓ_1-basis No.	2	3	4	7	6	9	11	8

Fig. 15 End-members for the Hubble satellite data with noise and blur. NMU with ℓ_2 (gray solid) and NMU with ℓ_1 (dashed) versus eight true end-members (black solid).

should be sufficiently larger than the number of materials (r). This is clearly satisfied by the Urban data used here, because for $n = 162$ and $r = 6$, the probability to have disjoint sparsity patterns is larger than $1 - 10^{-12}$. However, it has also been explained why this condition is not necessary; see Sec. 2.2. What happens then if we reduce the number of bands? For example, suppose we keep only nine bands from the 162 original clean bands of the urban data. [Note: We selected the bands so that they are well distributed in the spectral domain, starting from the first band and using a constant step. However, one might use more sophisticated techniques (e.g., subset selection algorithms such as in Ref. 26) or using dimensionality reduction techniques preserving non-negativity (such as NMF) as a preprocessing step. We know of no good dynamical way using NMU to determine the minimal number of bands sufficient to capture selected end members.] Figure 10 displays the basis elements for ℓ_2-NMU. Surprisingly, the algorithm is still able to separate the materials: trees are recovered in basis element 2, roofs 3, road 4 (mixed with dirt), grass 5, metal 7, and dirt 13.

When less than six bands are kept, the algorithm cannot detect all the materials individually. Figure 11 displays basis elements using five bands. The grass and trees are extracted together (basis element 2) because their spectral signatures are similar; the roads and the dirt also are extracted together (basis element 4), as are the road and roofs (basis element 7), whereas the roofs and the metal (basis elements 3 and 6 respectively) are extracted individually.

Finally, it seems that, as long as the spectral signatures of the different materials can be distinguished, the number of spectral bands does not need to be significantly larger than the number of materials in order for NMU to be able to perform classification (this will also be illustrated in Sec. 4.4.3). However, when this is not the case (e.g., when more noise and blur are present, or when spectral signatures look alike), more spectral bands are needed to distinguish the different materials, which seems to be a natural requirement (this will be illustrated in Sec. 4.3).
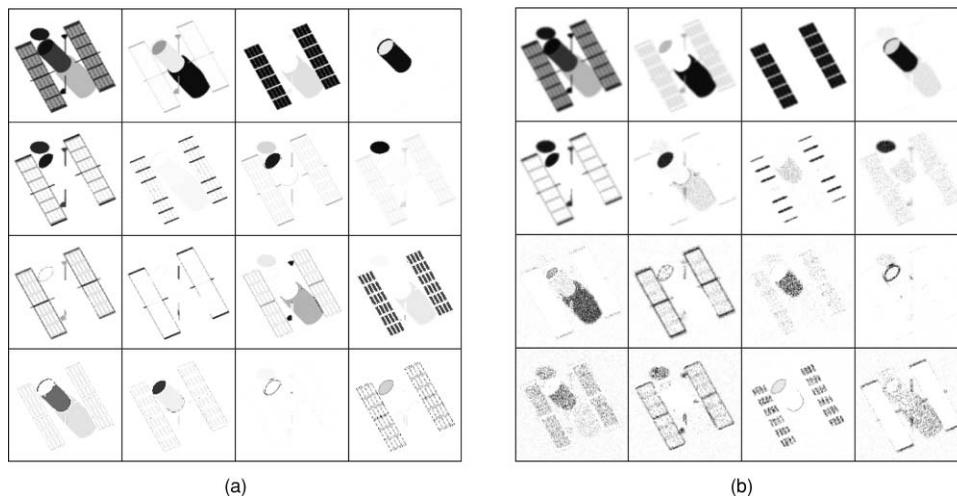

Fig. 16 Basis elements obtained with ℓ_2-NMU on the Hubble telescope hyperspectral image using only 12 spectral bands ($1 + 9i$, $0 \leq i \leq 11$): (a) clean image and (b) noisy and blurry image (same settings as before).
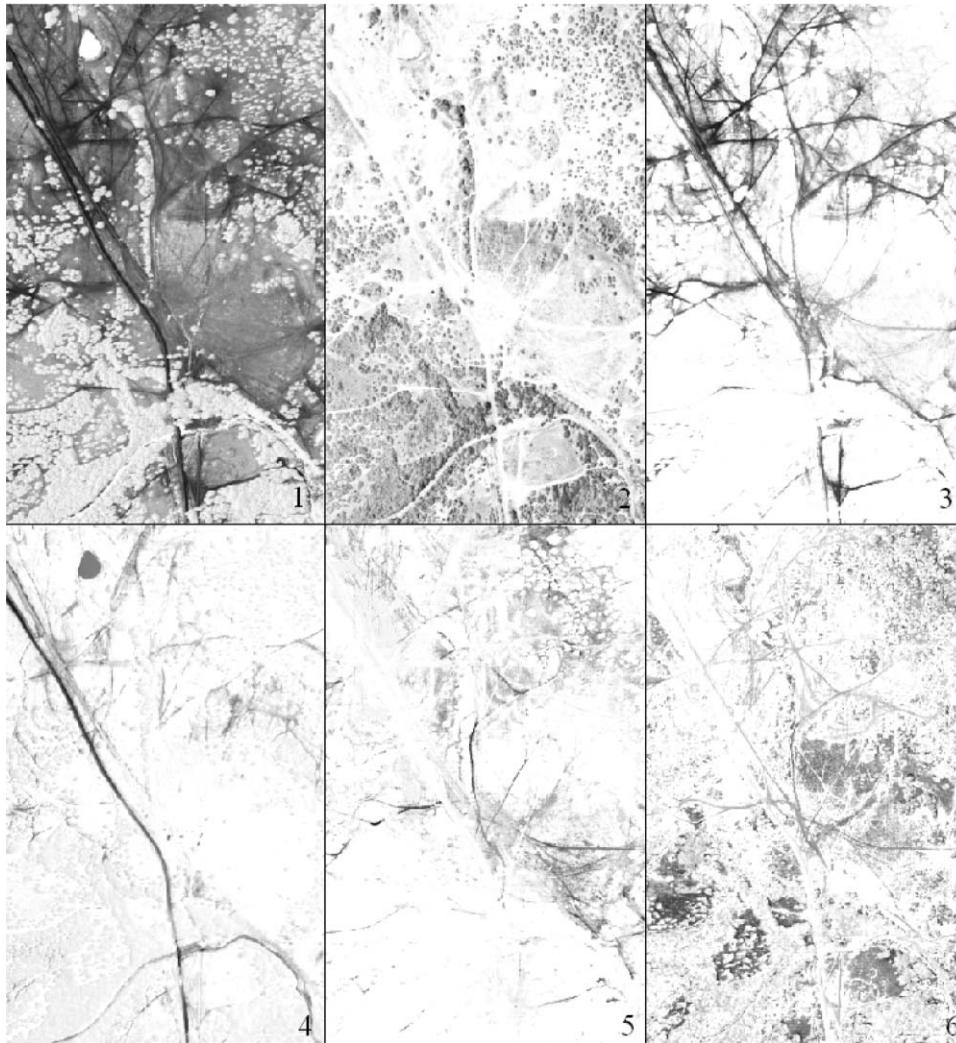
Fig. 17 Aerial image of a desert region basis elements obtained with ℓ_2-NMU.

4.3 Simulated Hubble Space Telescope Data

Figure 12 displays some sample images of the simulated Hubble database, which consists of 100 spectral images of the Hubble telescope, 128×128 pixels each, with added blur and noise.[5] [Note: Point spread function on 5×5 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.]. It is composed of eight materials; see Fig. 13. (Note: These are true Hubble satellite material spectral signatures provided to us by the NASA Johnson Space Center.)

Figure 14 shows the basis elements obtained with NMU, and Table 2 gives the classification of the basis elements. Figure 15 shows the end-member extraction: original versus 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 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, ℓ_2-NMU extracts more mixed materials and therefore needs more recursions to get all the parts separated than ℓ_1-NMU, which does a better job (especially for the black rubber edge).

4.3.1 Number of spectral bands

Let's reduce the number of spectral bands to 12, and compare the basis elements obtained by ℓ_2-NMU on the clean versus the noisy and blurry images. Figure 16 displays the basis elements. Clearly, in the noisy and blurry case, the algorithm is no longer able to extract all the materials. The reason is that there are not enough spectral bands left. Because of blur (spectral signatures of materials are mixed together) and noise (spectral signatures are perturbed), 12 bands is not enough to distinguish all the materials, as already explained in Sec. 4.2.1. Quite naturally, the larger the number of bands is, the more robust the algorithm will be with respect to noise and blur.

4.4 Visual Experiments

In this section, we first provide some visual results for datasets analyzed in a recent comparative study of dimensionality reduction techniques for hyperspectral images,[27] and for which 'ground truth' data is not available. [We will only

Fig. 18 San Diego Airport basis elements obtained with ℓ_2-NMU.

display basis elements obtained with Algorithm ℓ_2-NMU because results obtained with ℓ_1-NMU are comparable.] However, it allows one to experimentally reinforce our claims about the properties of NMU; namely that it is indeed able to detect materials or, at least, to separate some of them.

We then display results for a multispectral image of an eye with four spectral bands and propose a way to postprocess the basis elements obtained with NMU in order to achieve clustering. This biometric identification example is included in order to illustrate the diversity of applications of NMU in spectral imaging.

4.4.1 *Aerial image of a desert region*

This is a HYDICE terrain data set with 166 clean bands (originally 210), each containing 500×307 pixels. Figure 17 displays the first six basis elements. NMU is easily able to extract trees (basis element 2), roads (basis element 3), and grass (basis element 6).

4.4.2 *San Diego airport*

The San Diego airport hyperspectral image contains 158 clean bands, and 400×400 pixels for each spectral image. Figure 18 displays the first eight basis elements obtained by the NMU decomposition. In this case, it is less clear what the different materials are. It should probably be necessary to apply more sophisticated postprocessing techniques to classify the pixels. However, we observe the following:

1. Basis elements 4, 7, and 8 contain the roofs.
2. Basis element 6 mostly contains roads (including the parking lots).

3. Basis element 2 contains the grass and some roofs.
4. Basis element 3 is mostly composed of another type of road surface (including boarding and landing zones).

4.4.3 *Eye image*

We are given (only) four spectral images with 1040×1392 pixels (spectral bands correspond to IR, red, green, and blue channels). [Note: The data come from the West Virginia University multispectral image iris database. The circle around the pupil and the matrix inside the pupil were embedded in the image.] The data are from the West Virginia University multispectral image iris database[28] and is part of a biometric identification project involving Carnegie Mellon, Wake Forest, and West Virginia, using the ocular region of the face. Figure 19 displays the basis elements obtained with ℓ_2-NMU. The first basis element represents the pupil, part of the iris, and the eyelashes; the second represents some kind of substructure in the iris and the skin; and the third represents the pupil.

Segmentation and clustering multispectral eye images are useful in the analysis of iris-recognition algorithms in biometrics (see, e.g., Refs. 29 and 30). A possible way to postprocess the NMU basis elements in order to achieve clustering is to compare their support. Recall that each basis element should represent a set of materials. Therefore, if we want to identify these materials, then the following simple procedure can be used:

1. Compare the supports of each pair of basis elements; that is, compute $\text{supp}(u_i) \cap \text{supp}(u_j)$, $\forall i \neq j$.
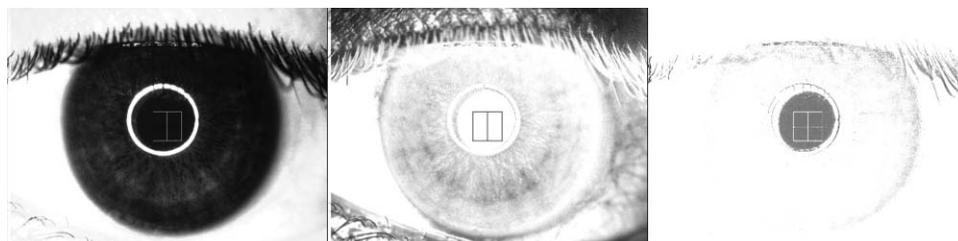

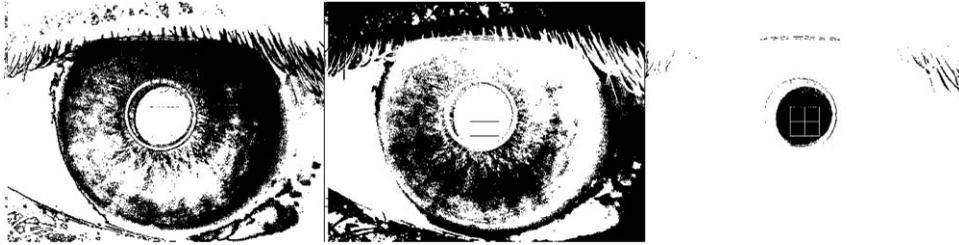**Fig. 19** Eye basis elements obtained with ℓ_2-NMU.

Fig. 20 Clustering of the eye, based on the NMU decomposition.

2. Define nonempty intersections as new basis elements. If no new basis elements are identified, then go to step 3; otherwise go back to step 1.

3. $\forall i \neq j$ such that $\text{supp}(u_i) \subset \text{supp}(u_j)$, set $\text{supp}(u_j) \leftarrow \text{supp}(u_j) \backslash \text{supp}(u_i)$. If some basis elements have been modified, then go back to step 1; otherwise go to step 4.

4. Materials are identified as the remaining basis elements, which define disjoint clusters.

Note that, in practice, some kind of thresholding should be used in order to define the supports, and the intersections containing a small number of pixels should be considered as empty.

In the example of Fig. 19, we have the following (after thresholding):

$$\text{supp}(u_1) \cap \text{supp}(u_2) \approx \emptyset, \quad \text{supp}(u_3) \subset \text{supp}(u_1) \text{ and}$$

$$\text{supp}(u_2) \cap \text{supp}(u_3) = \emptyset$$

so that the above procedure provides us with the clustering displayed in Fig. 20.

5 Summary and Further Work

We have studied an approximate non-negative matrix factorization (NMF) problem with underapproximation constraints, called non-negative matrix underapproximation (NMU), which can be used to design a recursive procedure to approximate non-negative data. We then gave theoretical and experimental evidence 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.

In further work, it would be interesting to compare NMU to other dimensionality reduction techniques such as PCA, NMF, ICA, etc. (see Ref. 27). Another direction of research would be to design automatic classification algorithms, based on the properties of NMU, to classify the pixels; as we proposed in Sec. 4.4.3. It would be particularly interesting to study these properties in more depth and see if it is possible to obtain stronger theoretical guarantees for the factors generated by NMU. In other work, comparisons of our NMU method will be made with the recent development of variational iterative methods for deblurring, denoising, and segmentation by Li et al.[31,32]

Finally, NMU can be generalized to higher order tensors, which could be called non-negative tensor underapproximation. For example, for a third-order tensor $\mathcal{T}$ of dimension $m \times n \times p$, we would use a third variable $z \in \mathbb{R}_+^p$ in order to approximate $\mathcal{T} \approx x \circ y \circ z$ [note: $(x \circ y \circ z)_{ijk} = x_i y_j z_k$].

The optimal solutions for x, y, and z separately can still be written in closed-forms and Algorithm ℓ_2-NMU can be easily generalized.

References

1. D. Lee and H. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature* **401**, 788–791 (1999).
2. M. Berry, M. Browne, A. Langville, P. Pauca, and R. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Comput. Stat. Data Anal.* **52**, 155–173 (2007).
3. J. Boardman, "Geometric mixture analysis of imaging spectrometry data," in *Proc. IGARSS 4*, Pasadena, pp. 2369–2371 (1994).
4. M. Craig, "Minimum-volume tranforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.* **32**(3), 542–552 (1994).
5. P. Pauca, J. Piper, and R. Plemmons, "Nonnegative matrix factorization for spectral data analysis," *Linear Algebra Appl.* **406**(1), 29–47 (2006).
6. Q. Zhang, H. Wang, R. Plemmons, and P. Pauca, "Tensor methods for hyperspectral data analysis: a space object material identification study," *J. Opt. Soc. Am. A* **25**(12), 3001–3012 (2008).
7. S.A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM J. Optim.* **20**(3), 1364–1377 (2009).
8. A. Cichocki, S. Amari, R. Zdunek, and A. Phan, *Non-negative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley and Sons, Chichester, UK (2009).
9. H. Laurberg, M. Christensen, M. Plumbley, L. Hansen, and S. Jensen, "Theorems on positive data: On the uniqueness of NMF," *Comput. Intell. Neurosci.* **2008**, 764206 (2008).
10. P. Hoyer, "Nonnegative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.* **5**, 1457–1469 (2004).
11. C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix tri-factorizations for clustering," in *Proc. SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 126–135 (2006).
12. H. Li, C. Adal, W. Wang, D. Emge, and A. Cichocki, "Non-negative matrix factorization with orthogonality constraints and its application to Raman spectroscopy," *J. VLSI Signal Process.* **48**, 83–97 (2007).
13. L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.* **45**(3), 765–777 (2007).
14. Y. Masalmah and M. Veléz-Reyes, "A full algorithm to compute the constrained positive matrix factorization and its application in unsupervised unmixing of hyperspectral imagery," *Proc. SPIE* **6966**, 779444, (2008).
15. A. Ifarraguerri and C.-I. Chang, "Multispectral and hyperspectral image analysis with convex cones," *IEEE Trans. Geosci. Remote Sens.* **37**(2), 756–770 (1999).
16. N. Gillis and F. Glineur, "Using underapproximations for sparse nonnegative matrix factorization," *Pattern Recogn.* **43**(4), 1676–1687 (2010).

17. N. Gillis and R. Plemmons, "Dimensionality reduction, classification, and spectral mixture analysis using nonnegative underapproximation," *Proc. SPIE* **7695**, 76951A (2010).

18. G. Golub and C. Van Loan, *Matrix Computation*, 3rd ed., Johns Hopkins University Press, Baltimore (1996).

19. N. Gillis, "Approximation et sous-approximation de matrices par factorisation positive: algorithmes, complexité et applications," MS Thesis, Université catholique de Louvain (2007). In French.

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

21. L. Ford, Jr. and D. Fulkerson, *Flows in Networks*, Princeton University Press, Princeton (1962).

22. C. Gurwitz, "Weighted median algorithms for ℓ_1 approximation," *BIT Numer. Math.* **30**(2), 301–310 (1990).

23. D. Chen and R. Plemmons, "Nonnegativity constraints in numerical analysis," in *Proc. of Symp. on the Birth of Numerical Analysis*, A. Bultheel and R. Cools, Eds., World Scientific Press, Leuven, pp. 109–140 (2009).

24. A. Szlam, Z. Guo, and S. Osher, "A split bregman method for nonnegative sparsity penalized least squares with applications to hyperspectral demixing," UCLA Dept. of Mathematics, CAM Report No. 10-6 (2010).

25. Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," *Proc. SPIE* **7334**, 73341M (2009).

26. B. Klingenberg, J. Curry, and A. Dougherty, "Non-negative matrix factorization: Ill-posedness and a geometric algorithm," *Pattern Recogn.* **42**(5), 918–928 (2009).

27. K. Burgers, Y. Fessehatsion, S. Rahmani, J. Seo, and T. Wittman, "A comparative analysis of dimension reduction algorithms on hyperspectral data," LAMDA Research Group, August 2009, http://www.math.ucla.edu/~wittman (2009).

28. A. Ross, Iris dataset obtained by Wake Forest from West Virginia University, Available at http://www.wvu.edu/ (2010).

29. C. Boyce, A. Ross, M. Monaco, L. Hornak, and X. Li, "Multispectral iris analysis: a preliminary study," in *Proc. of Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pp. 51–60, (2006).

30. M. Forkin, P. Pauca, R. Plemmons, and Q. Zhang, "Clustering and data fusion for ocular recognition," Preprint, Dept. of Computer Science, Wake Forest University (2010).

31. F. Li, M. Ng, and R. Plemmons, "Coupled segmentation and denoising/deblurring models for hyperspectral material identification," *Numer. Linear Alg. Appl.* (in press).

32. F. Li, M. Ng, R. Plemmons, S. Prasad, and Q. Zhang, "Hyperspectral image segmentation, deblurring, and spectral analysis for material identification," *Proc. SPIE* **7701**, 770103 (2010).

33. M. Winter, "N-findr: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," *Proc. SPIE* **3753**, 266 (1999).

Nicolas Gillis received his MS in applied mathematics from Université catholique de Louvain (UCL) in 2007. He is a PhD student in applied mathematics associated with the Center for Operations Research and Econometrics and the Department of Mathematical Engineering at UCL. He is also a F.R.S.-FNRS research fellow.

Robert J. Plemmons is currently the Z. Smith Reynolds Professor of Mathematics and Computer Science at Wake Forest University. His research is in computational mathematics applied to problems arising in image and signal processing and optics. His work is currently supported by grants from the AFOSR, the Intelligence Advanced Research Projects Activity, and the National Geospatial-Intelligence Agency. Plemmons received his BS in mathematics and physics from Wake Forest in 1962 and his PhD in applied mathematics from AUburn University in 1965. Before joining the Wake Forest faculty, while at the University of North Carolina, he founded its Center for Research in Scientific Computation. In U.S. Department of Defense research for over 35 years, Plemmons is the author of more than 200 papers and three books on computational mathematics and applications.