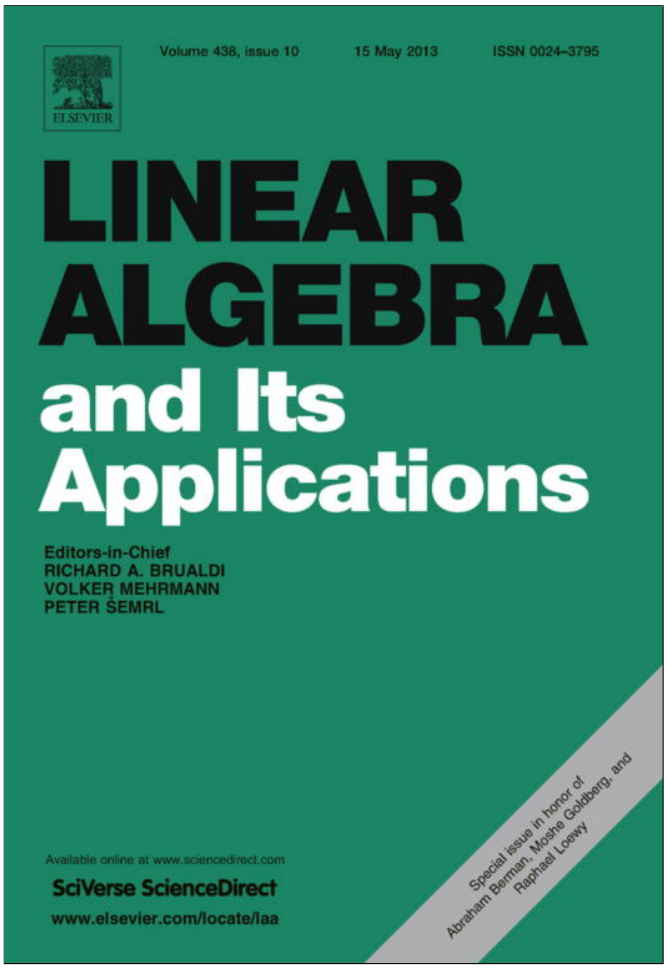

Linear Algebra and its Applications 438 (2013) 3991–4007

Contents lists available at SciVerse ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis ☆

Nicolas Gillis [a,*,1], Robert J. Plemmons [b,c]

[a] University of Waterloo, Department of Combinatorics and Optimization, Waterloo, Ontario, Canada N2L 3G1
[b] Department of Mathematics, Wake Forest University, Winston-Salem, NC 27106, United States
[c] Department Computer Science, Wake Forest University, Winston-Salem, NC 27106, United States

ARTICLE INFO

Article history:
Received 15 June 2011
Accepted 9 April 2012
Available online 16 May 2012

Submitted by E. Tadmor

AMS classification:
15A23
15B48
65F30

Keywords:
Nonnegative matrix factorization
Underapproximation
Sparsity
Hyperspectral images
Dimensionality reduction
Classification
Spectral mixture analysis

ABSTRACT

Dimensionality reduction techniques such as principal component analysis (PCA) are powerful tools for the analysis of high-dimensional data. In hyperspectral image analysis, nonnegativity of the data can be taken into account, leading to an additive linear model called nonnegative matrix factorization (NMF), which improves interpretability of the decomposition. Recently, another technique based on underapproximations (NMU) has been introduced, which allows the extraction of features in a recursive way, such as PCA, but preserving nonnegativity, such as NMF. Moreover, in some situations, NMU is able to detect automatically the materials present in the scene being imaged. However, for difficult hyperspectral datasets, NMU can mix some materials together, and is therefore not able to separate all of them properly. In this paper we introduce sparse NMU by adding a sparsity constraint on the abundance matrix and use it to extract materials individually in a more efficient way than NMU. This is experimentally demonstrated on the HYDICE images of the San Diego airport and the Urban dataset.

1. Introduction

Hyperspectral images are digital images, often taken either from an airplane or satellite, in which each pixel has not just the usual three visible bands of light (red at 650nm, green at 550nm, and blue

☆ Research supported in part by the U.S. Air Force Office of Scientific Research (AFOSR), under Grant Nos. FA9550-08-1-0151 and FA9550-11-1-0194, and by the U.S. National Geospatial-Intelligence Agency under Contract No. HM1582-10-C-0011.

* Corresponding author. Tel.: +1 519 888 4567x37529; fax: +1 519 725 5441.
E-mail addresses: ngillis@uwaterloo.ca (N. Gillis), plemmons@wfu.edu (R.J. Plemmons).

[1] This work was carried out the author was a Research Fellow of the Fonds de la Recherche Scientifique (F.R.S.-FNRS) at Université catholique de Louvain.

0024-3795/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.laa.2012.04.033

at 450nm), but on the order of hundreds of wavelengths so that spectroscopy can be conducted on the materials on the ground. The user is then able to identify, for instance, the species of trees and other vegetation, crop health, mineral and soil composition, moisture content of soils and vegetation, and pollution quantities. The technology also has clear military and homeland-security applications, as it enables identification of targets such as buildings and vehicles, even with attempts to camouflage. It is also possible to detect and identify gas plumes such as those arising from leaks even when the gases are invisible to the human eye. In fact, hyperspectral imaging was used following the attack on the twin towers and the hurricane Katrina disaster to identify dangerous gas leaks, providing guidance and protection to rescuers. See [4,2] for comprehensive overviews of recent hyperspectral image analysis trends.

The standard linear mixing model for hyperspectral image processing and analysis is based on the following general assumption: the spectral signature of each pixel results from the additive linear combination of the spectral signatures of the constitutive materials present in this pixel (called end-members). More precisely, assume we construct a matrix M such that each entry M_{ij} of M is equal to the reflectance[2] of the ith pixel at the jth wavelength (i.e., each row $M_{i:}$ of M corresponds to the spectral signature of a pixel and each column $M_{:i}$ of M to an image at a given wavelength), then the model reads

$$M_{i:} \approx \sum_{k=1}^{r} U_{ik} V_{k:} \quad \forall i,$$

where r is the number of materials present in the hyperspectral image. In fact, the spectral signature of each pixel ($M_{i:}$, a row of M) is approximated by a nonnegative linear combination (with weights $U_{ik} \geqslant 0$, representing abundances) of end-members' signatures ($V_{k:} \geqslant 0$) approximating the true signatures of the constituent materials of the hyperspectral image. This corresponds exactly to the nonnegative matrix factorization (NMF) model: given $M \in \mathbb{R}^{m\times n}_{+}$ and an integer $r \leqslant \min(m, n)$, solve

$$\min_{U\in\mathbb{R}^{m\times r}, V\in\mathbb{R}^{r\times n}} ||M - UV||_F^2 \quad \text{such that } U \geqslant 0 \text{ and } V \geqslant 0. \tag{1.1}$$

However, NMF is typically not able to separate all end-members correctly because of the *non-uniqueness* of the solution [19]. In order to improve NMF performances for hyperspectral image analysis, one should incorporate prior information into the model, and take into account the characteristics of the solutions to make the problem more well-posed, e.g., sparsity of the abundance matrix and piecewise smoothness of the spectral signatures [15], orthogonality [20], minimum-volume [22], and sum-to-one constraint on the abundances [21].

Another approach introduced recently in [8] is based on the introduction of additional underapproximation constraints, enabling one to extract features in a recursive way, like PCA, but preserving nonnegativity, like NMF. At the first step of the recursion, the following problem, called (rank-one) nonnegative matrix underapproximation (NMU),

$$\min_{u\geqslant 0, v\geqslant 0} ||M - uv^T||_F^2 \text{ such that } uv^T \leqslant M, \tag{1.2}$$

is solved, and a nonnegative residual matrix $R = M - uv^T \geqslant 0$ is obtained. Hence the same procedure can be applied on R. After r steps, this technique provides us with a rank-r NMF of the data matrix M. Even though NMU will feature a higher approximation error than NMF for the same factorization rank r (because of the greedy approach), it has the following three advantages:

- The factorization rank r does not have to be known a priori: the recursion can go on as long as the approximation error is too large (as for PCA). In contrast, NMF requires the rank r to be chosen a priori, and the solutions have to be recomputed when the factorization rank is modified.

[2] The reflectance is the fraction of the incident electromagnetic power that is reflected by a surface at a given wavelength.

- Problem (1.2) in well-posed in the sense that the optimal solution is, under some mild assumptions, unique [10] (this is similar to PCA, where the optimal solution is unique if the singular values of M are distinct, see [12]).
- As for NMF, NMU leads to a separation by parts. Furthermore, the additional underapproximation constraints enhance this property of the decomposition [8].

Moreover, it has been explained and experimentally shown how NMU is able to extract materials automatically in hyperspectral images [10,17,18]. Another advantage of NMU is that no parameters have to be tuned (i.e., it is completely unsupervised), as opposed to most constrained NMF techniques (see above). However, in some situations when blurring and noise, limited resolution, and mixed materials are present, NMU can also have some difficulties in separating all end-members properly. A possible way to overcome this is to combine NMU with standard clustering techniques, such as k-means or nearest neighbor.

In this paper, we propose to incorporate sparsity into the NMU model, leading to improved performances for hyperspectral image analysis in terms of end-members separation. The paper is organized as follows. In Section 2, we introduce the problem of interest, namely sparse NMU. In Section 3, we prove that sparse NMU is able to identify materials perfectly in ideal cases, and explain how it is able to deal with non-ideal situations. In Section 4, we propose an algorithm to find approximate solutions to sparse NMU and discuss its link to compressive sensing. In Section 5, we apply the new technique to the analysis of HYDICE images, and show qualitatively that it is able to separate end-members better than NMU. A preliminary and abbreviated conference version of this paper was published by IEEE in the conference proceedings [11].

Notation. The set of m-by-n real matrices is denoted $\mathbb{R}^{m\times n}$; for $A \in \mathbb{R}^{m\times n}$, $A_{:j}$ is the jth column of A, $A_{i:}$ is the ith row of A, and A_{ij} the entry at position (i, j), and $A(I, J)$ is the submatrix of A with row (resp. column) indices in the set I (resp. J). For $b \in \mathbb{R}^m$, we denote b_i as the ith entry of b. The set $\mathbb{R}^{m\times n}$ with component-wise nonnegative entries is denoted $\mathbb{R}^{m\times n}_+$. The ℓ_0-'norm' of vector x denoted $||x||_0$ is the cardinality of the set $\{i|x_i \neq 0\}$. The ℓ_2-norm of vector x is $||x||_2 = \sqrt{\sum x_i^2}$; the ℓ_∞-norm of vector x is $||x||_\infty = \max_i |x_i|$; the ℓ_1-norm of vector x is $||x||_1 = \sum_i |x_i|$; the squared Frobenius norm of matrix A is $||A||_F^2 = \sum_{i,j} a_{ij}^2$.

2. Enhancing sparsity in NMU

A possible way to enforce separation of the end-members is to add sparsity constraints on the abundance matrix U in NMF [15]. In fact, each material is typically present in a relatively small number of pixels and each pixel contains only a small number of constitutive materials. Thus U should be sparse, since each entry U_{ik} of U corresponds to the abundance of material k in pixel i. The reason why the original NMU approach sometimes mixes several materials together is because the vector u obtained when solving (1.2) is not sparse enough. In order to enforce sparsity, a ℓ_0-pseudo-norm penalty term can be incorporated into the NMU model. Ideally, we then would like to solve at each step of the recursion the following problem:

$$\min_{u\geqslant 0, v\geqslant 0} \quad ||M - uv^T||_F^2 + \mu||u||_0 \ \text{ such that } \ uv^T \leqslant M, \tag{2.1}$$

for some penalty parameter $\mu \geqslant 0$. The ℓ_0-norm $||u||_0$ of u is equal to the number of non-zero entries of u, hence incites u to be sparser. We will refer to this problem as *sparse NMU*.

There are three main pitfalls in using sparse NMU in practice, each of them will be addressed in the subsequent sections:

1. We will see that sparse NMU is able to detect the materials present in hyperspectral images in ideal conditions, see Section 3. It is not clear however how sparse NMU can deal with blurry and noisy

data. Even though no theoretical guarantees will be provided, we will demonstrate on numerous selected examples that sparse NMU is able to deal with non-ideal conditions.

2. The original NMU problem (1.2) was proved to be NP-hard [8], and it is not likely that the additional penalty term in the objective function will make the problem easier. We will propose in Section 4 an algorithm to find approximate solution to sparse NMU using the ℓ_1-norm heuristic approach combined with standard nonlinear optimization schemes. Because of NP-hardness, it is unlikely that a guarantee of global optimality can be provided (unless $\mathcal{P} = \mathcal{NP}$). However, we will see in the numerical experiments that our sparse NMU algorithm seems to be robust in the sense that it generates very similar solutions for different initializations.
3. A good sequence of parameters μ has to be chosen. In fact, at each step of the recursion, one has to decide on a penalty parameter μ and then solve the corresponding sparse NMU problem, see Equation (2.1). For similar problems (such as sparse PCA and sparse NMF), a trial and error approach is typically used in practice,[3] which will also be our framework. Notice that this is a crucial step, which is not easily handled in practice. This issue will be discussed in Section 5, along with the numerical experiments.

3. What can we expect from sparse NMU?

In this section, we analyze the sparse NMU formulation and characterize some of its properties. We start with the ideal case, and then give some clues of how sparse NMU behaves in non-ideal cases.

3.1. Ideal case

In the ideal case, i.e., when each pixel of the hyperspectral image contains only one material (see Assumption 1 below), the unmixing problem can be solved easily (simply compare the different spectral signatures). However, it is still interesting to verify that our model is able to identify the different materials in these settings (for example, NMF is not guaranteed to recover the materials because of the non-uniqueness issues, see the introduction):

Assumption 1. *Let $M \in \mathbb{R}^{m\times n}_+$ correspond to a hyperspectral image with m pixels and n spectral bands, i.e., M_{ij} is equal to the reflectance of the ith pixel at the jth wavelength. Moreover, M is constituted of r materials with linearly independent spectral signatures $H_{k:} \in \mathbb{R}^n_+$ $1 \leqslant k \leqslant r$, and each pixel contains only one material, i.e., for all i there exists k such that $M_{i:} = H_{k:}$. Therefore, there exists a matrix $W \in \{0, 1\}^{m\times r}$ with one and only one non-zero entry in each row such that $M = WH = \sum_{k=1}^{r} W_{:k}H_{k:}$ and* $\operatorname{rank}(M) = r \leqslant \min(m, n)$.

Theorem 1. *Let $M = WH$ be a matrix satisfying Assumption 1. Then there exists $\mu \in \mathbb{R}^k_+$ such that the following procedure:*

1: $R_1 = M$;
2: ***for*** $k = 1 : r$ ***do***
3: $[u, v] =$ *optimal solution of* (2.1) *with* $M = R_k$ *and* $\mu = \mu_k$;
4: $K = \max_i(u_i)$; $U_{:k} = u/K$; $V_{k:} = Kv^T$; $R_{k+1} = R_k - uv^T$.
5: ***end for***

recovers the r original materials in r steps, i.e., $U = W$ and $V = H$ (up to a permutation).

Proof. We are going to prove that there exists $\mu \in \mathbb{R}_+$ such that any optimal solution (u^*, v^*) of (2.1) for matrix M satisfying Assumption 1 corresponds to one of the materials, i.e., $u^* = W_{:k}$ and $v^* = H^T_{k:}$ for some $1 \leqslant k \leqslant r$ (up to a multiplicative factor). The first residual R will then be equal to $R = M - u^*v^{*T} = \sum_{i\neq k} W_{:i}H_{i:}$, and all rows of R corresponding to the first extracted material

[3] Except if one wants a solution of a specific sparsity, in which case the parameter μ can be easily tuned in the course of the optimization process.

will be identically zero. Hence, the next step of the recursion reduces to the same problem with $r-1$ materials, and, by recursion, the proof will be complete.

Without loss of generality, let us reorder the columns of W and the rows of H in such a way that:

$$||H_{1:}||_2 = ||H_{2:}||_2 = \cdots = ||H_{p:}||_2 > ||H_{(p+1):}||_2 \geqslant ||H_{(p+2):}||_2 \geqslant \cdots \geqslant ||H_{r:}||_2,$$

and $||W_{:1}||_2 \geqslant ||W_{:2}||_2 \geqslant \cdots \geqslant ||W_{:p}||_2$. Let then (u^*, v^*) be an optimal solution of (2.1), and note $I_k = \{1 \leqslant i \leqslant m \mid W_{ik} = 1\}$ the set of pixels containing the kth material, with $|I_k| = ||W_{:k}||_2^2$, and $K = \{1 \leqslant k \leqslant m \mid u_k^* > 0\}$ the index set of the non-zero entries of u^*. We observe that, for $\mu < ||H_{1:}||_2^2$,

1. The feasible solution $(u = W_{:1}, v = H_{1:}^T)$ is strictly better than the trivial solution with
$$||M - W_{:1}H_{1:}||_F^2 + \mu||W_{:1}||_0 = ||M||_F^2 - ||W_{:1}||_2^2(||H_{1:}||_2^2 - \mu) < ||M||_F^2,$$
hence $u^* \neq 0$ and $v^* \neq 0$ at optimality and K is non-empty.
2. The objective function can be decoupled into m independent terms: $\sum_{i=1}^m ||M_{i:} - u_i^* v^{*T}||_2^2 + \mu||u_i^*||_0$, we then have either
   (a) $u_i^* = 0$, and the corresponding term in the objective function is equal to $||M_{i:}||_2^2$, or
   (b) $u_i^* > 0$, and the corresponding term in the objective function is larger than μ.

   Since u^* is optimal, we then must have
   $$u_i^* = 0 \text{ for all } i \text{ such that } \mu > ||M_{i:}||_2^2.$$
   Therefore, for $||H_{(p+1):}||_2^2 < \mu$, we have $K \cap I_k = \emptyset$ for any $p+1 \leqslant k \leqslant r$ since $u_i^* = 0$ for all i such that $||M_{i:}||_2 < ||H_{1:}||_2$.
3. If $K \subseteq I_k$ for some $1 \leqslant k \leqslant p$ (i.e., K contains only one material), then by optimality we have $v^* = H_{k:}$ and $u_i^* = 1$ $i \in K$ (up to a multiplicative factor). Moreover, we must have $K = I_k$ since taking $u_i^* > 0$ for $i \in I_k$ is optimal for $\mu < ||H_{k:}||_2^2$ (see point 2. above).

It remains to show that K does not contain two indices corresponding to two different materials, i.e., that there does not exist $i, j \in K$ such that $M_{i:} \neq M_{j:}$. Assume that K does contain two such indices, and let us denote $d = ||M||_F^2 + \mu|K|$. We can obtain a lower bound for the optimal value e^* of (2.1) by relaxing the underapproximation constraints: the optimal rank-one approximation of $M(K, :)$ has an error of $||M(K, :)||_F^2 - \sigma_1^2(M(K, :))$ [12] (where $\sigma_1(X)$ denotes the largest singular value of matrix X), hence

$$e^* = ||M - u^* v^{*T}||_F^2 + \mu|K| \geqslant d - \sigma_1^2(M(K, :)).$$

Notice that $\sigma_1^2(M(K, :)) < \sigma_1^2(M(K, :)) + \sigma_2^2(M(K, :)) \leqslant ||M(K, :)||_F^2 = |K|||H_{1:}||_2^2$ by the linear independence of the rows of H; in fact, this assumption implies that the second largest singular value $\sigma_2(M(K, :))$ of $M(K, :)$ is positive. Moreover using the value of the objective function of the trivial solution $u = 0$, we obtain the following upper bound

$$e^* \leqslant d - \mu|K|.$$

Choosing μ such that $\frac{\sigma_1^2(M(K,:))}{|K|} < \mu < ||H_{1:}||_2^2$ leads to a contradiction.

Finally, taking μ such that

$$\max\left(||H_{(p+1):}||_2^2, \max_{\substack{K \subseteq \cup_{1 \leqslant k \leqslant p} I_k, \\ \exists i, j \in K \text{ s.t. } M_{i:} \neq M_{j:}}} \frac{\sigma_1^2(M(K, :))}{|K|} \right) < \mu < ||H_{1:}||_2^2,$$

implies that K contains only one material, i.e., $K \subset I_k$ for some $1 \leqslant k \leqslant p$.

The objective function value achieved by selecting a single material $K = I_k$ for some $1 \leqslant k \leqslant p$ is given by $||M||_F^2 - |I_k|(||H_{k:}||_2^2 - \mu)$. The lowest value is attained by the materials with largest $|I_k|$ $(= ||W_{:k}||_F^2)$; we therefore have $q \leqslant p$ optimal solutions (up to multiplicative factor) for $||W_{:1}||_F^2 = \cdots = ||W_{:q}||_F^2$. This means that when there are different materials whose spectral signatures have the same norm (i.e., $||H_{1:}||_2 = ||H_{2:}||_2 = \cdots = ||H_{p:}||_2$) then solving sparse NMU will recover any of the materials contained by the largest number of pixels. □

It is worth noting that Theorem 1 also applies to sparse NMF, i.e., solving the following problem

$$\min_{u \in \mathbb{R}_+^m, v \in \mathbb{R}_+^n} ||M - uv^T||_F^2 + \mu ||u||_0,$$

instead of (2.1) will also allow to recover the r materials in r steps. In fact, the proof of Theorem 1 does not consider explicitly the underapproximation constraint (but shows that the optimal solution satisfies it automatically). However, in practice, we cannot get rid of this constraint because we are dealing with non-ideal datasets.

More importantly, Theorem 1 shows that sparse NMU is a candidate for extracting materials in hyperspectral images. At least it works potentially better than standard NMU, which might require much more recursion steps to extract each material individually [10].

3.2. Non-ideal case

Theorem 1 sheds light on the choice of μ when solving the sparse NMU problem: this choice is very sensitive to the difference between the spectral signatures. The optimal solution also depends on the abundance of each material present in the scene being imaged. In particular, we have observed that if μ is chosen smaller than the value prescribed in Theorem 1, then a material whose spectral signature does not have the largest norm but is contained by many pixels will be extracted (more precisely, the material k for which $||W_{:k}||_2^2(||H_{k:}||_2^2 - \mu)$ is maximum). Moreover, solving (2.1) approximately or dealing with noisy and/or blurry data may lead to mixing materials together. It is interesting to observe that a pixel whose spectral signature s is very similar to the spectral signature h of the extracted material (more precisely such that $\min_{x \geqslant 0, xh \leqslant s} ||s - xh||_2^2 + \mu < ||s||_2^2$, cf. proof of Theorem 1) will also be extracted in the first step of the sparse NMU recursion. Therefore, one can show that Theorem 1 still holds if the spectral signatures $M_{i:}$'s are affected by a small noise level. Finally, the robustness of sparse NMU will be directly related to the differences between the spectral signatures and the distribution of materials among the pixels: more precise characterization of its behavior in non-ideal cases is a topic for further research.

Let us however show a simple example which illustrates the behavior of sparse NMU in non-ideal conditions: we construct the matrix $M = UV \in \mathbb{R}^{9 \times 12}$ with

$$U = \begin{pmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.9 & 0.1 \\ 0.1 & 0 & 0.9 \\ 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.50 & 0 & 0.5 \end{pmatrix}, \quad V = \begin{pmatrix} 8 & 0 & 7 & 5 & 9 & 10 & 1 & 1 & 4 & 0 & 2 & 2 \\ 2 & 3 & 9 & 4 & 2 & 1 & 1 & 5 & 8 & 6 & 9 & 9 \\ 4 & 8 & 1 & 3 & 4 & 3 & 2 & 8 & 8 & 1 & 1 & 7 \end{pmatrix}. \tag{3.1}$$

Table 1
Basis elements obtained with NMF, NMU and sparse NMU.

NMF	NMU	Sparse NMU
$\begin{pmatrix} 0.11 & 0.90 & 0 \\ 0.90 & 0 & 0.43 \\ 0 & 0.23 & 0.90 \\ 0.11 & 0.81 & 0.11 \\ 0.80 & 0.10 & 0.39 \\ 0.10 & 0.21 & 0.84 \\ 0.50 & 0.49 & 0.16 \\ 0.50 & 0.06 & 0.68 \\ 0 & 0.57 & 0.48 \end{pmatrix}$	$\begin{pmatrix} 0.43 & 0.90 & 0.02 & 0 \\ 0.80 & 0 & 0.90 & 0.08 \\ 0.64 & 0.11 & 0 & 0.90 \\ 0.53 & 0.79 & 0.01 & 0 \\ 0.88 & 0 & 0.76 & 0.02 \\ 0.75 & 0.06 & 0 & 0.76 \\ 0.71 & 0.41 & 0.36 & 0 \\ 0.90 & 0 & 0.29 & 0.27 \\ 0.70 & 0.35 & 0 & 0.24 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.90 & 0 \\ 0.90 & 0 & 0 \\ 0.15 & 0.02 & 0.90 \\ 0 & 0.86 & 0 \\ 0.82 & 0 & 0 \\ 0.26 & 0 & 0.75 \\ 0.38 & 0.18 & 0 \\ 0.67 & 0 & 0.12 \\ 0 & 0.62 & 0.24 \end{pmatrix}$

The multispectral image corresponding to M therefore contains three end-members whose spectral signatures are the rows of V, and the corresponding abundances for each pixel are given by the rows of U. Observe that there is no pure pixel, even though some pixels contain mostly one material. Table 1 displays the basis elements obtained by NMF (using the accelerated HALS algorithm [9]), NMU and sparse NMU (using the algorithm described in Section 4).

We observe the following:

- Even though NMF gives an optimal solution (i.e., $||M - UV||_F = 0$) and is able to extract perfectly the second material, it mixes the two other ones (because of non-uniqueness). We note that the solution is very sensitive to initialization (sometimes it finds back the three materials, sometimes it mixes all of them).
- NMU first extracts a positive factor, which is always the case for a positive matrix M. This amounts to subtracting from each spectral signature the same component, see [10]. Then, the next three factors of NMU provide a very good (soft) clustering of the pixels, being close to the original columns of the abundance matrix U.
- Sparse NMU clusters the pixels containing mainly one material (i.e., abundance larger than 0.8) in only one category (except for the 3rd and 6th pixel). It essentially considers low abundances as noise (of course, this will strongly depend on the choice of the parameter μ).[4]

We have observed that, in general, NMU and sparse NMU are less sensitive to initialization and typically find solution with identical sparsity patterns.

4. Algorithm for sparse NMU

We now present an algorithm to find approximate solutions to Problem (2.1), i.e., sparse NMU. Of course the ℓ_0-norm is very difficult to work with, and we propose to use the standard ℓ_1-norm heuristic approach (see, e.g., [16]):

$$\begin{aligned} \min_{u\geqslant 0, v\geqslant 0, \sigma\geqslant 0} \quad & ||M - \sigma uv^T||_F^2 + \mu||u||_1 \\ \text{such that} \quad & \sigma uv^T \leqslant M, \\ & ||u||_2 = 1, ||v||_2 = 1. \end{aligned} \tag{4.1}$$

The theoretical reason for this choice is the following. The convex envelope of a function $f(x)$ over the convex set X is the largest convex function $g(x)$ such that $g(x) \leqslant f(x)\ \forall x \in X$. The convex envelope of the ℓ_0-norm $f(x) = ||x||_0$ over the set $X = \{x \in \mathbb{R}^n \mid ||x||_\infty \leqslant 1\}$ is given by the ℓ_1-norm $g(x) = ||x||_1$ (see, e.g., [24]). That is, $||x||_1$ is the best convex function approximating $||x||_0$ from below for $||x||_\infty \leqslant 1$.

[4] We chose $\lambda = (0.8\ 0.5\ 0.2)$, see Algorithm 1.

4.1. Link with compressed sensing

Notice that problem (4.1) shares some similarities with compressed sensing (CS) [3,6,5]. When noise is present, the optimization problem to be solved in CS is as follows

$$\min_{x\in\mathbb{R}^q} ||x||_1 \text{ such that } ||Ax-b||_2 \leqslant \epsilon, \tag{4.2}$$

for some parameter $\epsilon \geqslant 0$, $b \in \mathbb{R}^p$ and $A \in \mathbb{R}^{p\times q}$ where $p \gg q$. Under some assumptions on the matrix A (roughly, any subset of k columns of A can only be loosely correlated; where k depends on the sparsity of the sought solution), one can show that solving (4.2) (which can be done efficiently with convex optimization solvers) will indeed recover the sparsest solution x satisfying $||Ax-b||_2 \leqslant \epsilon$.

Our problem (4.1) can be equivalently reformulated as

$$\min_{u\in\mathbb{R}^m, ||u||_2=1, u\geqslant 0} ||u||_1 \text{ such that } f(u) \leqslant \delta, \tag{4.3}$$

for some δ depending on μ, and where $f(u) = \min_v ||M-uv^T||_F^2$ such that $uv^T \leqslant M$. Unfortunately, it seems that applying the results from the CS literature to our formulation (4.1) would require extensive investigation and might not apply to hyperspectral images. In fact,

1. The function $f(u)$ is highly non-convex, and the feasible domain $\{u \in \mathbb{R}^m \mid ||u||_2 = 1, u \geqslant 0\}$ is also non-convex so that, if a proof of optimality would exist, it could be much more complicated to establish. Moreover, even if one were able to show that any optimal solution of (4.3) is the sparsest possible, we would, in general, not be able to compute such a solution since Problem (4.3) is NP-hard [8].
2. In hyperspectral images, the columns of matrix M are highly correlated (columns of M are images of the same scene taken at different wavelengths) and $m \gg n$ (m represents the number of pixels while n is the number of spectral bands), so that even if one could generalize the results from CS to (4.3), it would most likely not apply directly to our particular hyperspectral imaging application of finding abundance maps using sparse NMU. However, ℓ_1 methods in unmixing and template matching problems for hyperspectral imaging can be directly related to compressive sensing, as shown in [14] and [13], respectively.

Hence, our use of the ℓ_1-norm as a surrogate for the ℓ_0-norm is, at this point, only a heuristic motivated by the fact that the former is the convex envelope of the latter (see above). In the following, we propose a simple approach to find good approximate solutions to (4.1).

4.2. Algorithm based on Lagrangian relaxation

Following the work of [8] (see also [10]), we propose a Lagrangian dual approach to solve Problem (4.1). Let us introduce Lagrangian dual variables $\Lambda \in \mathbb{R}_+^{m\times n}$ associated with the underapproximation constraints and write the Lagrangian dual:

$$\sup_{\Lambda\geqslant 0} L(\Lambda), \tag{4.4}$$

where $L(\Lambda)$ is the Lagrangian dual function:

$$L(\Lambda) = \min_{\sigma,u,v\geqslant 0, ||u||_2=||v||_2=1} ||M-\sigma uv^T||_F^2 + \mu||u||_1 - 2\left\langle \Lambda, M-\sigma uv^T\right\rangle. \tag{4.5}$$

Problem (4.4) is a non-smooth convex optimization problem (see, e.g., [1]) and can be solved using the following subgradient scheme [25]:

0. Initialize $\Lambda^{(0)}$. Then for $k = 0, 1, 2, \ldots$
1. Find a solution (σ, u, v) of (4.5) for $\Lambda = \Lambda^{(k)}$.

2. Update $\Lambda^{(k+1)} \leftarrow \max\left(0,\ \Lambda^{(k)} + \alpha_k(\sigma uv^T - M)\right)$.

If the Lagrangian relaxation subproblems (4.5) are solved exactly at each step and, for example, $\alpha_k = 1/k$, then the convergence of the above scheme is guaranteed [1].

4.3. Solving the Lagrangian relaxation

Unfortunately, the Lagrangian relaxation Problem (4.5) is NP-hard [8]. We then propose to use a simple exact block-coordinate descent scheme to find good solutions to Problem (4.5). This amounts to successively optimizing over one bock of variables while keeping the other fixed:

0. Initialize (σ, u, v) (see Section 4.4 below). Then for $k = 0, 1, 2, \ldots$
1. $u \leftarrow \text{argmin}_{u \geqslant 0, ||u||_2=1} ||(M - \Lambda) - \sigma uv^T||_F^2 + \mu||u||_1$.
2. $v \leftarrow \text{argmin}_{v \geqslant 0, ||v||_2=1} ||(M - \Lambda) - \sigma uv^T||_F^2$.
3. $\sigma \leftarrow \text{argmin}_{\sigma \geqslant 0} ||(M - \Lambda) - \sigma uv^T||_F^2 = \max(0, u^T(M - \Lambda)v)$.

One advantage of this approach is that the optimal solutions of these subproblems can be written in closed form (cf. Algorithm 1 to follow[5]) and are uniquely attained when σ, u and v are different from zero (which guarantees the limit points to converge to a stationary point [23]). In practice, we observed that updating only σ, u and v once between each update of Λ seems to give good results. This approach is implemented in Algorithm 1.

4.4. Initialization and choice of regularization parameter μ

At each step of the recursion, variables σ, u, v and Λ have to be initialized. Following [10], we initialize

1. The variables (σ, u, v) with an optimal nonnegative rank-one approximation (σ^*, u^*, v^*) of $M \geqslant 0$. This can be computed, for example, using the power method [12].
2. The Lagrangian dual variables Λ with the nonnegative part of the residual, given by $\max\left(0, -(M - \sigma^* u^* v^{*T})\right)$.

The next update of u would then be given by (before projection onto the ℓ_2-ball)

$$\text{argmin}_{u \geqslant 0} ||M - \Lambda - \sigma^* uv^{*T}||_F^2 + \mu||u||_1 = \max\left(0, \sigma^*(M - \Lambda)v^* - \mu\right).$$

Therefore parameter μ must be chosen smaller than $a = ||\sigma^*(M - \Lambda)v^*||_\infty$, otherwise u is set to zero at the first step. Introducing parameter λ such that $\mu = \lambda a$, we then must have $\lambda \in [0, 1)$.

Notice that the value of σ can only decrease in the course of the algorithm. In fact, σ is equal to $u^T Mv$ with $||u||_2 = ||v||_2 = 1$ hence σ is smaller than the largest singular value $\sigma_1(M)$ of M since, by definition, $\sigma_1(M) = \max_{||u||_2=||v||_2=1} u^T Mv$ [12]. This maximal value $\sigma_1(M)$ is attained with our initialization. Therefore, if the value of μ is kept fixed, it often happens that it becomes larger than $||\sigma(M - \Lambda)v||_\infty$ and u is then set to zero. We found it appropriate to change the value of μ proportionally to σ (i.e., if σ is multiplied by a factor, μ is multiplied by the same factor). This is equivalent to updating u as done in steps 6 and 10 of Algorithm 1. As a safety procedure, if we have $\max\left(0, \sigma(M - \Lambda)v\right) < \mu$ at some iterate, we reduce the value of μ, see steps 7–9 of Algorithm 1.

4.5. Lower and upper bounds on the sparsity of u

In case our parameter λ is chosen too large (resp. too small), we might obtain a solution which is too sparse (resp. too dense), hence containing only a very small (resp. large) subset of the pixels. In

[5] The code is available at http://sites.google.com/site/nicolasgillis/.

Algorithm 1 Sparse NMU

Require: $M \in \mathbb{R}^{m\times n}_+$, $r \in \mathbb{N}$, $\lambda \in [0,1)^r$, $0 \leqslant \delta < \Delta \leqslant 1$, maxiter.
Ensure: $(U, V) \in \mathbb{R}^{m\times r}_+ \times \mathbb{R}^{r\times n}_+$ s.t. $UV \lesssim M$ with U sparse.
1: **for** $k = 1 : \mathrm{r}$ **do**
2: $[\sigma, u, v]$ = optimal rank-one approximation(M);
3: $U(:,k) \leftarrow u$; $V(k,:) \leftarrow \sigma v^T$; $\Lambda \leftarrow \max(0, -(M - \sigma uv^T))$;
4: $\mu = \lambda_k ||(M-\Lambda)v||_\infty$;
5: **for** $p = 1 : \text{maxiter}$ **do**
6: $u \leftarrow \max(0, (M-\Lambda)v)$;
7: **if** $||u||_\infty \leqslant \mu$ **then**
8: $\mu \leftarrow 0.99||u||_\infty$;
9: **end if**
10: $u \leftarrow \max(0, u-\mu)$; $u \leftarrow u/||u||_2$;
11: **if** $||u||_0 \leqslant \delta m$ **then**
12: $\mu \leftarrow 0.95\mu$;
13: **else if** $||u||_0 > \Delta m$ **then**
14: $\mu \leftarrow 1.05\mu$;
15: **end if**
16: $v \leftarrow \max(0, (M-\Lambda)^T u)$; $v \leftarrow v/||v||_2$;
17: $\sigma \leftarrow u^T(M-\Lambda)v$;
18: **if** $\sigma > 0$ **then**
19: $U(:,k) \leftarrow u$; $V(k,:) \leftarrow \sigma v^T$;
20: $\Lambda \leftarrow \max\left(0, \Lambda - \frac{1}{p+1}(M - U(:,k)V(k,:))\right)$;
21: **else**
22: $\Lambda \leftarrow 0.95\Lambda$; $v \leftarrow V(k,:)^T$;
23: **end if**
24: **end for**
25: $M \leftarrow \max(0, M - U(:,k)V(k,:))$;
26: **end for**

fact, the support of factor u should ideally be the set of pixels containing a single material. Therefore, if available or if one is only interested in clusters whose number of pixels is in a given interval, a lower and upper bound on the sparsity of u can be imposed. We define new parameters $0 \leqslant \delta < \Delta \leqslant 1$ as these lower and upper bounds in percent of the total size of the image. If this lower (resp. upper) bound is reached for some iterate u, the value of parameter λ is decreased (resp. increased), see steps 11–15 of Algorithm 1. For $\lambda = \delta = 0$ and $\Delta = 1$, Algorithm 1 reduces to the algorithm from [10]. Moreover, it has *exactly* the same computational cost, with $O(Kmnr)$ operations where $K = \text{maxiter}$.

5. Numerical experiments

In this section, we qualitatively assess the good performance of sparse NMU on two real-world datasets for which the ground truth is not available (hence a rigorous quantitative comparison is not possible). We use 100 iterations of Algorithm 1 to solve the sparse NMU problems, i.e., maxiter $= 100$, and we visually adjusted the value of parameter λ (see also Section 5.1.2).

5.1. San Diego airport

The San Diego airport hyperspectral image is taken from HYper-spectral Digital Imagery Collection Experiment (HYDICE) air-borne sensors. It contains 158 clean bands, and 400×400 pixels for each spectral image (i.e., $M \in \mathbb{R}^{160000\times 158}_+$), see Fig. 1. There are four basic types of materials under

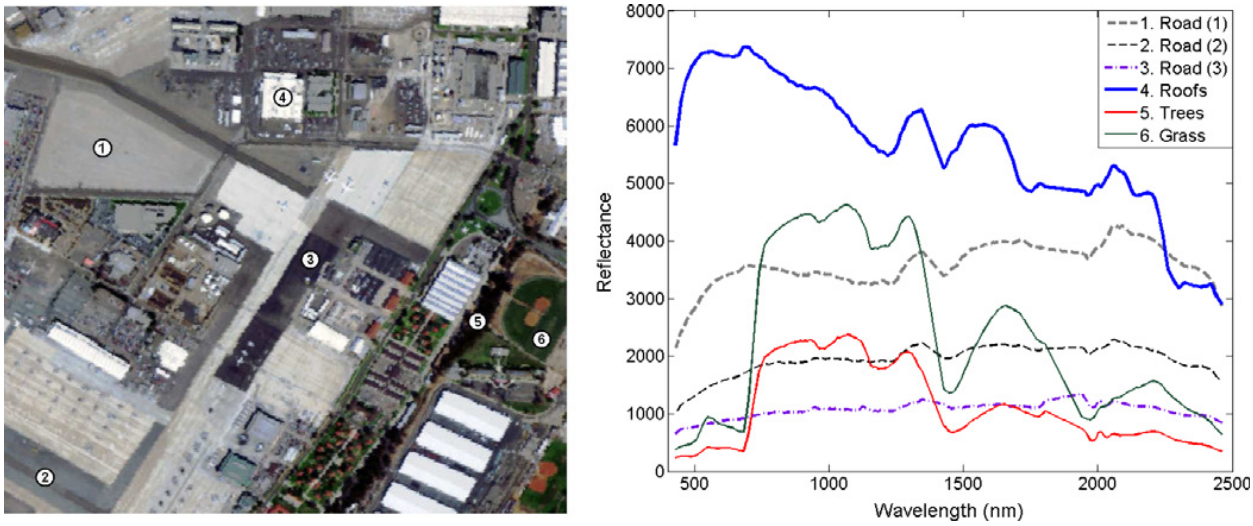

Fig. 1. San Diego airport.

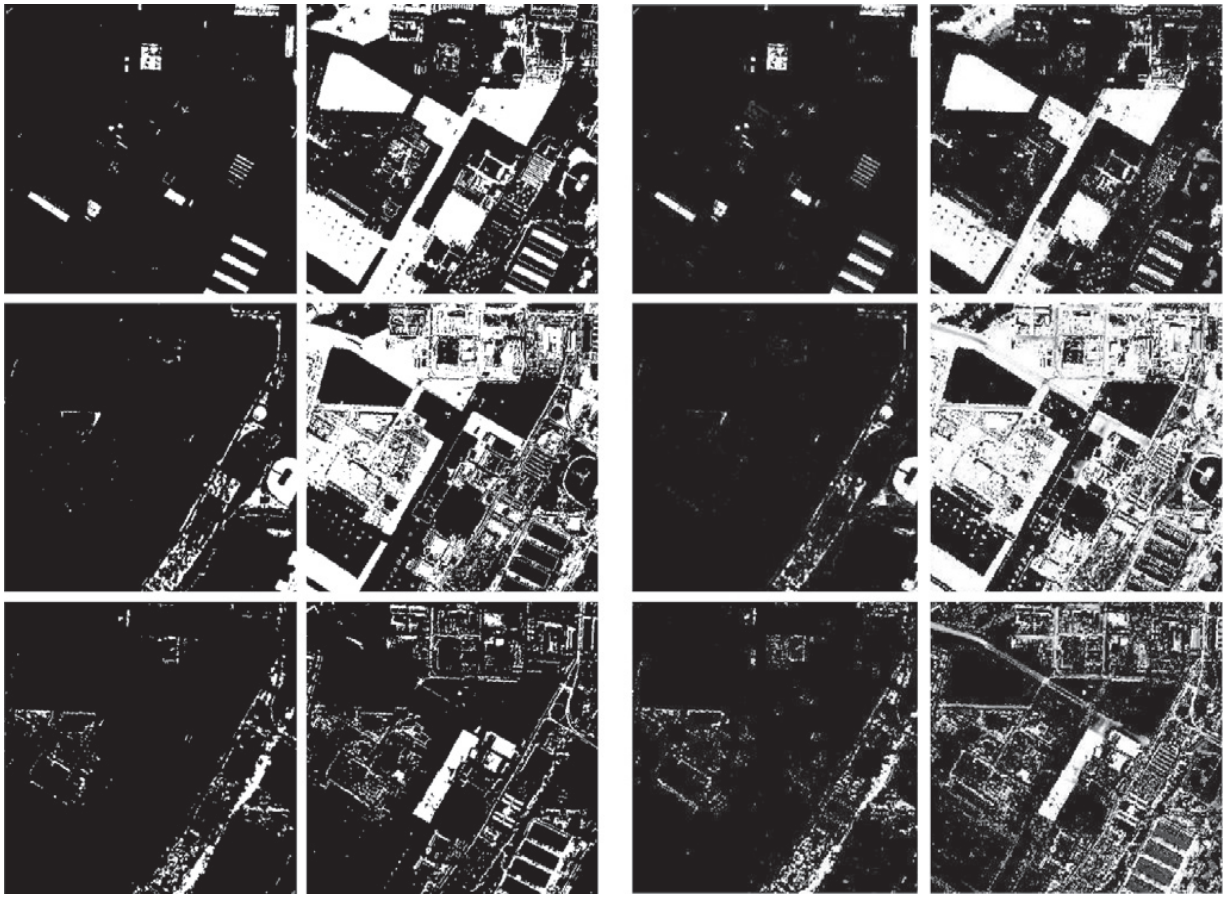

Fig. 2. Abundance maps obtained with a manually adjusted nearest neighbor technique [7] (left) and with sparse NMU (right).

consideration: road surface, roofs, trees and grass. Moreover, there are (mainly) three different types of road surfaces including boarding and landing zones, parking lots and streets. The spectral signatures of the end-members are shown on Fig. 1 and have been computed using a manually adjusted nearest neighbor technique from the HYPERACTIVE toolkit [7], see the left side of Fig. 2 for the corresponding abundance maps.

For this dataset, standard NMU is not able to separate the materials properly, see Fig. 3: it extracts materials together (e.g., the grass, the trees and the roofs in the second basis element) and cannot

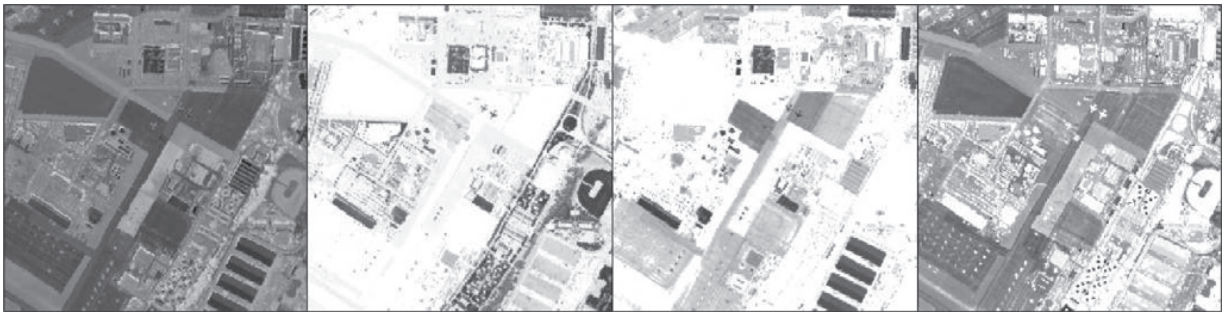

Fig. 3. First four basis elements obtained with NMU.

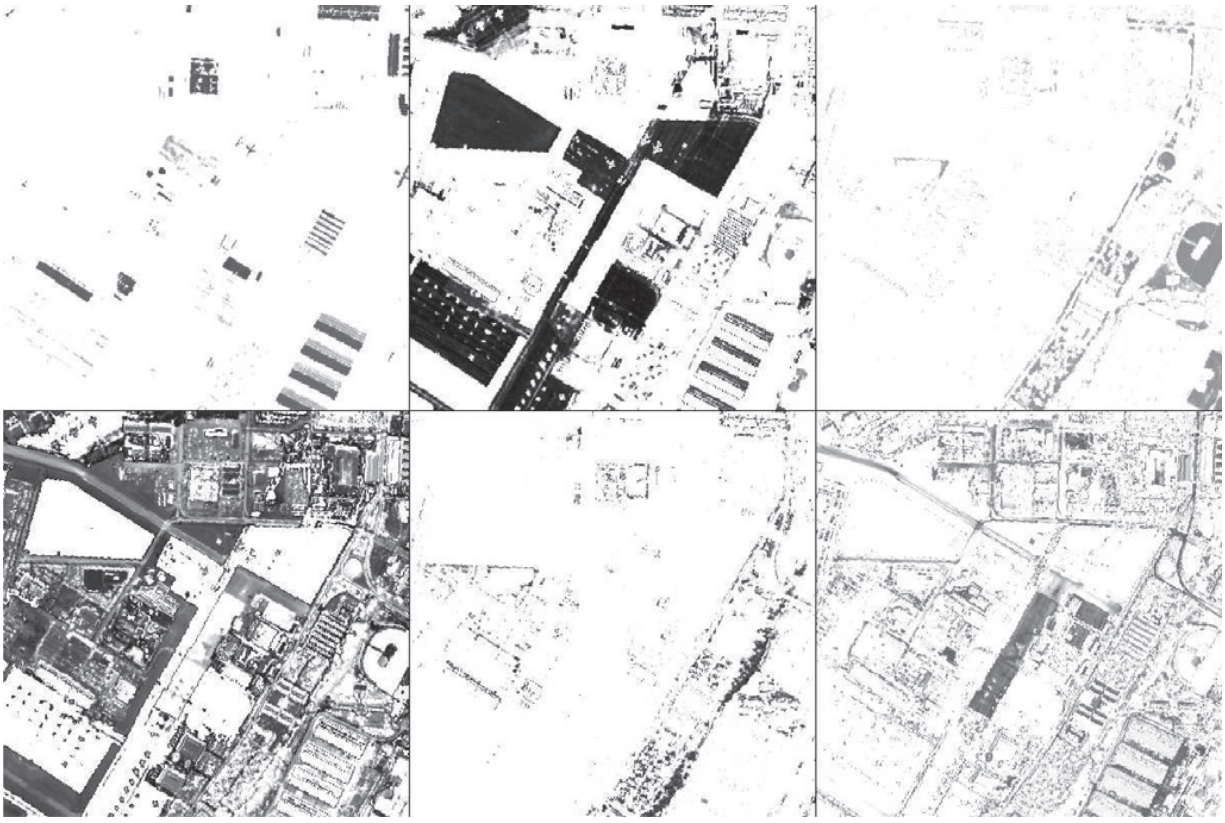

Fig. 4. First six basis elements obtained with sparse NMU with $\lambda = (0.2\ 0.3\ 0.2\ 0.2\ 0.2\ 0.05)$, $\delta = 0.01$, and $\Delta = 1$. Left to right and top to bottom: roofs, road type 1, grass, road type 2, trees, road type 3.

identify them individually (only the fourth basis element corresponds to the different types of road surfaces).

5.1.1. Spectral unmixing using sparse NMU

Fig. 4 displays the abundance maps (i.e., rows of matrix U) extracted with sparse NMU, which is able to extract the different materials: roofs in the first basis element, road type 1 in the second, grass in the third, road type 2 in the fourth, trees in the fifth and road type 3 in the sixth (after which we stopped the recursion since the six main clusters were identified). It is interesting to observe the ordering of the extracted materials: it follows exactly the theoretical results obtained in Theorem 1, i.e., the materials whose spectral signature have larger norm are extracted first (see Fig. 1). As mentioned in the introduction, the recursive approach leads to higher approximation errors: the normalized error $\frac{||M-UV||_F}{||M||_F}$ for $r = 6$ of NMF is 4.4% (using the accelerated HALS algorithm), of NMU is 6.8% and of sparse NMU is 13.2%.

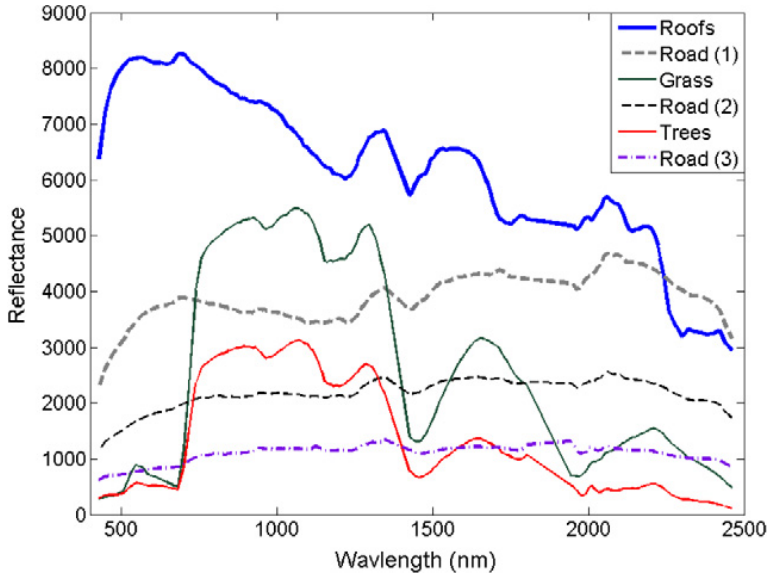

Fig. 5. Spectral unmixing with sparse NMU.

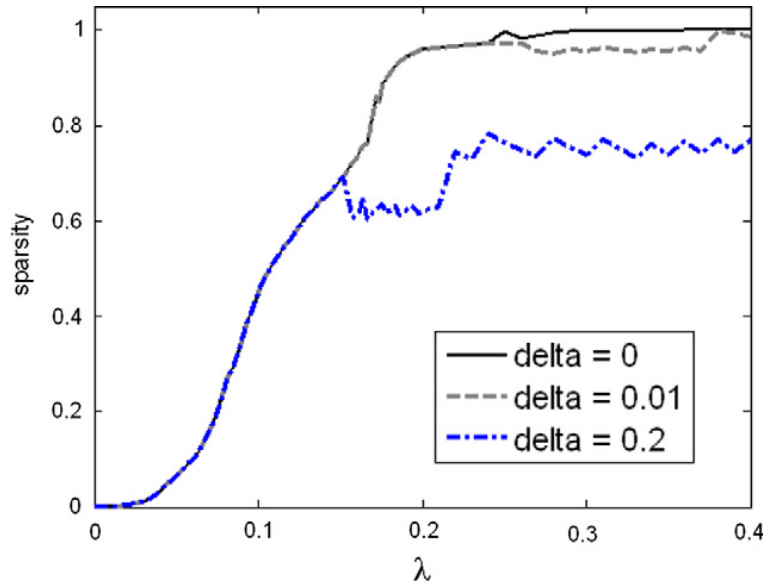

Fig. 6. Sparsity of the first basis element (i.e., $1 - ||u||_0/m \in [0, 1]$) with respect to λ, and for different values of δ ($\Delta = 1$).

Fig. 2 (right) shows the abundance maps obtained with the sparse NMU decomposition. Fig. 5 displays the spectral signatures corresponding to these six extracted materials. We observe that sparse NMU and the nearest neighbor technique perform similarly, and are both able to separate all end-members properly (see also Fig. 1).

5.1.2. Choice and sensitivity of λ

In practice, the parameter λ has to be tuned in order to extract the end-members properly: this requires some (human) supervision (as opposed to standard NMU which is completely unsupervised). This procedure can be rather long and expensive, especially if the dataset is large. In practice, we found that an efficient way for finding proper values for λ is to first run the algorithm on a reduced "training" dataset, considering only a subset of the columns of the input matrix M. For example, for the San Diego image, we kept ten wavelengths equally spaced. In all the tests, we observed that good values for the parameter λ for the reduced dataset also lead to good values for the full dataset.

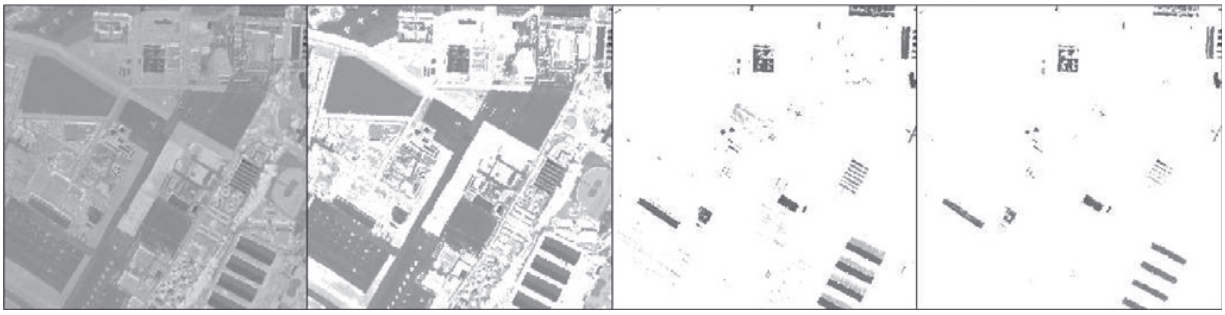

Fig. 7. Abundance maps for the first basis element extracted by sparse NMU with δ=0.01 and for different values of λ (from left to right): 0 (original NMU), 0.1, 0.2 and 0.4.

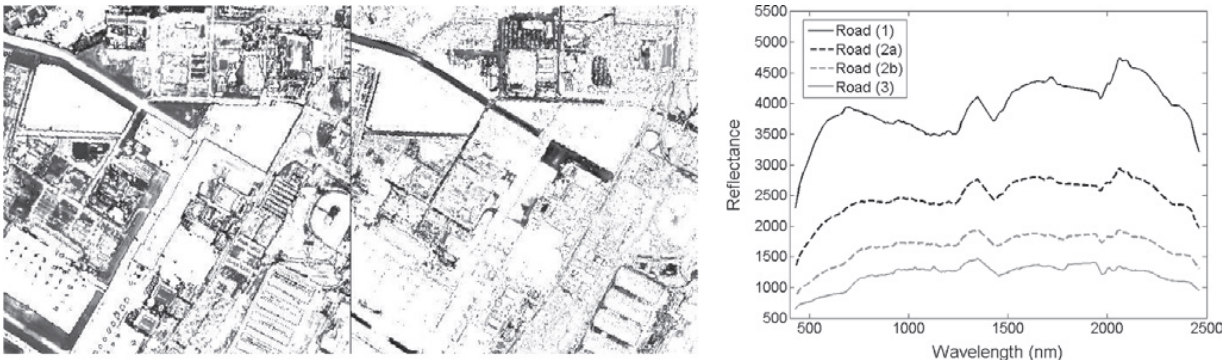

Fig. 8. Larger values of λ_4 (from 0.2 to 0.3) in sparse NMU leads to the splitting of the second type of road surface into two distinct clusters (left) with different spectral signatures (right).

To illustrate the sensitivity of sparse NMU with respect to λ, Fig. 6 displays the evolution of the sparsity of the first basis element extracted by sparse NMU with respect to λ, and for different values of δ. For λ larger than 0.2, the sparsity of the solution is stabilizing. For $\delta = 0$ and $\delta = 0.01$, the solution hence obtained corresponds to the roofs, see Fig. 7; for $\delta = 0.2$, it corresponds to the roofs and the road type 1 because the first basis element is inclined to contain at least 20% of all pixels.

Another interesting observation is that if we increase the parameter λ at the fourth step of the sparse NMU recursion (from 0.2 to 0.3), then the second type of road surface is separated into two different clusters as shown on Fig. 8. We observe that all types of road surface have a very similar spectral signature, the main difference being the intensity of the reflectance.

5.2. Urban dataset

The HYDICE Urban hyperspectral image[6] contains 210 spectral bands, and the data cube has dimension $307 \times 307 \times 210$. The Urban data is mainly composed of 6 types of materials (road, dirt, trees, roofs, grass and metal) as reported in [14] (see Fig. 9) and our implementation of NMU was quite successfully able to detect these six materials, as reported in [10].

In order to make the problem more difficult, we use a subset of only five bands, corresponding to the following wavelengths: 412, 610, 1102, 2003 and 2382nm (i.e., $M \in \mathbb{R}_+^{94249\times 5}$). NMU is no longer able to extract all materials in this situation, see Fig. 10. In fact, it mixes the grass with the trees in the second basis element and the road and dirt in the third. Fig. 11 displays the basis elements obtained with sparse NMU which is able, even with only five bands, to extract each material individually. As mentioned previously, sparse NMU has a larger approximation error: for $r = 8$, the normalized error $\frac{||M-UV||_F}{||M||_F}$ of NMF is zero (since r is larger than n), of NMU is 4.2% and of sparse NMU is 6.16%.

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

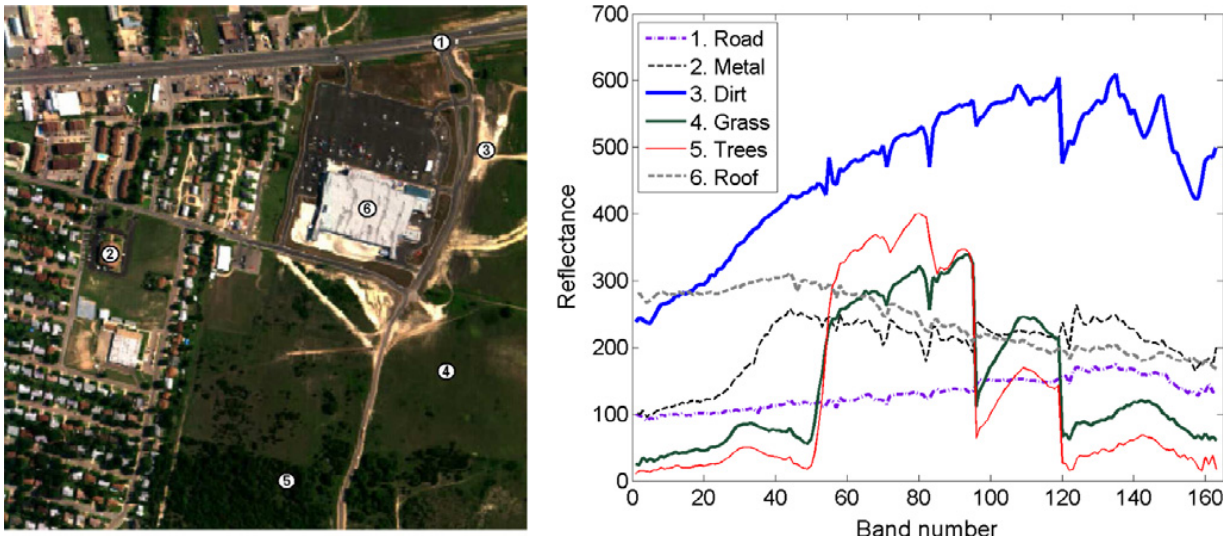

Fig. 9. Urban dataset.

Fig. 10. NMU basis elements obtained for the Urban hyperspectral image with only five bands.

Let us compare sparse NMU with the following standard procedure [7] : the 6 end-members (i.e., the rows of matrix V) were obtained using the N-FINDR5 algorithm [26] plus manual adjustment from [14] (see right image in Fig. 9), and the abundance matrix was computed as the optimal solution of the nonnegative least squares problem $\text{argmin}_{U\geqslant 0} ||M - UV||_F^2$, see the left image in Fig. 12. We observe that sparse NMU, although using only five bands, is able to perform a rather good clustering, comparable to the N-FINDR5 based technique (which used all the 162 clean bands): only the road remains rather difficult to extract properly and is still mixed up with some dirt and roofs.

Finally, we would like to point out that, even though we are using a nonlinear optimization scheme, sparse NMU is quite robust to different initializations (as NMU is [8]). In fact, we have also run sparse NMU with randomly generated initial matrices (instead of the best rank-one approximation of M, cf. step 2 of Algorithm 1) on the Sand Diego Airport and the Urban datasets, and have obtained similar results (which are typically difficult to differentiate visually).

[7] The nearest neighbor technique used for the San Diego image in the previous section was not able to perform a good clustering.

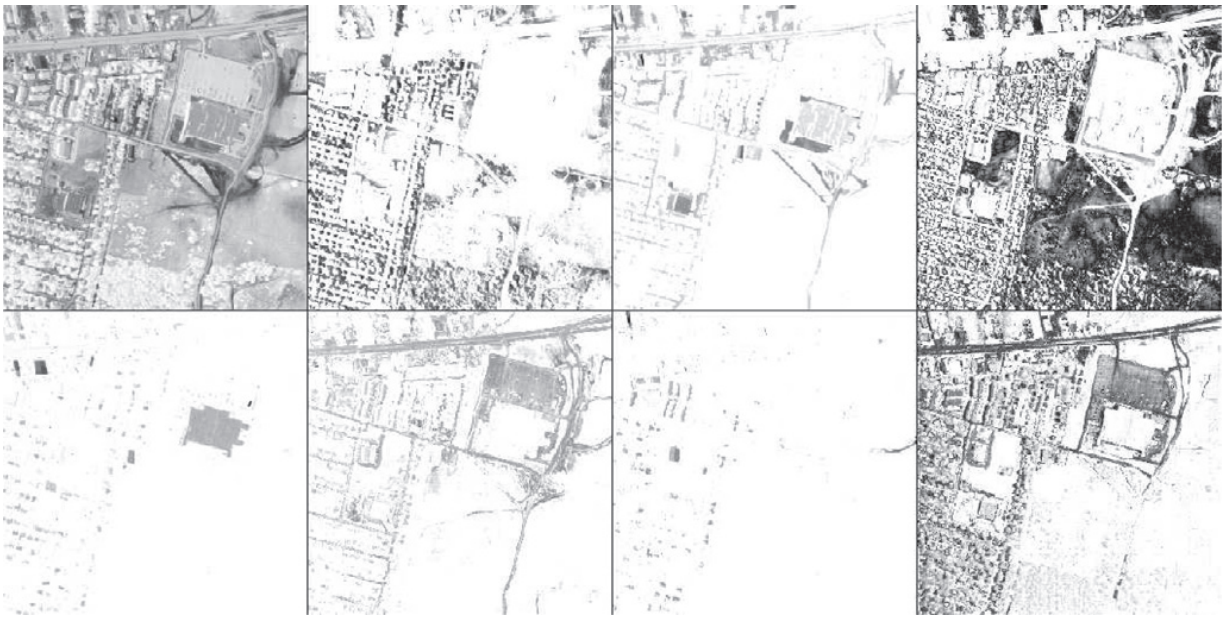

Fig. 11. Sparse NMU basis elements obtained for the Urban hyperspectral image with only five bands, with $\lambda = (0\ 0.25\ 0.2\ 0\ 0.1\ 0.1\ 0.55\ 0)$, $\delta = 0$ and $\Delta = 1$.

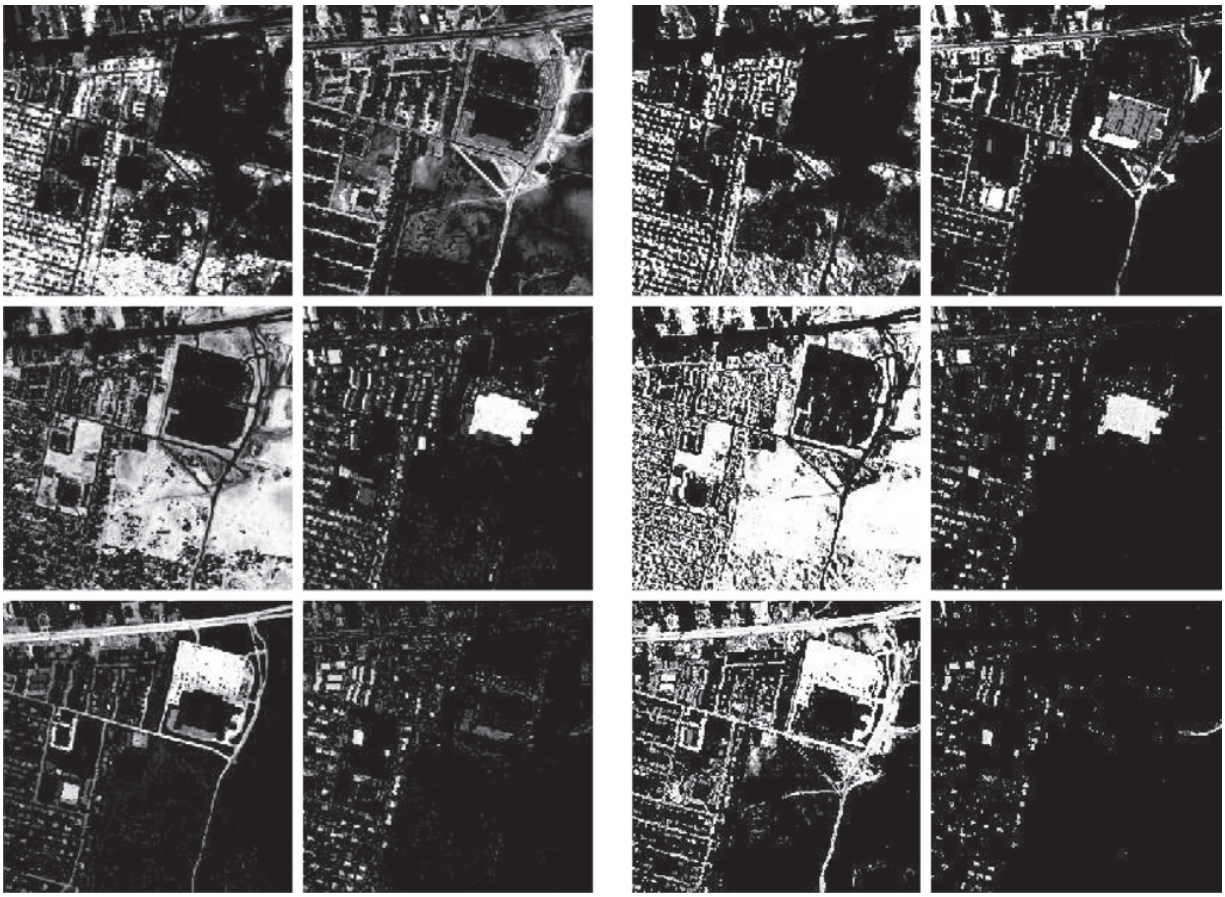

Fig. 12. Abundance maps based on NFINDR5 algorithm with all spectral bands (left), and based on sparse NMU with only five spectral bands (right).

6. Summary and further work

In this paper we have presented a way to incorporate sparsity into the NMU problem, with applications to material identification provided by abundance maps for hyperspectral images. Our approach is based on the use of the standard ℓ_1-norm sparsity inducing term added to the minimization problem. We then experimentally showed that sparse NMU is able to identify materials in situations where standard NMU is not, because sparsity enhances separation of the end-members.

As global convergence and recovery guarantee results are largely unknown for sparse NMU, we hope the promising results in this paper will motivate future theoretical studies. We are also interested in the addition of other penalization terms in order to take other prior information into account, e.g., spatial information or piecewise smoothness of the spectral signatures [15]. In addition, we have in mind testing sparse NMU on difficult data obtained from compressive sensing, using e.g., data collected by the Coded Aperture Snapshot Spectral Imaging (CASSI) system studied in [27].

References

[1] K.M. Anstreicher, L.A. Wolsey, Two well-known properties of subgradient optimization, Math. Program. 120 (1) (2009) 213–220.
[2] J. Bioucas-Dias, A. Plaza, Hyperspectral unmixing: geometrical, statistical and sparse regression-based approaches, in: SPIE Remote Sensing Europe, Image and Signal Processing for Remote Sensing Conference, Toulouse, France, 2010.
[3] E.J. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2006) 489–509.
[4] A. Castrodad, Z. Xing, J.B. Greer, E. Bosch, L. Carin, G. Sapiro, Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 49 (11) (2011) 4263–4281.
[5] A. Cohen, W. Dahmen, R. Devore, Compressed sensing and best *k*-term approximation, J. Amer. Math. Soc 22 (2009) 211–231.
[6] D.L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (2006) 1289–1306.
[7] M. Fong, Z. Hu, Hyperactive: a matlab tool for visualization of hyperspectral images, 2007. Available from: <http://www.math.ucla.edu/∼wittman/lambda/software.html>.
[8] N. Gillis, F. Glineur, Using underapproximations for sparse nonnegative matrix factorization, Pattern Recognit. 43 (4) (2010) 1676–1687.
[9] N. Gillis, F. Glineur, Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization, Neural Comput. 24 (4) (2012) 1085–1105.
[10] N. Gillis, R.J. Plemmons, Dimensionality reduction, classification, and spectral mixture analysis using nonnegative underapproximation, Opt. Eng. 50 (2011) 027001.
[11] N. Gillis, R.J. Plemmons, Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis, in: Third Worskshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, 2011.
[12] G.H. Golub, C.F. Van Loan, Matrix Computation, third ed., The Johns Hopkins University Press, Baltimore, 1996.
[13] Z. Guo, S. Osher, Template matching via l1 minimization and its application to hyperspectral data, Inverse Problems Imaging 5 (1) (2011) 19–35.
[14] Z. Guo, T. Wittman, S. Osher, 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.
[15] S. Jia, Y. Qian, Constrained nonnegative matrix factorization for hyperspectral unmixing, IEEE Trans. Geosci. Remote Sens. 47 (1) (2009) 161–173.
[16] H. Kim, H. Park, Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis, Bioinformatics 23 (12) (2007) 1495–1502.
[17] I. Kopriva, X. Chen, Y. Jao, Nonlinear band expansion and nonnegative matrix underapproximation for unsupervised segmentation of a liver from a multi-phase CT image, SPIE Medical Imaging-Image Processing, Orlando, vol. 7962, 2011.
[18] I. Kopriva, M. Hadzija, M.P. Hadzija, M. Korolija, A. Cichocki, Rational variety mapping for contrast-enhanced nonlinear unsupervised segmentation of multispectral images of unstained specimen, Amer. J. Pathol. 179 (2) (2011) 547–554.
[19] H. Laurberg, M.G. Christensen, M.D. Plumbley, L.K. Hansen, S.H. Jensen, Theorems on positive data: on the uniqueness of NMF, Comput. Intel. Neurosci. (2008), ID 764206.
[20] H. Li, C. Adal, W. Wang, D. Emge, A. Cichocki, Non-negative matrix factorization with orthogonality constraints and its application to raman spectroscopy, J. VLSI Signal Process. 48 (2007) 83–97.
[21] Y.M. Masalmah, M. Vélez-Reyes, A full algorithm to compute the constrained positive matrix factorization and its application in unsupervised unmixing of hyperspectral imagery, in: Proc. SPIE, vol. 6966, 2008, http://dx.doi.org/10.1117/12.779444.
[22] L. Miao, H. Qi, Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization, IEEE Trans. Geosci. Remote Sens. 45 (3) (2007) 765–777.
[23] M.J.D. Powell, On search directions for minimization algorithms, Math. Program. 4 (1973) 193–201.
[24] G.B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (3) (2010) 471–501.
[25] N.Z. Shor, Minimization Methods for Non-differentiable Functions, Springer Series in Computational Mathematics, 1985.
[26] M. Winter, N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data, in: Proc. SPIE Conference on Imaging Spectrometry V, 1999.
[27] Q. Zhang, R. Plemmons, D. Kittle, D. Brady, S. Prasad, Joint segmentation and reconstruction of hyperspectral data with compressed measurements, Appl. Opt. 50 (2011) 4417–4435.