SPARSE NONNEGATIVE MATRIX UNDERAPPROXIMATION AND ITS APPLICATION TO HYPERSPECTRAL IMAGE ANALYSIS

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

[a] CORE and ICTEAM, Université catholique de Louvain, B-1348 Louvain-La-Neuve, Belgium
[b] Computer Science and Mathematics, Wake Forest University, Winston-Salem, NC 27106

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. However, for difficult hyperspectral datasets, even NMU can mix some materials together, and is therefore not able to separate of all materials properly for accurate target identification. 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 a HYDICE image of the San Diego airport.

Index Terms— Hyperspectral images, nonnegative matrix factorization, underapproximation, sparsity, dimensionality reduction, classification, spectral mixture analysis.

1. INTRODUCTION

Hyperspectral images (HSI) 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 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 [1] for a comprehensive overview of recent HSI analysis for target identification trends.

A standard linear model for HSI 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 row m^i of M corresponds to the spectral signature of a pixel (and therefore, each column of M corresponds to an image at a given wavelength); then model reads

$$m_i \approx \sum_k u_{ik} \, v_k^T \quad \forall i,$$

i.e., the spectral signature of each pixel (m_i, a row of M) is approximated by a nonnegative linear combination (with weights $u_{ik} \geq 0$, representing abundances) of end-members' signatures ($v_k \geq 0$), approximating the true signatures of the constituent materials of the HSI. This corresponds exactly to the nonnegative matrix factorization (NMF) model $\min_{U \geq 0, V \geq 0} ||M - UV||_F^2$. However, NMF is typically not able to separate all end-members correctly because of the *non-uniqueness* of the solution [2]. In order to improve NMF performance, one should incorporate prior information into the model, and take into account the characteristics of the solutions (such as sparsity and smoothness) to make the problem more well-posed, see [1, 3] and the references therein.

Another approach introduced recently in [4] is based on the introduction of additional underapproximation constraints, enabling one to extract features in a recursive way, like PCA, but preserving nonnegativity. At each step of the recursion, the following problem, called nonnegative matrix underapproximation (NMU),

$$\min_{u \geq 0, v \geq 0} ||M - uv^T||_F^2 \text{ such that } uv^T \leq M, \qquad (1)$$

is solved, and a nonnegative residual matrix $R = M - uv^T \geq 0$ is obtained, so that the same procedure can be applied on R.

The first author is a Research Fellow of the Fonds de la Recherche Scientifique (F.R.S.-FNRS). This research is supported in part by the U.S. Air Force Office of Scientific Research (AFOSR), grant number FA9550-08-1-0151. Email: Nicolas.Gillis@uclouvain.be, plemmons@wfu.edu

978-1-4577-2201-1/11/$26.00

It has been explained and experimentally shown why NMU performs much better than standard NMF for the analysis of HSI [5]. 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.

In this paper, we propose to incorporate sparsity into the NMU model, leading to improved performance for HSI 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 (sNMU), and propose an algorithm to find approximate solutions. In Section 3, we apply this new technique to the analysis of a HYDICE image of the San Diego Airport, and show qualitatively that it is able to separate end-members better than NMU.

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_i is the i^{th} row of A, and a_{ij} the entry at position (i,j); for $b \in \mathbb{R}^m$, we denote b_i as the i^{th} 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 Frobenius norm of matrix A is $||A||_F^2 = \sum_{i,j} a_{ij}^2$.

2. SPARSE NMU

A possible way to enforce separation of the end-members is to add sparsity constraints on the factor U in NMF [3]. In fact, each end-member 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) is not sparse enough. In order to enforce sparsity, a ℓ_0-pseudo-norm penalty term can be incorporated into the NMU model. We then would like to solve at each step of the recursion the following problem:

$$\min_{u\geq 0, v\geq 0} ||M - uv^T||_F^2 + \mu||u||_0 \text{ such that } uv^T \leq M, \quad (2)$$

for some regularization parameter $\mu \geq 0$. We now present an algorithm to find approximate solutions to (2).

Following the work of [4] (see also [5]), we introduce Lagrangian dual variables $\Lambda \in \mathbb{R}^{m\times n}_+$ associated with the underapproximation constraints and write the Lagrangian dual (LD): $\sup_{\Lambda\geq 0} \; L(\Lambda)$, where $L(\Lambda)$ is the Lagrangian dual function of (2):

$$L(\Lambda) = \min_{u\geq 0, v\geq 0} ||M - uv^T||_F^2 + \mu||u||_0 - 2\left\langle \Lambda, M - uv^T \right\rangle. \quad (3)$$

The LD is a non-smooth convex optimization problem (see, e.g., [6]) and can be solved using the following subgradient scheme [7]: initialize $\Lambda^{(0)}$, then for $k = 0, 1, 2, \ldots$ do

1. Find a solution (u, v) of (3) for $\Lambda = \Lambda^{(k)}$.
2. Update $\Lambda^{(k+1)} \leftarrow \max\left(0, \Lambda^{(k)} + \alpha_k(uv^T - M)\right)$.

If the Lagrangian relaxation subproblems (3) are solved exactly at each step and for example $\alpha_k = 1/k$, then the convergence of the above scheme is guaranteed [6].

2.1. Solving the Lagrangian Relaxation (3)

It has been shown that (3) is NP-hard for $\mu = 0$, see [4]; while when $\mu > 0$ the ℓ_0-penalty term is very difficult to work with since it is of a combinatorial nature. Therefore, in order to find an approximate solution to (3) we use the standard ℓ_1-norm heuristic with regularization parameter μ (see, e.g., [8] and the references therein), and solve instead[1]

$$\min_{u\geq 0, v\geq 0, ||v||_2=1} ||(M - \Lambda) - uv^T||_F^2 + \mu||u||_1. \quad (4)$$

Then we apply an exact two-block coordinate descent scheme to (4), which guarantees the convergence to a stationary point of (4), see [9]. This amounts to successively fixing u and computing the corresponding optimal solution for v, and vice versa: initialize (u, v), then for $k = 0, 1, 2, \ldots$ do

1. $v \leftarrow \text{argmin}_{v\geq 0, ||v||_2=1} ||(M - \Lambda) - uv^T||_F^2$.
2. $u \leftarrow \text{argmin}_{u\geq 0} ||(M - \Lambda) - uv^T||_F^2 + \mu||u||_1$.

One advantage of this approach is that the optimal solutions of these subproblems can be written in closed form, cf. steps 6 and 8 of Alg. 1 to follow.

However, the penalty term $||u||_1$ in (4) can force u to have smaller entries than desired. Therefore the quantity $||M-\Lambda-uv^T||_F$ can typically be reduced by multiplying uv^T with a constant factor $\alpha \geq 1$. Since the sparsity of u is unchanged by this procedure, we obtain a better solution for the original problem (3). The optimal value of this factor can be computed easily (see step 10 of Alg. 1). Finally, we observed that updating only u and v once between each update of Λ seems to give good results. This is implemented in Alg. 1.

2.2. Choice of Regularization Parameter μ

At each step of the recursion, Alg. 1 consists of initializing

1. the variables (u, v) with an optimal nonnegative rank-one approximation (u^*, v^*) of $M \geq 0$, with $||v^*||_2$=1;
2. the Lagrangian dual variables Λ with the nonnegative part of the residual $\max\left(0, -(M - u^*v^{*T})\right)$.

[1] Notice that we have also added the normalizing constraint $||v||_2 = 1$, otherwise solution u of (4) would go to zero while v would go to infinity.

Algorithm 1 Sparse NMU

Require: $M \in \mathbb{R}_+^{m \times n}$, $r > 0$, $\lambda \in [0,1)$, $\delta \in [0,1)$, maxiter.

1: **for** $k = 1 : \mathrm{r}$ **do**
2: $[x, y] =$ optimal rank-one approximation(M);
3: $u_k \leftarrow x$; $v_k \leftarrow y$; $\Lambda \leftarrow \max(0, -(M - xy^T))$;
4: $y \leftarrow y/||y||_2$; $\mu = \lambda||(M - \Lambda)y||_\infty$;
5: **for** $p = 1 :$ maxiter **do**
6: $x \leftarrow \max\left(0, (M - \Lambda)y - \mu\right)$;
7: **if** $||x||_0 \leq \max(1, \delta m)$, **then** $\mu \leftarrow 0.95\mu$; **end if**
8: $y \leftarrow \max\left(0, (M - \Lambda)^T x\right)$; $y \leftarrow y/||y||_2$;
9: **if** $x \neq 0$ and $y \neq 0$ **then**
10: $\alpha = \frac{x^T(M-\Lambda)y}{||x||_2^2||y||_2^2}$; $u_k \leftarrow \alpha x$; $v_k \leftarrow y$;
11: $\Lambda \leftarrow \max(0, \Lambda - \frac{1}{p+1}(M - u_k v_k^T))$;
12: **else**
13: $\Lambda \leftarrow \frac{\Lambda}{2}$; $y \leftarrow v_k$;
14: **end if**
15: **end for**
16: $M \leftarrow \max(0, M - u_k v_k^T)$;
17: **end for**

Hence the first update of u will be given by $\max\left(0, (M - \Lambda)v^* - \mu\right)$ (see step 6 of Alg. 1). Therefore parameter μ must be chosen smaller than $||(M - \Lambda)v^*||_\infty$ otherwise u is set to zero. Introducing parameter λ such that $\mu = \lambda||(M - \Lambda)v^*||_\infty$, we then must have $\lambda \in [0, 1)$.

2.3. Lower Bound on the Sparsity of u

In case our parameter λ is chosen too large, we might obtain either a trivial solution (u=0), or a solution which is too sparse, and hence containing only a very small subset of the pixels. In 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 relatively large clusters, a lower bound on the sparsity of u can be imposed. We define a new parameter $\delta \in [0, 1)$ as this lower bound in percent of the total size of the image. If this lower bound is reached for some iterate u, the value of parameter λ is decreased, see step 7 of Alg. 1. For λ=0 and δ=0, Alg. 1 reduces to the algorithm from [5]. Moreover, it has *exactly* the same computational cost, with $O(Kmnr)$ operations where K=maxiter.

3. SAN DIEGO AIRPORT HYPERSPECTRAL IMAGE

The San Diego airport HYDICE image consists of 158 clean bands with 400×400 pixels. It contains the four following main types of surfaces: roofs, grass, road type 1 (including boarding zones) and road type 2 (including roads and parking lots). NMF and NMU are not able to extract the materials properly [5]: they extract materials together and can therefore not identify them individually, essentially because they do not incorporate enough prior constraints, such as sparsity.

3.1. Spectral Unmixing

Fig. 1 displays the abundance maps for the basis elements obtained for λ=0.2, δ=0.01, and maxiter=100, for which sparse NMU (sNMU) extracts the different clusters[2]: roofs in the first basis element, road type 1 in the second, grass in the third, and road type 2 in the fourth (after which we stopped the recursion since the four main clusters were identified). Fig. 2 displays the four clusters obtained with a nearest neighbor technique manually initialized. Fig. 3 displays the spectral signatures corresponding to these four extracted materials (following the same procedure as in [5]), with a comparison with nearest neighbor. Both methods give similar results, while sNMU does not need manual adjustment.

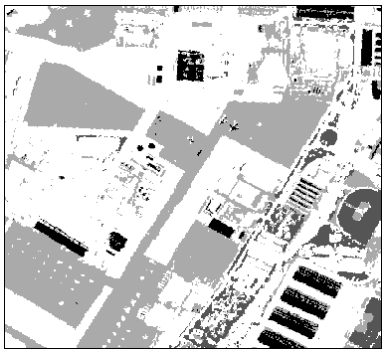

Fig. 2. Clustering with manually adjusted nearest neighbor.

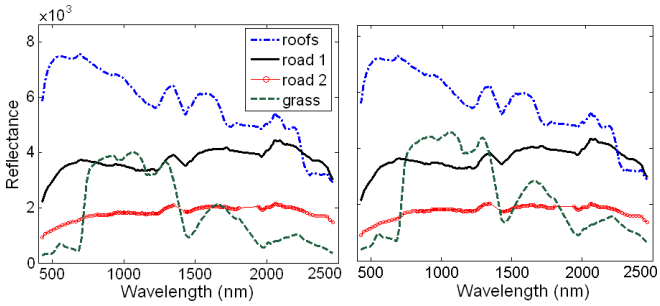

Fig. 3. Spectral unmixing: (left) sNMU (λ=0.2, δ=0.01), (right) manually adjusted nearest neighbor.

3.2. Sensitivity to λ

Fig. 4 displays the evolution of the sparsity of first basis element extracted by sNMU with respect to λ, and for different values of δ. For λ larger than 0.2, the sparsity of the solution is stabilizing. For δ=0 and δ=0.01, the solution hence obtained

[2] As expected, sNMU has a larger approximation error: the normalized error $\frac{||M-UV||_F}{||M||_F}$ of NMF is 5.6%, NMU 7.2% and sNMU 18.0% (for a fair comparison, since $UV \leq M$, V=argmin$_{V \geq 0} ||M - UV||_F^2$ [4]).

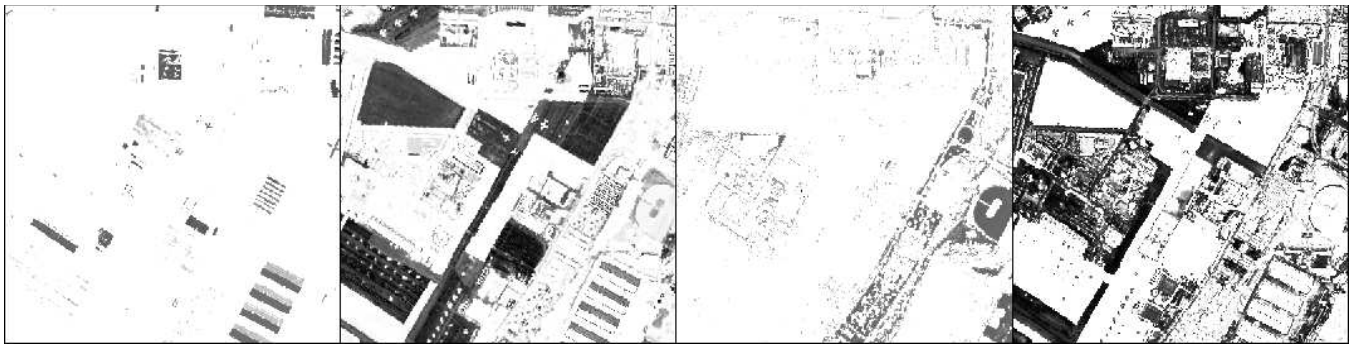

Fig. 1. First four basis elements of sNMU with λ=0.2 and δ=0.01. Left to right: roofs, roads type 1, grass, roads type 2.

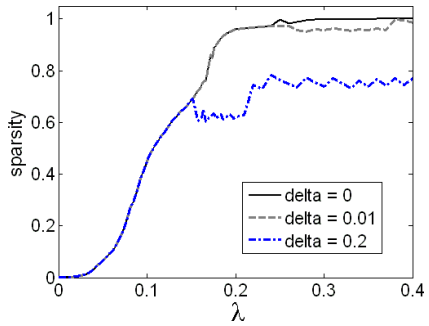

Fig. 4. Sparsity of the first basis element (i.e., $1 - ||u||_0/m \in [0, 1]$) with respect to λ, and for different values of δ.

corresponds to the roofs, see Fig. 5; for δ=0.2, it corresponds to the roofs and the road type 1 since the first basis element is inclined to contain at least 20% of all pixels.

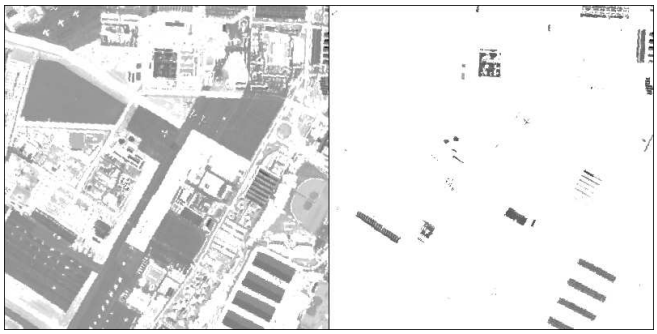

Fig. 5. Abundance maps for the first basis element extracted by sNMU with δ=0.01: (left) λ=0.1 and (right) λ=0.3.

4. SUMMARY AND FURTHER WORK

In this paper, we have presented a way to enhance sparsity of the solutions to the NMU problem, with applications to target identification. 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 sNMU is able to identify objects in a HYDICE image (while standard NMU and NMF are not), because sparsity enhances separation of the end-members. In further work we will test sNMU on additional HSI and we expect to develop a theoretical argument to explain why sNMU works better than standard NMU. We will also compare the results with other approaches such as complexity constrained NMF [3], and dictionary modeling methods with constrained NMF initialization [1].

5. REFERENCES

[1] A. Castrodad, Z. Xing, J. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning discriminative sparse models for source separation and mapping of hyperspectral imagery," IMA Preprint Series No. 2341, University of Minnesota, 2010.

[2] H. Laurberg, M.G. Christensen, M.D. Plumbley, L.K. Hansen, and S.H. Jensen, "Theorems on positive data: On the uniqueness of NMF," *Comput. Intel. and Neurosci.*, vol. 2008, ID 764206.

[3] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47(1), pp. 161–173, 2009.

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

[5] N. Gillis and R.J. Plemmons, "Dimensionality reduction, classification, and spectral mixture analysis using nonnegative underapproximation," *Optical Engineering*, vol. 50, 027001, 2011.

[6] K.M. Anstreicher and L.A. Wolsey, "Two "well-known" properties of subgradient optimization," *Mathematical Programming*, vol. 120(1), pp. 213–220, 2009.

[7] N.Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer Series in Computational Mathematics, 1985.

[8] H. Kim and H. Park, "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis," *Bioinformatics*, vol. 23(12), pp. 1495–1502, 2007.

[9] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," *Operations Research Letters*, vol. 26, pp. 127–136, 2000.