Priors in Sparse Recursive Decompositions of Hyperspectral Images

Nicolas Gillis[a], Robert J. Plemmons[b] and Qiang Zhang[c]

[a]University of Waterloo, Department of Combinatorics and Optimization, Waterloo, Ontario N2L 3G1, Canada;
[b]Departments of Mathematics and Computer Science, Wake Forest University, Winston-Salem, NC 27106.
[c]Department of Biostatistical Sciences, Wake Forest University Health Sciences, Winston-Salem, NC 27106.

ABSTRACT

Nonnegative matrix factorization and its variants are powerful techniques for the analysis of hyperspectral images (HSI). Nonnegative matrix underapproximation (NMU) is a recent closely related model that uses additional underapproximation constraints enabling the extraction of features (e.g., abundance maps in HSI) in a recursive way while preserving nonnegativity. We propose to further improve NMU by using the spatial information: we incorporate into the model the fact that neighboring pixels are likely to contain the same materials. This approach thus incorporates structural and textural information from neighboring pixels. We use an ℓ_1-norm penalty term more suitable to preserving sharp changes, and solve the corresponding optimization problem using iteratively reweighted least squares. The effectiveness of the approach is illustrated with analysis of the real-world cuprite dataset.

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

1. INTRODUCTION

A hyperspectral image (HSI) is a three-dimensional tensor datacube, providing the electromagnetic reflectance of a scene at varying wavelengths. HSI are digital images 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 hundreds of wavelengths. The spatial information in the scene is generally a two dimensional grayscale image, while the wavelength parameter is recorded in the third dimension. HSI analysis is widely used in remote sensing, and applications also exist in areas such as medical imaging and environmental studies. The spectral vector at every spatial location can be (approximately) linearly expressed as combination of a set of common vectors, called endmembers or spectral signatures of materials in the scene. A hyperspectral tensor datacube is a set of nonnegative mixtures of endmembers, which are also nonnegative. A primary objective in analyzing HSI is to recover the endmembers (unmixing) and mixture coefficients (abundance maps) from a hyperspectral datacube using matrix factorization methods. This can enable identification of, for example, objects, species of trees and vegetation, crop health, mineral and soil composition, moisture content, and pollution quantities. Three-dimensional tensor factorization methods can also be used,[1,2] but we concentrate here on matrix models.

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 materials present in this pixel (called endmembers). More precisely, assume we construct a matrix M such that each entry m_{ij} of M is equal to the reflectance* of the i^{th} pixel at the j^{th} wavelength (i.e.,

E-mail addresses: ngillis@uwaterloo.ca (corresponding author), plemmons@wfu.edu and qizhang@wakehealth.edu.
Research by Robert Plemmons and Qiang Zhang supported in part by the U.S. Air Force Office of Scientific Research (AFOSR), with award numbers FA9550-08-1-0151 and FA9550-11-1-0194, and by the U.S. National Geospatial-Intelligence Agency under Contract HM1582-10-C-0011. The scientific responsibility is assumed by the authors.

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

Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, edited by Sylvia S. Shen, Paul E. Lewis, Proc. of SPIE Vol. 8390, 83901M · © 2012 SPIE
CCC code: 0277-786X/12/$18 · doi: 10.1117/12.918333

each row m_i^T of M corresponds to the spectral signature of a pixel and each column m_j 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 $U_{ik} \geq 0$, representing abundances) of endmembers' signatures ($V_{k:} \geq 0$) 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 $r \leq \min(m, n)$, solve

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

However, NMF is typically not able to separate all end-members correctly because of the *non-uniqueness* of the solution.[3] 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,[4] orthogonality,[5] minimum-volume,[6] and sum-to-one constraints on the abundances.[7]

Nonnegative matrix underapproximation (NMU)[8] is a recent closely related model that uses additional underapproximation constraints enabling the recursive extraction of abundance maps in HSI. At the first step, NMU requires solving the following optimization problem

$$\min_{x \geq 0, y \geq 0} \|M - xy^T\|_F^2 \quad \text{such that } xy^T \leq M, \tag{1.2}$$

where $M \in \mathbb{R}^{m \times n}$ is the given nonnegative data matrix. Then, the residual matrix $R = M - xy^T \geq 0$ is computed and the same procedure is applied on R. After r steps, we obtain an approximate NMF solution of rank r. NMU has several advantages over NMF:

1. It is well-posed (the solution is, under some mild assumption, unique[9]).
2. The factorization rank does not have to be chosen a priori.
3. It produces sparser solution leading to a better decomposition into parts.[8]

Moreover, it was theoretically and experimentally shown to be able to extract automatically constitutive materials in HSI.[9–11] However, sometimes, NMU fails at extracting all endmembers and mixes some of them. A possible way to improve NMU performances is to add prior information into the model. It was shown[12] how adding sparsity constraint on the abundance matrix makes it able to extract endmembers in a more efficient way.

In this paper, we propose to add spatial information into NMU, i.e., we incorporate in the model the fact that neighbor pixels are likely to contain the same materials, and can thus incorporate structural and textural information from neighboring pixels. We refer to this method as spatial NMU, and apply the technique on the cuprite dataset and show that it performs better than NMU, and competes favorably with the popular Vertex Component Analysis (VCA) method.[13]

2. SPATIAL NONNEGATIVE MATRIX UNDERAPPROXIMATION

The objective of this section is to design an algorithm to solve NMU problem (1.2) while taking the local information into account. The typical approach is to add a regularization term into the objective function. We propose here to use the following

$$\sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} |x_i - x_j|,$$

where $\mathcal{N}(i)$ is the set of neighboring pixels of pixel i. The ℓ_1-norm is suitable for image analysis because it is able to preserve the edges (using the ℓ_2-norm would smooth them out), see[14,15] and the references therein where the same penalty term is used for HSI.

Let us construct the 'neighbor' matrix N as follows. Each pair (i,j) of neighboring pixels with $1 \leq i < j \leq m$, $i \in \mathcal{N}(j)$ and $j \in \mathcal{N}(i)$ corresponds to a row (say at position k) of matrix $N \in \mathbb{R}^{K \times m}$ with

$$N(k,i) = 1 \text{ and } N(k,j) = -1, \tag{2.1}$$

so that

$$\sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} |x_i - x_j| = 2\|Nx\|_1.$$

Notice that K, the number of neighboring pairs, is proportional to the number of pixels in the image. In this paper, we define the neighboring pixels as the adjacent pixels (hence each pixel has at most four neighbors and $K \leq 4m$).

2.1 Lagrangian Dual

In Ref.,[8] approximate solutions of NMU (1.2) are obtained by solving the Lagrangian dual

$$\max_{\Lambda \geq 0} \min_{x \geq 0, y \geq 0} L(x,y,\Lambda) = \|M - xy^T\|_F^2 + 2\sum_{i,j}(xy^T - M)_{ij}\Lambda_{ij}, \tag{2.2}$$

where $\Lambda \in \mathbb{R}_+^{m \times n}$ is the matrix containing the Lagrangian multipliers. This is achieved by applying the following alternating scheme

1. $x \leftarrow \operatorname{argmin}_{\bar{x} \geq 0} L(\bar{x}, y, \Lambda) = \max\left(0, \frac{(M-\Lambda)y}{y^T y}\right)$.
2. $y \leftarrow \operatorname{argmin}_{\bar{y} \geq 0} L(x, \bar{y}, \Lambda) = \max\left(0, \frac{(M-\Lambda)^T x}{x^T x}\right)$.
3. $\Lambda \leftarrow \max(0, \Lambda + \alpha_k(xy^T - M))$, where α_k is a square summable but not summable sequence (e.g., $\alpha_k = \frac{1}{k}$), as used in online algorithms.[16] This is projected gradient step since $\nabla_\Lambda L(\bar{x}, y, \Lambda) = 2(xy^T - M)$.

In order to increase the spatial coherence, we are going to modify the update of x to take into account the regularization term $\|Nx\|_1$.

Equivalent Lagrangian Relaxation

For a fixed Λ, the problem, $\min_{x \geq 0, y \geq 0} L(x,y,\Lambda)$, is called a Lagrangian relaxation of (1.2). Let us introduce a variable $\sigma \geq 0$ and fix the norms of x and y to one to obtain the following equivalent problem

$$\min_{\sigma \geq 0, x \geq 0, y \geq 0} -2\sigma x^T(M - \Lambda)y + \sigma^2 \underbrace{\|xy^T\|_F^2}_{=1}, \text{ such that } \|x\|_2 = \|y\|_2 = 1. \tag{2.3}$$

For any fixed (x,y), the minimum of (2.3) is achieved at $\sigma^*(x,y) = \max(0, x^T(M-\Lambda)y)$. Assuming $x^T(M-\Lambda)y \geq 0$ (otherwise σ is equal to zero and we have a trivial solution), (2.3) is equivalent to

$$\max_{x \geq 0, y \geq 0} x^T(M - \Lambda)y, \text{ such that } \|x\|_2 = \|y\|_2 = 1. \tag{2.4}$$

In order to improve the spatial coherence in x, we add the regularization term described in the previous section and, for a fixed y, our goal is to solve

$$\max_{x \geq 0, \|x\|_2 = 1} x^T(M - \Lambda)y - \frac{\mu}{2}\|Nx\|_1.$$

In the following, we use iteratively re-weighted least squares and a simple gradient scheme to find an approximate solution to this subproblem.

2.2 Iteratively Re-Weighted Least Squares (IRWLS)

Consider problems of the type

$$\max_{x\geq 0, \|x\|_2=1} \; x^T A y - \frac{\mu}{2}\|Nx\|_1, \tag{2.5}$$

and assume for now that the optimal objective function value is positive (see Section 2.3 for a discussion). Then we can replace the constraint $\|x\|_2 = 1$ with $\|x\|_2 \leq 1$ (in fact, the constraint will be active at optimality since multiplying x by a constant factor multiplies the objective function by the same factor). Unfortunately, the term $\|Nx\|_1$ is non-differentiable and computationally difficult to work with. In this paper, we propose to use iteratively re-weighted least squares (IRWLS):[17] assuming the vector z is positive, we have that

$$\|z\|_1 = \sum_i |z_i| = \sum_i \frac{z_i^2}{|z_i|} = \|w \circ z\|_2^2,$$

where $w_i = \frac{1}{\sqrt{|z_i|}}$, $\forall i$, and '$\circ$' is the component-wise product. IRWLS fixes the weight w based on the value of z at the previous iteration. Denote $z^{(k)}$ as z at the k^{th} iteration. Then $\|z\|_1$ is approximated at the $(k+1)^{\text{th}}$ iteration with $\|z\|^2_{w^{(k)}}$, where $w_i^{(k)} = (|z_i^{(k)}| + \epsilon)^{-0.5}$ for a small positive constant ϵ. In (2.5), $\|Nx\|_1$ is then replaced at the $(k+1)^{\text{th}}$ iteration with

$$\|Nx\|_1 \approx \|w^{(k)} \circ (Nx)\|_2^2 = \|W^{(k)} Nx\|_2^2 = x^T ({N^T {W^{(k)}}^T W^{(k)} N}) x,$$

where $w_i^{(k)} = (|Nx^{(k)}|_i + \epsilon)^{-1/2}$ $\forall i$, $x^{(k)}$ is the k^{th} iterate, and $W^{(k)} = \text{diag}(w^{(k)})$ (i.e., $W^{(k)}$ is a diagonal matrix whose diagonal entries are given by the entries of $w^{(k)}$). Hence the subproblem (2.5) is now a convex quadratic program:

$$\max_{x\geq 0, \|x\|_2\leq 1} \quad f(x) = x^T A y - \frac{\mu}{2}\, x^T B x, \tag{2.6}$$

where $B = N^T {W^{(k)}}^T W^{(k)} N \succeq 0$ at the $(k+1)^{\text{th}}$ iteration. For this preliminary work, we propose to solve problem (2.6) with a simple projected gradient scheme[18] which guarantees the objective function to decrease:

$$x \leftarrow \mathcal{P}\left(x + \frac{1}{L}\nabla f(x)\right), \quad \nabla f(x) = Ay - \mu B x,$$

where L is the Lipschitz constant of $\nabla f(x)$ (which is equal to the largest eigenvalue of B), and $\mathcal{P}$ is the projection onto the set $\{x \in \mathbb{R}^m \mid x \geq 0, \|x\|_2 \leq 1\}$, given by

$$\mathcal{P}(z) = \begin{cases} \frac{\max(z,0)}{\|\max(z,0)\|_2} & \text{if } \|\max(z,0)\|_2 > 1, \\ \max(z,0) & \text{otherwise.} \end{cases}$$

Since B is a very large (but sparse) K-by-K matrix, it is computational costly to compute exactly its largest eigenvalue. Instead, since B is sparse, we propose to use several steps of the power method (see step 12 of Algorithm 1). Notice that the power method underestimates the Lipschitz constant, which allows the algorithm to initially take larger steps. In the worst-case, this scheme requires $\mathcal{O}(\frac{L}{\epsilon})$ iterations, where ϵ is the desired accuracy for the objective function. In further work, we plan to use a fast gradient method requiring only $\mathcal{O}\left(\sqrt{\frac{L}{\epsilon}}\right)$ iterations.[18]

REMARK 1. *We have also implemented a subgradient method to solve* (2.5) *similarly as in Ref.,*[14] *but convergence was much slower and the step length strategy is much more difficult to handle.*

REMARK 2. *Because the ℓ_1-norm penalty term $\|Nx\|_1$ is a regularization term and since the ultimate goal is to increase the spatial coherence in the solution, the fact that $\|\|Nx\|_1$ is somehow loosely approximated is not crucial, while it makes the subproblems much more easily solvable.*

Finally, Algorithm 1 gives the pseudocode for the proposed alternating scheme for increasing the spatial coherence in NMU solutions: Λ and y are updated similarly as in the original NMU algorithm while x is updated as described above by taking into account the ℓ_1-norm penalty term. We initialize (x, y, Λ) with an approximate solution of (2.2) using the algorithm from Ref.[9] The code is available at `https://sites.google.com/site/nicolasgillis/`.

Algorithm 1 NMU incorporating spatial information

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

Generate the matrix N; % see Equation (2.1)
for $k = 1 :$ r **do**
  $z = rand()$; % Estimate of the eigenvector of B associated with the largest eigenvalue
  % Intialization of (x, y) with an approximate solution to NMU (1.2), see Ref.[9]
  $[x, y, \Lambda]$ = rank-one underapproximation(M);
  $u_k \leftarrow x$; $v_k \leftarrow y$; $x \leftarrow \frac{x}{\|x\|_2}$; $y \leftarrow \frac{y}{\|y\|_2}$;
  $w_i = (|Nx|_i + \epsilon)^{-0.5}$; $W = \mathrm{diag}(w)$; % Initialization of IRWLS weights
  for $p = 1 :$ maxiter **do**
    $A = M - \Lambda$;
    % Update of x
    $B = (WN)^T(WN)$;
    for $l = 1 :$ iter, $z = Bz$; $z = \frac{z}{\|z\|_2}$; **end** % Power method
    $L = z^T Bz$; % Approximated Lipschitz constant
    for $l = 1 :$ iter **do**
      $\mu = \lambda \frac{\|Ay\|_\infty}{\|Bx\|_\infty}$;
      $\nabla f(x) = Ay - \mu Bx$;
      $x \leftarrow \mathcal{P}\left(x + \frac{1}{L}\nabla f(x)\right)$;
    end for
    % Update of y
    $y \leftarrow \max\left(0, (M - \Lambda)^T x\right)$; **if** $y \neq 0$, $y \leftarrow \frac{y}{\|y\|_2}$; **endif**
    % Update of Λ and save (x, y)
    if $x \neq 0$ and $y \neq 0$ **then**
      $\sigma = x^T Ay$; $u_k \leftarrow x$; $v_k \leftarrow \sigma y$;
      $\Lambda \leftarrow \max\left(0, \Lambda - \frac{1}{p}\left(M - u_k v_k^T\right)\right)$;
      % Update the IRWLS weights
      $w_i = (|Nx|_i + \epsilon)^{-0.5}$; $W = \mathrm{diag}(w)$;
    else
      $\Lambda \leftarrow \frac{\Lambda}{2}$; $x \leftarrow \frac{u_k}{\|u_k\|_2}$; $y \leftarrow \frac{v_k}{\|v_k\|_2}$;
    end if
  end for
  $M = \max(0, M - u_k v_k^T)$;
end for

2.3 Heuristic for the Choice of the Penalty Parameter μ

In the previous section, we assumed that the optimal objective function value of problem (2.5) was nonnegative so that the solution x is not forced to zero (which would lead to a trivial solution). If the current iterate $x^{(k)}$ has norm one, then we ideally would like the next iterate $\mathcal{P}(x^{(k)} + \frac{1}{L}\nabla f(x^{(k)}))$ to have norm one as well, i.e., the gradient direction does not point towards zero. A necessary and sufficient (but rather strong) condition is to require ${x^{(k)}}^T \nabla f(x^{(k)}) \geq 0$ (in fact, the gradient of the constraints $\sum_i x_i^2 - 1 \leq 0$ is $2x$) from which we would have to take μ such that $\mu {x^{(k)}}^T B x^{(k)} \leq {x^{(k)}}^T A y^{(k)}$. Based on this observation, we have tried different heuristics and it seems that the following choice works best:

$$\mu = \lambda \frac{\|Ay^{(k)}\|_\infty}{\|Bx^{(k)}\|_\infty}, \quad \text{for some } \lambda \in [0, 1].$$

In contrast, we observed that fixing the value of μ a priori is rather difficult and requires one to take different values at each step of the NMU recursion. A value of λ between 0.1 and 0.5 seems to work well in practice (see Section 3). A more detailed analysis and a better choice for the parameter μ are topics for further research.

3. NUMERICAL EXPERIMENTS - CUPRITE DATASET

We now apply NMU and spatial NMU on the cuprite dataset[†]. Cuprite is a mining area in southern Nevada with mostly mineral and very little vegetation, located approximately 200km northwest of Las Vegas, see, e.g.,[13] for more information and `http://speclab.cr.usgs.gov/PAPERS.imspec.evol/aviris.evolution.html`. It consists of 188 images, each having 250×191 pixels, and is composed of about 20 different materials (minerals). We use the following parameters for Algorithm 1: maxiter = 100, $\lambda = 0.1$, $\epsilon = 10^{-3}$, and iter = 10. Figures 1 and 2 display the first 16 basis elements obtained with NMU and spatial NMU respectively, while Figure 3 displays other basis elements extracted by spatial NMU further in the recursion. Spatial NMU clearly generates a better decomposition, being able to separate more materials and obtaining basis elements with less noise and better spatial coherence. Moreover, spatial NMU preserves the edges of the constitutive materials. These properties can be easily observed, for example, on the seventh basis elements obtained with NMU and spatial NMU (Chalcedony).

Notice that some basis elements obtained with spatial NMU are still rather noisy; in particular the tenth and twelfth basis elements. Increasing the penalty μ allows one to improve upon these results, see Figure 4, where we used $\lambda = 0.3$ (instead of $\lambda = 0.1$) allowing the extracted materials to appear much more clearly.

Table 1 gives the materials present in the the basis elements extracted by spatial NMU. Some materials are particularly well separated: Hematite (a2), Alunite (a3), Chalcedony (b3), Kaolinite (b4), Muscovite (c1 and f2), Montmorillonite (c3), Alunite-K (d3) and Koalinite wxl (e4), while others remain mixed, e.g., jarosite (f1) is mixed with some Hematite. (In future work, we plan to include both sparsity and spatial information which will allow us to separate materials in a more efficient way.)

	(1)	(2)	(3)	(4)
(a)	Mixture	Hematite	Alunite	Kaolinite #1 Goethite
(b)	Muscovite Montmorillonite	Amorphous iron oxydes Goethite	Chalcedony	Kaolinite #1
(c)	Muscovite	Hematite Muscovite Buddingtonite	Montmorillonite	Kaolinite #2
(d)	Alunite-K Goethite	Muscovite Kaolinite #3	Alunite-K	Fine-grained Hematite
(e)	Hematite	Nontronite	Nontronite	Kaolinite wxl
(f)	Jarosite	Muscovite	Buddingtonite	Desert Varnish

Table 1. Classification: materials present in the abundance maps generated by spatial NMU, see Figures 2 and 3.

Because of the recursive procedure in generating U and V by NMU, the factor V does not correspond directly to the spectral signatures of the extracted materials, cf. the discussion in.[12] Many approaches are possible to obtain the spectral signatures from the abundance maps. We propose here a simple strategy: we take the weighted average of the spectral signatures of the pixels present in the corresponding abundance map, i.e., for the ith abundance map u_i, we take

$$s_i = \sum_k u_{ki} m_k^T.$$

Figure 5 displays some of the spectral signatures hence obtained, compared with the spectral signatures from the USGS library (`http://speclab.cr.usgs.gov/spectral.lib06/`) and the ones obtained with the Vertex Component Analysis (VCA) algorithm.[13] Spatial NMU performs similarly as VCA and is able to extract some

[†]Available at `http://aviris.jpl.nasa.gov/html/aviris.freedata.html`.

materials that were not extracted by VCA (e.g., Chalcedony and Jarosite) and generate spatially more coherent abundance maps (the ones obtained from the VCA spectral signatures are much more noisy and difficult to interpret, see[13]).

4. CONCLUSIONS

In summary, we have extended the original nonnegative matrix underapproximation model by formulating the spatial closeness of hyperspectral pixels using an l_1-norm regularization term. Thus the material constitutions of neighboring pixels are constrained to be similar, which effectively reduces the noise often seen in abundance maps produced by linear unmixing models. Since l_1-norm penalty terms are generally difficult to work with, we relaxed it with iteratively re-weighted least squares to convert it into a convex quadratic program, which is then solved with a projected gradient scheme. Experiments on the Cuprite dataset show comparatively better results, especially in terms of the sharpness of abundance maps, and even shows extracted materials not obtained by the VCA algorithm. Future work will include refinement of our spatial NMU algorithm and numerical evaluations on additional hyperspectral datasets.

REFERENCES

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

[2] Zhang, L., Zhang, L., Tao, D., and Huang, X., "A multifeature tensor for remote sensing target recognition," *IEEE Geoscience and Remote Sensing Letters* **8 (2)**, 374–378 (2011).

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

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

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

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

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

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

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

[10] Kopriva, I., Chen, X., and Jao, Y., "Nonlinear Band Expansion and Nonnegative Matrix Underapproximation for Unsupervised Segmentation of a Liver from a Multi-phase CT image," in [*SPIE Medical Imaging-Image Processing Volume 7962, Orlando*], (2011).

[11] Kopriva, I., Hadzija, M., Hadzija, M., Korolija, M., and Cichocki, A., "Rational variety mapping for contrast-enhanced nonlinear unsupervised segmentation of multispectral images of unstained specimen," *The American Journal of Pathology* **179(2)**, 547–554 (2011).

[12] Gillis, N. and Plemmons, R., "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).

[13] Nascimento, J. and Dias, J., "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing* **43(4)**, 898–910 (2005).

[14] Zymnis, A., Kim, S.-J., Skaf, J., Parente, M., and Boyd, S., "Hyperspectral image unmixing via alternating projected subgradients," in [*Signals, Systems and Computers, 2007*], 1164 –1168 (2007).

[15] Iordache, M.-D., Bioucas-Dias, J., and Plaza, A., "Total variation regulatization in sparse hyperspectral unmixing," in [*Third Worskshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon*], (2011).
[16] Bottou, L., [*Online algorithms and stochastic approximations*], Online Learning and Neural Networks, David Saad, Ed. Cambridge University Press, Cambridge, UK (1998).
[17] Daubechies, I., DeVore, R., Fornasier, M., and Gunturk, C., "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics* **63**(1), 1–38 (2010).
[18] Nesterov, Y., [*Introductory lectures on convex optimization: a basic course*], Kluwer, Boston (2004).

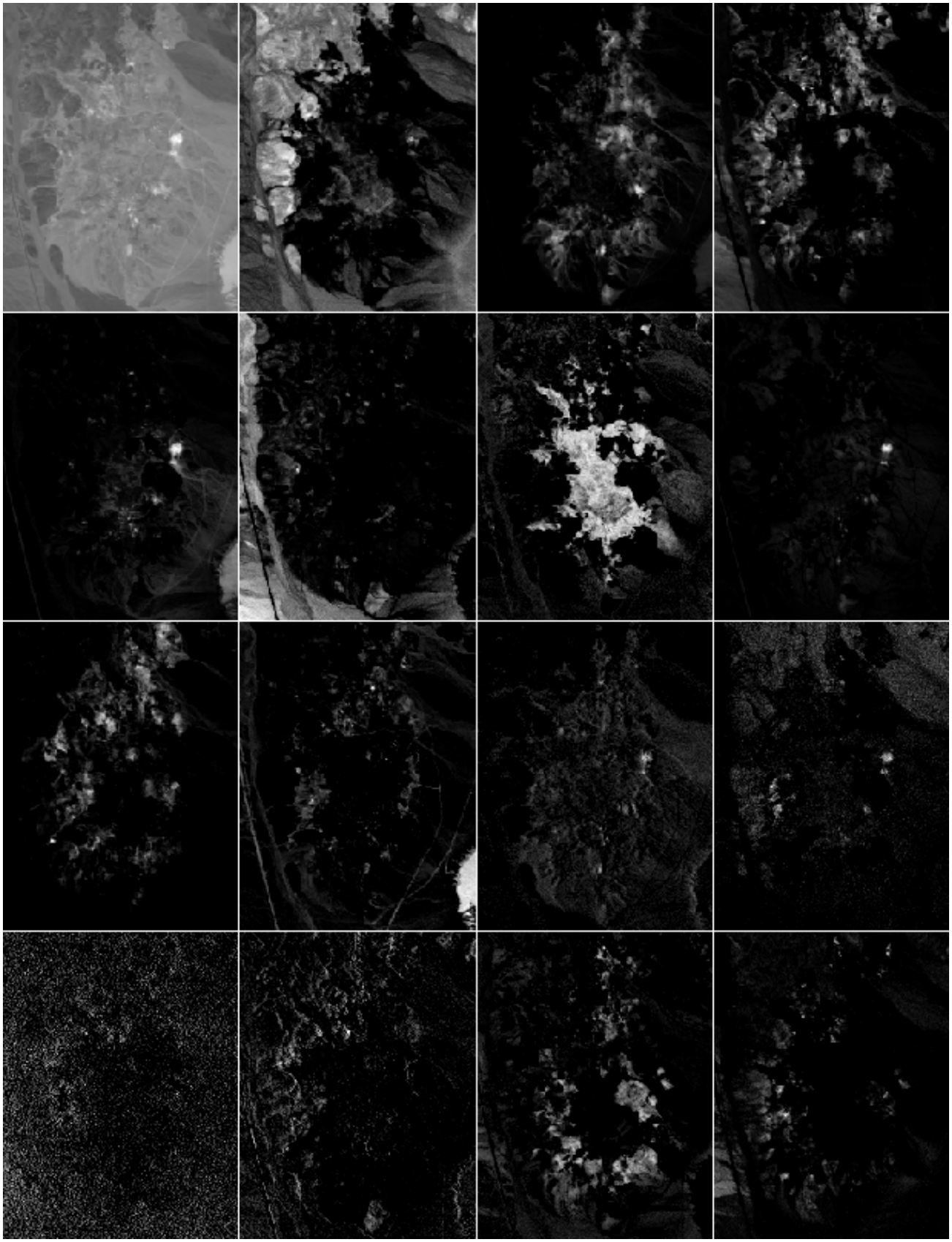

Figure 1. First 16 abundance maps extracted by NMU. (Light tones indicate a high degree of membership.)

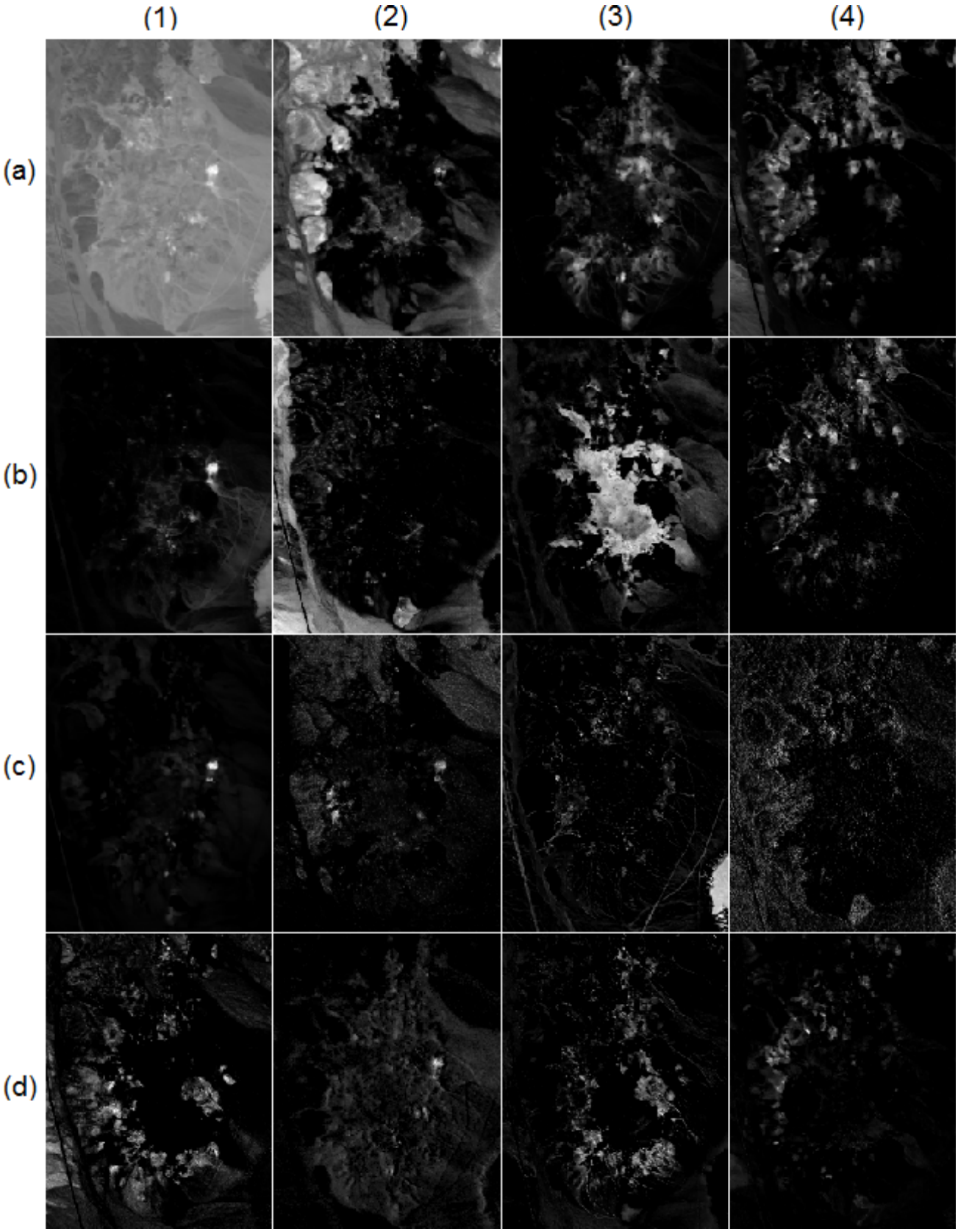

Figure 2. First 16 abundance maps extracted by spatial NMU ($\lambda = 0.1$), see Table 1 for the classification. (Light tones indicate a high degree of membership.)

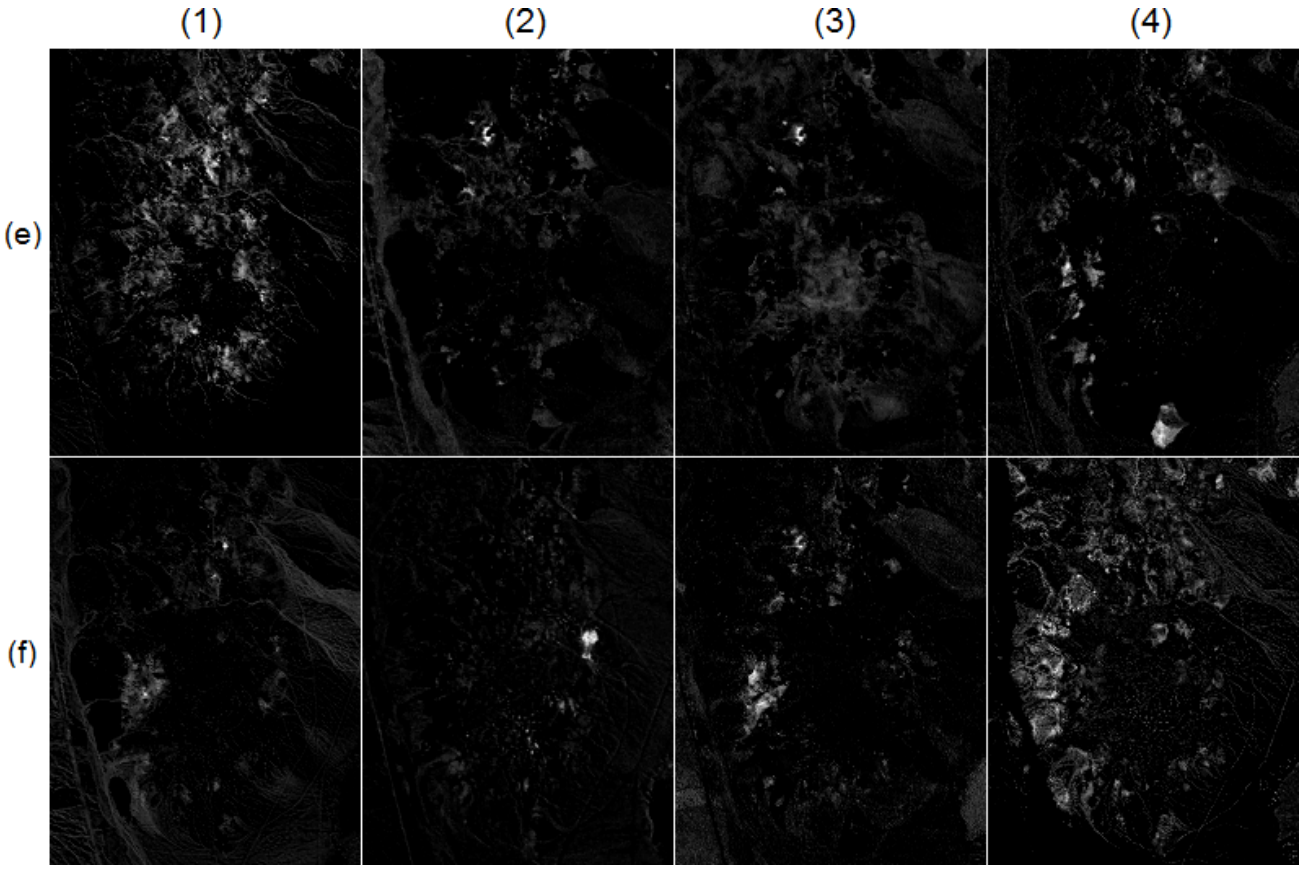

Figure 3. Other abundance maps extracted by Spatial NMU ($\lambda = 0.1$), see Table 1 for the classification. (Light tones indicate a high degree of membership.)

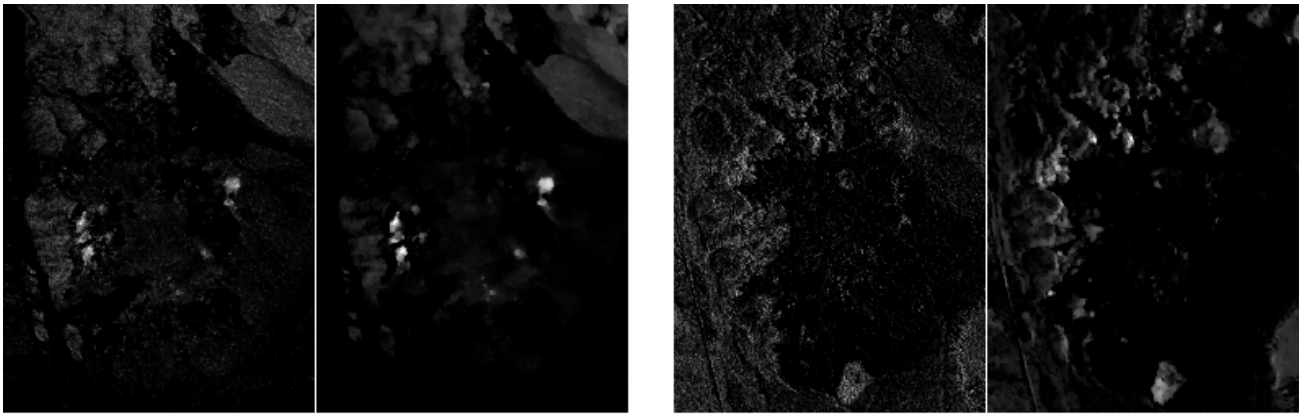

Figure 4. (i) Tenth basis element obtained with spatial NMU for $\lambda = 0.1$, and (ii) for $\lambda = 0.3$. (iii) Twelfth basis element obtained with spatial NMU for $\lambda = 0.1$, and (iv) for $\lambda = 0.3$.

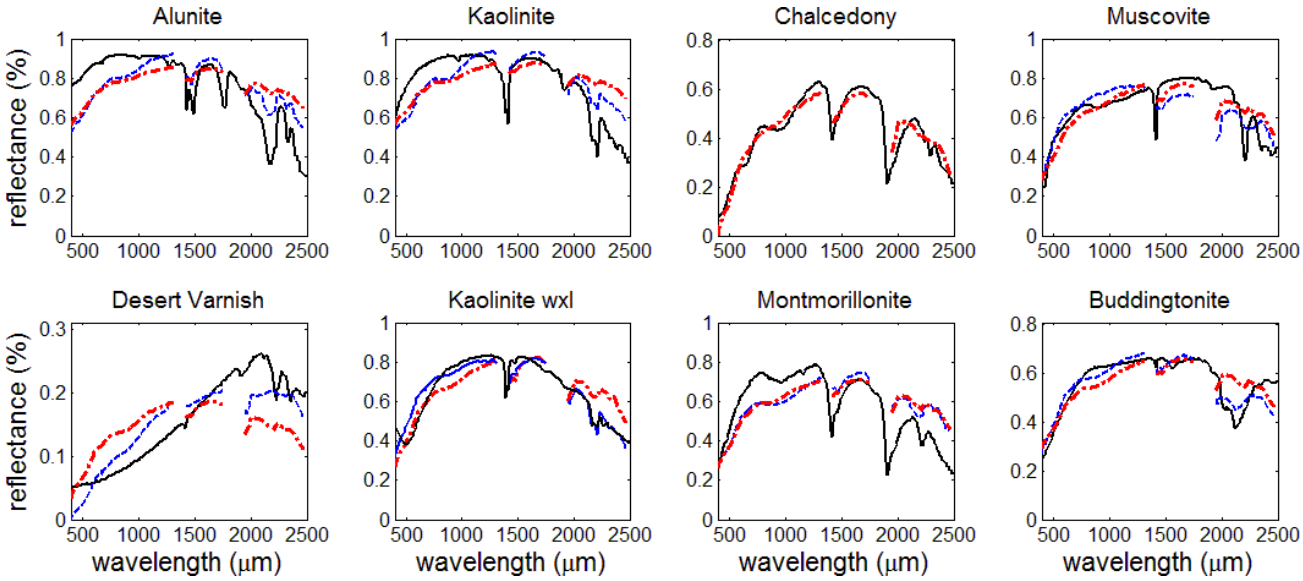

Figure 5. Spectral signatures from the USGS library (continuous black curves), spectral signatures obtained with spatial NMU (dashed-dotted red curves), and with VCA (dashed blue curves).