Non-Negative Sparse Mathematical Morphology

Jesus Angulo, Santiago Velasco-Forero

Non-Negative Sparse Mathematical Morphology

Jesús Angulo, Santiago Velasco-Forero

MINES ParisTech, PSL-Research University,
CMM-Centre de Morphologie Mathématique; France
jesus.angulo@mines-paristech.fr ; santiago.velasco@mines-paristech.fr

June 2017

Abstract

Sparse modeling involves constructing a succinct representation of initial data as a linear combination of a few typical atoms of a dictionary. This paper deals with the use of sparse representations to introduce new nonlinear image filters which efficiently approximate morphological operators. Reasons why non-negative matrix factorization (NMF) is a dimensional reduction (i.e., dictionary learning) paradigm particularly adapted to the nature of morphological processing are given. In particular, Sparse-NMF representations are studied and used to introduce first approximations to binary dilations/erosions and then to openings/closings. The idea behind consists of processing exclusively the image dictionary and then, the result of processing each image is approximated by multiplying the processed dictionary by the coefficient weights of the current image. These operators are then extended to gray-scale images and their interest for feature detection is illustrated. The practical relevance of our approach is considered for two applications on multivariate image processing. The first case deals with multispectral texture modeling using Boolean random set theory; the second case with multi-scale decomposition of hyperspectral images and its interest in spectral-spatial pixel classification.

Keywords: mathematical morphology; sparse modeling; non-negative matrix factorization ; sparse NMF ; non-linear filtering ; hyperspectral image processing.

1 Introduction

Modern image processing techniques should deal with images datasets of huge volume, great variety (e.g., optical, radar and hyperspectral sensors, sequence of images often multi-scale, etc.) and different veracity (i.e., incomplete images, or different uncertainty in the acquisition procedure). This paper addresses the problem of simultaneous treatment of a set of images via mathematical morphology. Our approach is motivated by sound developments in sparse signal representation (or *coding*), which suggest that the linear relationships among

high-resolution signals can be accurately recovered from their low-dimensional projections (also called *dictionary*) (Donoho, 2006). Sparse coding and dictionary learning, where data is assumed to be well represented as a linear combination of a few elements from a dictionary, is an active research topic which leads to state-of-the-art results in image processing applications, such as image denoising, inpainting, demosaicking or compression (Elad and Aharon, 2006; Mairal et al., 2008; Yu et al., 2011). For a detailed self-contained overview on sparse modeling for image and vision processing the reader is referred to (Mairal et al., 2012). Inspired by this paradigm of parsimony representation of images, the aim of this work is to explore how image sparse representations can be useful to efficiently calculate approximations to morphological operators.

Mathematical morphology (Serra, 1982; Heijmans, 1994) is a nonlinear image processing methodology based on the application of lattice theory to spatial structures. Morphological filters and transformations are useful for various image processing tasks (Soille, 1999), such as denoising, contrast enhancement, multi-scale decomposition, feature extraction and object segmentation. In addition, morphological operators are defined using very intuitive geometrical notions which allows us the perceptual development and interpretation of complex algorithms by combination of various operators.

Notation. Let E be a domain of points, which is considered here as a finite digital space of the pixels of the image, i.e., $E \subset \mathbf{Z}^2$ such that $N = |E|$ is the number of pixels. Image intensities are numerical values, which ranges in a closed subset $\mathcal{T}$ of $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, +\infty\}$; for example, for an image of discrete L values, it can be assumed $\mathcal{T} = \{t_1, t_2, \cdots, t_L\}$. A binary image X is modeled as a subset of E, i.e., $X \in \mathcal{P}(E)$; a grey-level image $f(p_i)$, where $p_i \in E$ are the pixel coordinates, is a numerical function $E \to \mathcal{T}$, i.e., $f \in \mathcal{F}(E, \mathcal{T})$. In this paper, we are interested in operators ψ as a map transforming an image into an image. There are thus operators on binary images, i.e., maps $\mathcal{P}(E) \to \mathcal{P}(E)$; or on grey-level images, i.e., maps $\mathcal{F}(E, \mathcal{T}) \to \mathcal{F}(E, \mathcal{T})$.

Motivation and outline of the approach. In order to illustrate our motivation, let us consider the diagrams depicted in Fig. 1. In many practical situations, a collection $\mathcal{F} = \{f_1(p_i), \cdots, f_M(p_i)\}$ of M binary or grey-level images (each image having N pixels) should be analyzed by applying the same morphological operator ψ (or a series of operators) to each image of $\mathcal{F}$, i.e., $\psi(\mathcal{F}) = \{\psi(f_j)\}_{1 \leq j \leq M}$, Fig. 1(a). If one considers that the content of the various images is relatively similar, we can expect that the initial collection can be efficiently projected into a R dimensionality reduced image space, using typically sparse modeling techniques, i.e., $\mathcal{F} \approx \Phi H$, in order to learn the dictionary composed of atom images $\Phi = \{\phi_1(p_i), \cdots, \phi_R(p_i)\}$ and the corresponding coding or weights H, Fig. 1(b). In this context, we would like to apply the morphological operator ψ (or an equivalent operator) to the reduced set of images of the projective space Φ, i.e., $\psi(\Phi) = \{\psi(\phi_k)\}_{1 \leq k \leq R}$, in such a way that the original processed images are approximately obtained by projecting back to

the initial space, i.e., $\psi(\mathcal{F}) \approx \psi(\Phi) H$, Fig. 1(c). Note that if the number of operators to be applied increases, e.g., Φ_A, Φ_B, Φ_C, etc., then such paradigm becomes much more efficient since $R \ll M$, Fig. 1(d). At this point, we should remark that our approach is different from the standard use of sparse modeling in image processing (Mairal et al., 2012), since in the classical approaches the dictionary is not processed. In fact, sparse representation itself leads to a solution for the tasks of denoising, regularization, coding, etc.

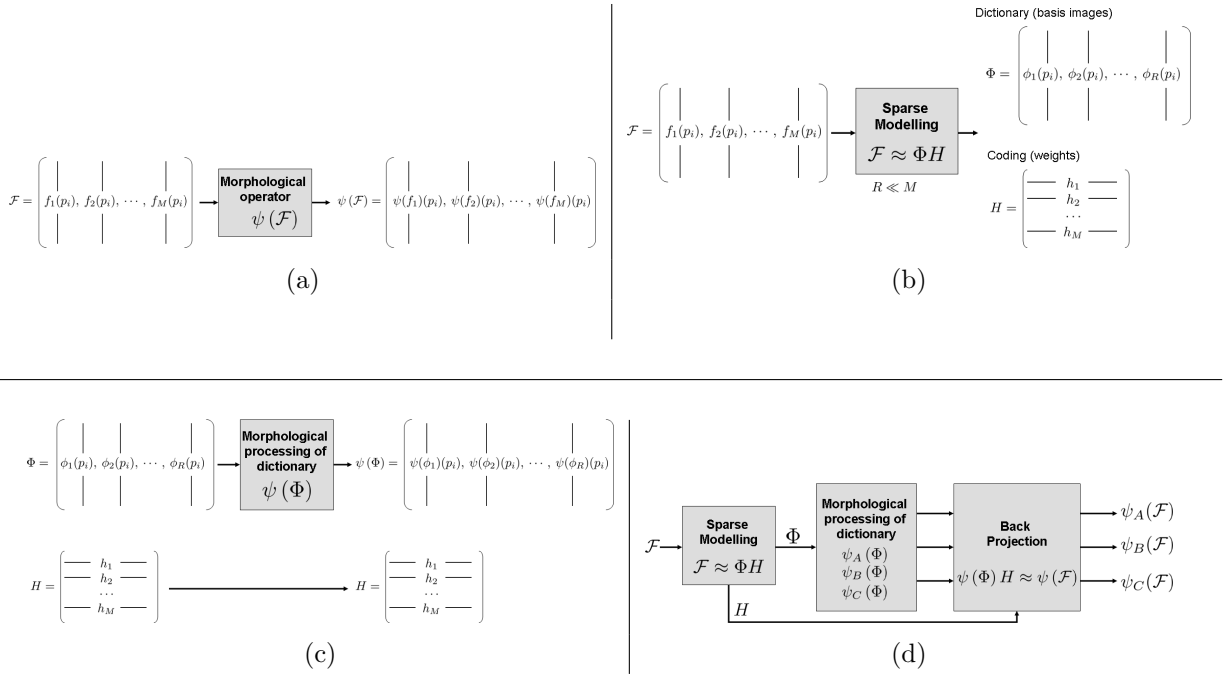


Figure 1: Sparse mathematical morphology: motivation and outline.

Typical examples of image families which can be fit in our framework are: i) a collection of binary shapes, ii) a database of registered grey scale images, for instance, faces, cells (from biomicroscopy applications) or galaxies (in astronomy), iii) a set of patches of a large image, iv) a time series of registered images, v) the spectral bands of a multi/hyper-spectral image, etc. Some of these examples are used to illustrate the relevance of our approach. The rationale behind this methodology is that the *intrinsic dimension*[1] of the image collection is lower than M. Usually the subspace representation involves deriving a set of basis components using linear techniques like Principal Component Analysis (PCA), Independent Component Analysis (ICA), or sparse dictionary learning techniques like K-SVD (Elad and Aharon, 2006). The projection coefficients for the linear combinations in the above methods can be either positive or negative, and such linear combinations generally involve complex cancellations

[1]Roughly, the intrinsic dimension of an image f is defined as the number of "independent" parameters needed to represent f.

between positive and negative numbers. Therefore, these representations lack the intuitive meaning of "adding parts to form a whole". This property is particularly problematic in the case of mathematical morphology since the basic binary operator, the dilation of a set, is defined as the operator which commutes with the union of parts of the set. This principle is widely discuss in the paper. In practice, that means that the paradigm discussed above in Fig. 1 cannot be used with sign unconstrained sparse modeling methods. Non-negative matrix factorization (NMF) (Lee and Seung, 1999) imposes the non-negativity constraints in learning basis images: the pixels values of resulting images as well as the coefficients for the reconstruction are all non-negative. This ensures that NMF is a procedure for learning a parts-based representation (Lee and Seung, 1999). In addition, by incorporating sparse constraint to NMF, it will be constructed a succinct representation of the image data as a combination of a few typical patterns (few atoms of the dictionary) learned from the data itself. Hence, our approach of sparse mathematical morphology is founded on sparse NMF modeling.

NMF state-of-the-art. The state-of-the on NMF is nowadays vast. Besides the pioneering work (Lee and Seung, 1999), several NMF variants have been developed by introducing additional constraints and properties to the original NMF. In particular, being relevant for our purpose, we can mention, on the one hand, the local NMF (Li et al., 2001), which learns spatially localized, parts-based subspaces for images (i.e., visual patterns); and on the other hand, the sparse NMF (Hoyer, 2004), which incorporates explicit sparseness constraints. These powerful and already classical algorithms are reviewed in next section. We note also that in hyperspectral imaging, there is a common task called *linear unmixing* (or factorization into physical space) which consists in detecting the spectra of the pure materials and estimating their relative abundances. Linear unmixing can be seen as an equivalent problem to NMF (Esser et al., 2012).

Formulation of PCA-like optimization with nonnegative and sparse constraints have also considered by several works, see for instance (Zass and Shashua, 2006). Motivated by a combination of NMF and kernel theory (emmbedding into polynomial feature space), various nonlinear variants of NMF have been proposed (Zafeiriou and Petrou, 2010). Similarly, algorithms have been also introduced for nonnegative ICA (Plumbley, 2003). NMF has been extended to factorization of non vector spaces, typically for data sampled from a low-dimensional manifold embedded in a high-dimensional space, using affinity graph matrix non-negative decomposition (Cai et al., 2011) and manifold regularized NMF (Guan et al., 2011). NMF has been also extended to tensor factorizations (Cichocki et al., 2009), and has considered for applications to exploratory multi-way data analysis and blind source separation. In general, NMF is a NP-hard problem (Vavasis, 2009). However, starting from the notion of separable NMF, introduced by (Donoho and Stodden, 2004), which in geometric terms involves that the conical hull of a small subsets of the columns of the data matrix contain the rest of the columns, an emerging line of work in NMF, based on computational

geometry approaches, is leading to efficient solutions. It can be then solved by linear programming (Arora et al., 2012), conical hull algorithms (Kumar et al., 2013) and random projections (Damle and Sun, 2014). When the data to be factorized is binary or Boolean (i.e., only "0" or "1" values), the NMF problem is called discrete basis problem (Miettinen et al., 2008). This discrete basis problem which is equivalent to the binary k-median problem is NP-hard and some greedy algorithms (Miettinen, 2010) have been proposed to deal with. Other binary matrix factorization approaches are straightforward related to classical NMF (Zhang et al., 2007) either by a thresholding algorithm (smooth approximation to Heaviside function) or by binary-constraint penalty function. A binary version of the sparse NMF can be also mentioned (Yuan et al., 2009), based on windowed image parts with Haar-like box functions. Extension of NMF to binary data factorization has been also considered by means of a Bernoulli distribution of binary observations (Schachtner et al., 2010), or more recently using a logistic PCA approach (Tomé et al., 2015).

Paper organisation. The methodology described here was initiated in (Angulo and Velasco-Forero, 2011) for binary erosion/dilation. In the present work, we develop also sparse approximations to binary opening/closing as well as the full counterpart of sparse approximation to morphological operators for gray-scale images. The applications considered in the last part of the paper are also new with respect to our previous work (Angulo and Velasco-Forero, 2011).

The rest of paper is structured as follows. Section 2 reviews the formal definition of NMF and various algorithms proposed in the state-of-the-art, including the sparse variant introduced in (Hoyer, 2004), which is the one used in the our study. The use of NMF representations for implementing sparse pseudo-morphological binary operators is introduced and theoretically justified in Section 3. The section discusses also why sparse NMF is appropriate in morphological image processing. Section 4 generalizes the framework of morphological sparse processing to gray-scale images using two alternatives. The first one is based on level set decomposition of numerical functions. The second approach, more useful in practice, uses a straightforward representation of sparse NMF basis from the gray-scale images. Applications are studied in Section 5. The first case-study deals with the sparse processing and modeling of a multivariate Boolean texture. The second application focusses on morphological sparse processing of hyperspectral images. Conclusions and perspectives close the paper in Section 6.

2 NMF and Sparse Variants

2.1 Definition of NMF on vector space

Let us assume that our data consists of M vectors of N non-negative scalar variables. Denoting the column vector $\mathbf{v}_j$, $j = 1, \cdots, M$, the matrix of data is obtained as $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_M)$

(each $\mathbf{v}_j$ is the j-th column of $\mathbf{V}$), with $|\mathbf{v}_j| = N$. If we analyze M images of N pixels, these images can be stored in linearized form, so that each image will be a column vector of the matrix.

Given the non-negative matrix $\mathbf{V} \in \mathbb{R}^{N \times M}$, $\mathbf{V}_{i,j} \geq 0$, NMF is defined as a linear non-negative approximate data decomposition into the two matrices $\mathbf{W} \in \mathbb{R}^{N \times R}$ and $\mathbf{H} \in \mathbb{R}^{R \times M}$ such that

$$\mathbf{V} \approx \mathbf{WH}, \quad \text{s.t. } \mathbf{W}_{i,k}, \mathbf{H}_{k,j} \geq 0, \tag{1}$$

where usually $R \ll M$, which therefore involves a dimensionality reduction. Each of the R columns of $\mathbf{W}$ contains a basis vector $\mathbf{w}_k$ and each row of $\mathbf{H}$ contains the coefficient vector (weights) $\mathbf{h}_j$ corresponding to vector $\mathbf{v}_j$: $\mathbf{v}_j = \sum_{k=1}^{R} \mathbf{w}_k \mathbf{H}_{k,j} = \mathbf{W}\mathbf{h}_j$. Using the modern terminology from sparse coding theory, the matrix $\mathbf{W}$ contains the dictionary and $\mathbf{H}$ the encoding.

A theoretical study of the properties of NMF representation has been achieved in (Donoho and Stodden, 2004) using geometric notions. Hence, NMF is interpreted as the problem of finding a simplicial cone which contains the data points in the positive orthant, or in other words, NMF is a conical coordinate transformation. It is interesting to note that NMF is somehow equivalent to Kernel k-means clustering and Laplacian-based spectral clustering (Ding et al., 2005).

2.2 Basic NMF algorithms

The factorization $\mathbf{V} \approx \mathbf{WH}$ is not necessarily unique, and the optimal choice of matrices $\mathbf{W}$ and $\mathbf{H}$ depends on the cost function that minimizes the reconstruction error. The most widely used is the Euclidean distance, i.e., minimize

$$\|\mathbf{V} - \mathbf{WH}\|_2^2 = \sum_{i,j} \left(\mathbf{V}_{i,j} - (\mathbf{WH})_{i,j}\right)^2,$$

with respect to $\mathbf{W}$ and $\mathbf{H}$, and subject to the constraints $\mathbf{W}, \mathbf{H} > 0$. Although the minimization problem is convex in $\mathbf{W}$ and $\mathbf{H}$ separately, it is not convex in both simultaneously. In (Lee and Seung, 2001) a multiplicative good performance algorithm to solve (1) is proposed. They proved that the cost function is nonincreasing at the iteration and the algorithm converges at least to a local optimal solution. More precisely, the update rules for both matrices are:

$$\mathbf{H}_{k,j} \leftarrow \mathbf{H}_{k,j} \frac{(\mathbf{W}^T\mathbf{V})_{k,j}}{(\mathbf{W}^T\mathbf{WH})_{k,j}}; \quad \mathbf{W}_{i,k} \leftarrow \mathbf{W}_{i,k} \frac{(\mathbf{VH}^T)_{i,k}}{(\mathbf{WHH}^T)_{k,j}}.$$

Another useful cost function, also considered in (Lee and Seung, 2001), is the Kullback-Leibler (KL) divergence, which leads also quite simple multiplicative update rules.

In (Li et al., 2001), a variant of KL divergence NMF was proposed, which is named Local NMF (LNMF), aiming at learning spatially localized components (by minimizing the number of basis R to represent $\mathbf{V}$ and by maximizing the energy of each retained components) as well

as imposing that different bases should be as orthogonal as possible (in order to minimize redundancy between the different bases). The multiplicative update rules for LNMF are given by

$$\mathbf{H}_{k,j} \leftarrow \sqrt{\mathbf{H}_{k,j} \sum_i \mathbf{V}_{i,j} \frac{\mathbf{W}_{i,k}}{(\mathbf{WH})_{i,k}}}; \quad \mathbf{W}_{i,k} \leftarrow \mathbf{W}_{i,k} \frac{\sum_j \mathbf{V}_{i,j} \frac{\mathbf{H}_{k,j}}{(\mathbf{WH})_{i,j}}}{\sum_j H_{k,j}}; \quad \mathbf{W}_{i,k} \leftarrow \frac{\mathbf{W}_{i,k}}{\sum_i \mathbf{W}_{i,k}}.$$

As we have discussed in the introduction, there are many other alternative NMF algorithms, but we focus on the following technique which provide a sparse variant.

2.3 NMF with sparseness constraints

A very powerful framework to add an explicit degree of sparseness in the basis vectors $\mathbf{W}$ and/or the coefficients $\mathbf{H}$ was introduced in (Hoyer, 2004). First of all, the sparseness measure σ of a vector $\mathbf{v} \in \mathbb{R}^{N \times 1}$ used in (Hoyer, 2004) is based on the relationship between the L_1 norm and the L_2 norm:

$$\sigma(\mathbf{v}) = \frac{\sqrt{N} - \|\mathbf{v}\|_1 / \|\mathbf{v}\|_2}{\sqrt{N} - 1}.$$

This function is maximal at one if and only if $\mathbf{v}$ contains only a single non-zero component, and takes a value of zeros if and only if all components are equal (up to signs). Then, matrix $\mathbf{W}$ and $\mathbf{H}$ are solved by the problem (1) under additional constraints $\sigma(\mathbf{w}_k) = S_w$ and $\sigma(\mathbf{h}_j) = S_h$, where S_w and S_h are respectively the desired sparseness of $\mathbf{W}$ and $\mathbf{H}$. The algorithm introduced in (Hoyer, 2004) is a projected gradient descent algorithm (additive update rule), which takes a step in the direction of the negative gradient

$$\mathbf{W}_{i,k} \leftarrow \mathbf{W}_{i,k} - \mu_{\mathbf{W}} \left(\mathbf{W}_{i,k}\mathbf{H}_{k,j} - \mathbf{V}_{i,j}\right) \mathbf{H}^T_{j,k}; \quad \mathbf{H}_{k,j} \leftarrow \mathbf{H}_{k,j} - \mu_{\mathbf{H}} \mathbf{W}^T_{k,i} \left(\mathbf{W}_{i,k}\mathbf{H}_{k,j} - \mathbf{V}_{i,j}\right);$$

and subsequently projects each column of $\mathbf{W}$ and each row of $\mathbf{H}$ onto the constraint space. The most sophisticated step is therefore how to find, for a given vector $\mathbf{v}$, the closest non-negative vector $\mathbf{u}$ with a given L_1 norm and a given L_2 norm. The algorithm works as follows (Hoyer, 2004). The vector is projected onto the hyperplane of L_1. Next, within this space, one projects to the closest point on the joint constraint hypersphere (intersection of the sum and the L_2 constraints), by moving radially outward for the center of the sphere (the center is the point where all the components have equal values). If the components of this projection point are not completely non-negative, the negative values must be fixed at zero and a new point found in similar way under the same constraints. Sparseness is controlled explicitly with a pair of parameters that is easily interpreted; in addition, the number of required iterations grows very slowly with the dimensionality of the problem. In fact, for all the empirical tests considered in this paper, we have used the MATLAB code for performing NMF and its various extensions (LNMF, sparse NMF) provided by Hoyer (2004).

Besides the sparseness parameters (S_w, S_h), a crucial parameter to be chosen in any NMF algorithm is the value of R, that is, the number of basis of projective reduced space.

Any dimensionality reduction technique, such as PCA, requires also to fix the number of components. In PCA, the components are ranked according to the second-order statistical importance of the components and each one has associated a value of the represented variance; whereas in NMF the selection R can be evaluated only *a posteriori*, by evaluating the error of reconstruction. This issue is not considered in the paper and we empirically fix R.

2.4 A few properties of NMF

We conclude this review on NMF algorithms by discussing some well-known propeorties which are useful for the sequel.

- **Boundedness property (for any NMF variant)** (Zhang et al., 2007): We say that $\mathbf{V}$ is bounded if $0 \leq \mathbf{V}_{i,j} \leq 1$. If $\mathbf{V}$ is bounded, then the factor matrices $\mathbf{W}$ and $\mathbf{H}$ are also bounded, i.e., $0 \leq \mathbf{W}_{i,k} \leq 1$ and $0 \leq \mathbf{H}_{k,j} \leq 1$.

- **Indeterminacies (for any NMF variant)** (Theis et al., 2005; Huang et al., 2014): Positivity and sparseness are invariant under permutation and scaling of columns of $\mathbf{W}$ (and correspondingly of the rows of $\mathbf{H}$), i.e., $\mathbf{V} = \mathbf{WH} = (\mathbf{WP}^{-1}\mathbf{L}^{-1})(\mathbf{LPH})$, where $\mathbf{P}$ is a permutation matrix and $\mathbf{L}$ a positive scaling matrix.

- **Uniqueness (only for Sparse-NMF)** (Theis et al., 2005; Huang et al., 2014): Under sparsity constraints, projection step proposed in (Hoyer, 2004) has a unique solution, which is found. Hence, under non-degenerate data, Sparse-NMF a.s. produces a unique factorization. This is another fundamental reason why we think that Sparse-NMF is an excellent choice for our learning our non-negative dictionaries.

3 Sparse approximation to binary morphological operators

Let $\mathcal{X} = \{X_1, \cdots, X_M\}$ be a collection of M binary shapes, called a *family of shapes*, i.e., $X_j \in \mathcal{P}(E)$. For each shape X_j, let $\mathbf{x}_j(i) : I \to \{0,1\}$, with $i \in I = \{1, 2, \cdots, N\}$ and $N = |E|$, be its *indicator vector*:

$$\forall X_j \in \mathcal{P}(E),\ \forall p_i \in E,\ \ \mathbf{x}_j(i) = \begin{cases} 1 & \text{if } p_i \in X_j \\ 0 & \text{if } p_i \in X_j^c \end{cases} \tag{2}$$

Then the family of shapes $\mathcal{X}$ has associated a data matrix $\mathbf{V} \in \{0,1\}^{N \times M}$, where each indicator vector corresponds to one of its columns, i.e., $\mathbf{V}_{i,j} = \mathbf{x}_j(i)$.

3.1 Sparse NMF approximations of binary sets

After computing the NMF representation on data $\mathbf{V}$, for a given constant $R > 0$, an approximation to $\mathbf{V}$ is obtained. More precisely, if we denote by $\phi_k(p_i) : E \to \mathbb{R}^+$ the basis images associated to the basis matrix $\mathbf{W}$, i.e., $\phi_k(p_i) = \mathbf{W}_{i,k}$, the following image is obtained as

$$a_{X_j}(p_i) = \sum_{k=1}^{R} \phi_k(p_i) \mathbf{H}_{k,j} \tag{3}$$

It is obvious that without any additional constrains, function $a_X(p_i)$ is not a binary image. By the boundedness property discussed above, we have $0 \leq a_X(p_i) \leq 1$. Hence, a thresholding operation at value α is required to impose a binary approximate set $\widetilde{X}_j$ to each initial shape X_j, i.e.,

$$X_j \xrightarrow{NMF} \widetilde{X}_j : \quad p_i \in \widetilde{X}_j \ \text{ if } \ a_X(p_i) > \alpha \tag{4}$$

We propose to fix, for all the examples of the paper, the threshold value to $\alpha = 0.45$, in order to favor the reconstruction of X_j against its complement.

Let us consider a practical example from a binary image collection $\mathcal{X}$, composed of $M = 100$ images from the Fish shape database ($N = 400 \times 200$), see in Fig. 2 a few examples of images. Fig. 3 depicts the corresponding basis images for various NMF algorithms: we have fixed $R = 10$ for all the cases (factor 10 of dimensionality reduction). We observe that standard NMF produces a partial part-based representation, which also includes almost complete objects for outlier shapes (basis 2-upper-center and 5-center-center). As expected, LNMF produces more local decompositions, however the orthogonality constraints involves also an atomization of some common parts. A similar problem arises for Sparse-NMF when $S_w \neq 0$ (constraint of sparsity in basis matrix $\mathbf{W}$). When the sparsity constraint is limited to the coding S_h, with a typical value around 0.6, the obtained dictionary of shapes is less local, but in exchange, this constraint involves that each binary shapes is reconstructed using a limited number of atoms. The various groups of fish shapes are therefore better approximated by the latter case than using the other NMF algorithms. The comparison of

Fig. 4 illustrates qualitatively the better performance of Sparse-NMF ($S_w = 0$, $S_h = 0.6$) with respect to the others. We have also included in Fig. 3(a) the 10 first eigenimages obtained by PCA (in red the positive values and in blue the negative values); as expected, the corresponding representation does not fit with a part-based decomposition needed for morphological operators.

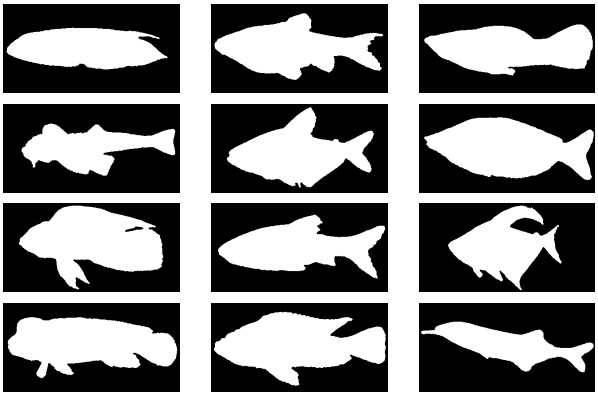


Figure 2: Examples of original binary images from the fish database, composed of 100 images.

3.2 Sparse max-approximation to binary dilation and erosion

Dilation and Erosion. The two fundamental morphological operators are the dilation and the erosion, which are defined respectively as the operators which commute with the union and the intersection. Given a *structuring element* $B \subseteq E$, i.e., a set defined at the origin which introduces the shape/size of the operator, the *dilation of a binary image* X by B and the *erosion of a binary image* X by B are defined respectively by (Serra, 1982; Heijmans, 1994; Bloch et al., 2007):

$$\delta_B(X) = \cup \{B(p_i) |\, p_i \in X\}\,, \tag{5}$$

and

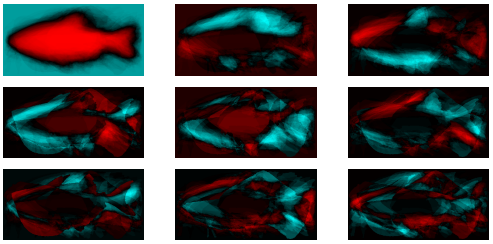
$$\varepsilon_B(X) = \{p_i \in E |\, B(p_i) \subseteq X\}\,, \tag{6}$$

where $B(p_i)$ is the structuring element centered at pixel p_i. In the case of numerical functions $f \in \mathcal{F}(E, \mathcal{T})$, which are considered in detail in next section, the *dilation of a grey-level image* is defined by (Heijmans, 1994; Soille, 1999; Bloch et al., 2007):

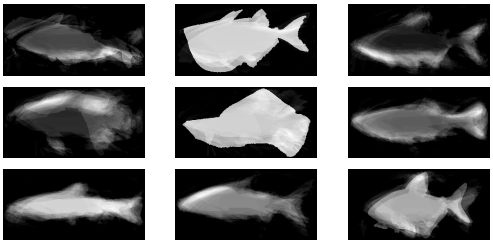
$$\delta_B(f)(p_i) = \left\{f(p_m) |\, f(p_m) = \sup\left[f(p_n)\right], p_n \in \check{B}(p_i)\right\} = \sup_{p_n \in \check{B}(p_i)} \left\{f(p_n)\right\}, \tag{7}$$

and the dual *grey-level erosion* is given by (Heijmans, 1994; Soille, 1999; Bloch et al., 2007)

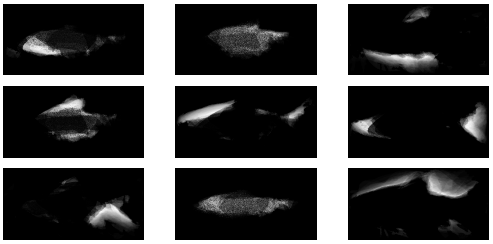
$$\varepsilon_B(f)(p_i) = \left\{f(p_m) |\, f(p_m) = \inf\left[f(p_n)\right], p_n \in B(p_i)\right\} = \inf_{p_n \in B(p_i)} \left\{f(p_n)\right\}, \tag{8}$$



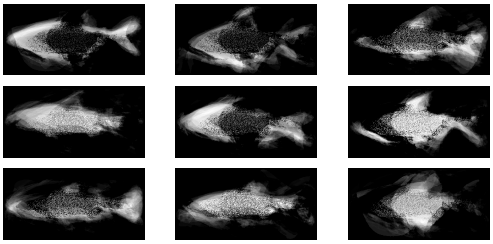
(a) PCA basis images



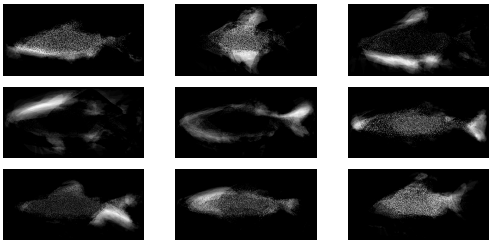
(b) NMF basis images



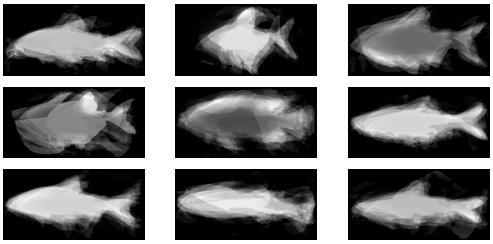
(c) LNMF basis images



(d) Sparse-NMF basis images ($S_w = 0.5$,$S_h = 0.5$)



(e) Sparse-NMF basis images ($S_w = 0.6$,$S_h = 0$)



(f) Sparse-NMF basis images ($S_w = 0$,$S_h = 0.6$)

Figure 3: PCA vs. non-negative representation of binary shapes. A collection of $M = 100$ shapes has been used in the experiments (see examples in Fig. 2), where the number of reduced dimensions has been fixed to $R = 10$ (in the examples are given the first 9 basis images).

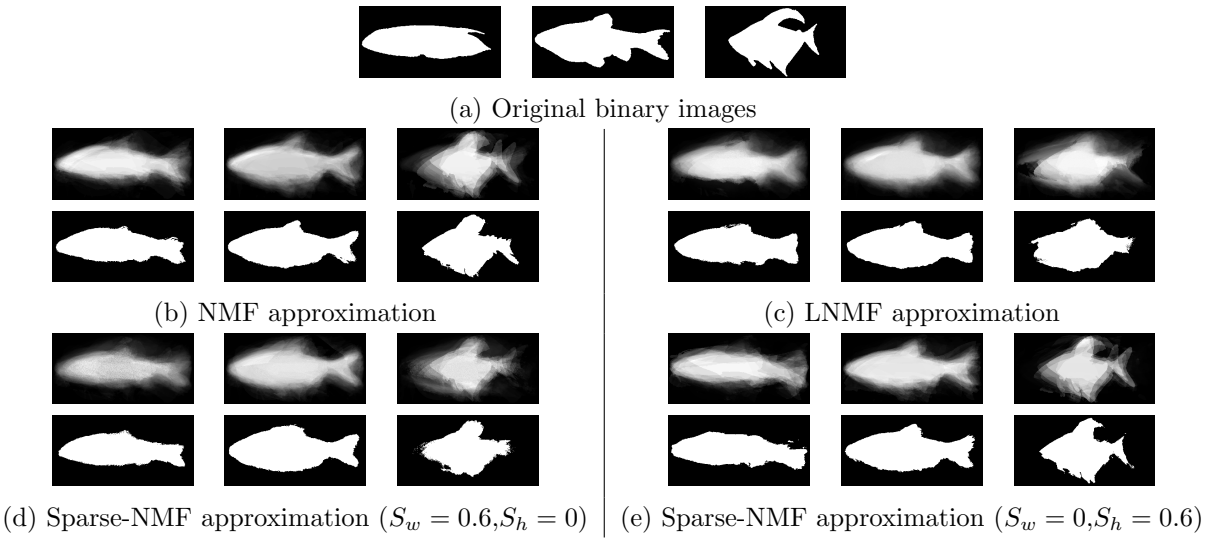
where $\check{B}(p_i)$ is the transposed structuring element centered at pixel p_i. If B is symmetric with respect to the origin, one has $\check{B} = B$.

Sparse max-approximation to binary dilation. Let us first introduce two basic notions. The *indicator function* of set X, denoted $\mathbb{1}_X : E \to \{0, 1\}$, is defined by

$$\forall X \in \mathcal{P}(E),\ \forall p_i \in E,\quad \mathbb{1}_X(p_i) = \begin{cases} 1 & \text{if } p_i \in X \\ 0 & \text{if } p_i \in X^c \end{cases} \tag{9}$$

Obviously, we have $\mathbb{1}_{X^c} = 1 - \mathbb{1}_X$. Given two sets $X, Y \in \mathcal{P}(E)$, one has the two following basic properties of indicator function:

$$\begin{aligned} \mathbb{1}_{X \cap Y} &= \min\{\mathbb{1}_X, \mathbb{1}_Y\} = \mathbb{1}_X \cdot \mathbb{1}_Y; \\ \mathbb{1}_{X \cup Y} &= \max\{\mathbb{1}_X, \mathbb{1}_Y\} = \mathbb{1}_X + \mathbb{1}_Y - \mathbb{1}_X \cdot \mathbb{1}_Y = \min\{1, \mathbb{1}_X + \mathbb{1}_Y\}. \end{aligned}$$



(a) Original binary images

(b) NMF approximation

(c) LNMF approximation

(d) Sparse-NMF approximation ($S_w = 0.6$,$S_h = 0$)

(e) Sparse-NMF approximation ($S_w = 0$,$S_h = 0.6$)

Figure 4: Sparse-NMF approximations to binary sets: (a) three original shapes X_j; (b)-(e) Top, reconstructed function a_{X_j} and Bottom, approximate set $\widetilde{X}_j$.

For a function $f : E \to \mathcal{T}$, the *thresholded set* at value $t \in \mathcal{T}$ is a mapping from $\mathcal{F}(E, \mathcal{T})$ to $\mathcal{P}(E)$ given by (Serra, 1982):

$$\varpi_t(f) = \{p_i \in E |\, f(p_i) \geq t\}\,. \tag{10}$$

Using these transformations it is obvious that the binary dilation (6) can be computed using the numerical operator (7), i.e.,

$$\delta_B(X) = \varpi_1\left(\delta_B(\mathbb{1}_X)(p_i)\right). \tag{11}$$

By the fundamental property of dilation (Serra, 1982), given a set defined as the union of a family of sets, i.e., $X = \cup_{k\in K} X_k$, its corresponding binary dilation by B is

$$\delta_B(X) = \delta_B\left(\cup_{k\in K} X_k\right) = \cup_{k\in K}\delta_B(X_k). \tag{12}$$

From (11) and the property of the indicator function for the union of sets, it is easy to see that (12) can be rewritten using the dilation of functions as

$$\delta_B(X) = \varpi_1\left(\delta_B\left(\min\left\{1, \sum_{k\in K}\mathbb{1}_{X_k}(p_i)\right\}\right)\right) = \varpi_1\left(\min\left\{1, \sum_{k\in K}\delta_B\left(\mathbb{1}_{X_k}(p_i)\right)\right\}\right),$$

and finally it is obtained that

$$\delta_B(X) = \varpi_1\left(\sum_{k\in K}\delta_B\left(\mathbb{1}_{X_k}(p_i)\right)\right). \tag{13}$$

Hence, the justification for using NMF-based part decompositions in sparse mathematical morphology arises from equations (12) and (13).

Coming back to the NMF reconstruction, expressions (3) and (4), we can now write

$$X_j \approx \widetilde{X}_j = \varpi_\alpha \left(\sum_{k=1}^{R} \phi_k(p_i) \mathbf{H}_{k,j} \right). \tag{14}$$

Hence, we introduce the following nonlinear operator, named *sparse max-approximation to binary dilation*:

$$D_B(X_j) = \varpi_\alpha \left(\sum_{k=1}^{R} \delta_B\left(\phi_k\right)(p_i) \mathbf{H}_{k,j} \right). \tag{15}$$

Note that by the nonnegativity of $\mathbf{H}_{k,j}$, we have $\delta_B\left(\phi_k(p_i)\mathbf{H}_{k,j}\right) = \delta_B\left(\phi_k\right)(p_i)\mathbf{H}_{k,j}$. We can say that

$$\text{if} \quad X_j \approx \widetilde{X}_j \quad \text{then} \quad \delta_B(X_j) \approx D_B(X_j),$$

or in other words, the degree of approximation to the dilation depends on the degree of approximation to the original set. However neither the increasiness nor the extensitivity of $D_B(X_j)$ w.r.t. X_j can be guaranteed and consequently, operator (15) is not a morphological dilation in an algebraic sense (Heijmans and Ronse, 1990).

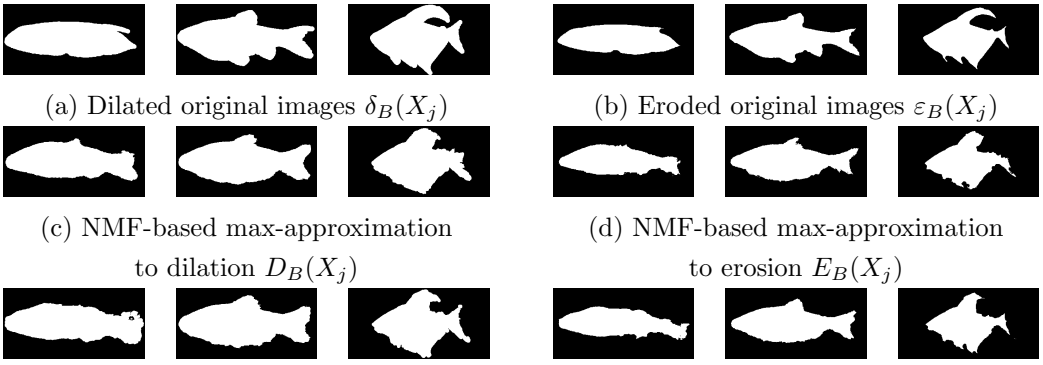
In conclusion, in order to approximate the dilation by B of any of the M sets X_j, we only need to calculate the dilation of the R basis images. In addition, if sparsity is imposed to $\mathbf{H}$, that involves that only a limited number of dilated atoms are required for each X_j.

Dual sparse max-approximation to binary erosion. One of the most interesting properties of mathematical morphology is the duality by the complement of pairs of operators, and in particular the duality between the dilation and the erosion. Hence, the binary erosion of set X by structuring element B can be defined as the dual operator to the dilation: $\varepsilon_{\check{B}}(X) = (\delta_B(X^c))^c$, where the complement set is $X^c = \complement X = E \setminus X$. Using this property, we propose to define the *sparse max-approximation to binary erosion* as

$$E_B(X_j) = \varpi_\alpha \left(\sum_{k=1}^{R} \complement \left[\delta_{\check{B}} \left(\complement[\phi_k] \right)(p_i) \right] \mathbf{H}_{k,j} \right), \tag{16}$$

where the complement basis images are defined by $\complement[\phi_k(p_i)] = \max(\mathbf{W}_{i,k}) - \phi_k(p_i) + \min(\mathbf{W}_{i,k})$.

The results of sparse max-approximations to $D_B(X_j)$ and $E_B(X_j)$ for three examples of the Fish shapes, compared to the exact binary dilation and erosion, are given in Fig. 5. We have compared in particular the sparse max-approximation for the standard NMF and for the Sparse-NMF.



(a) Dilated original images $\delta_B(X_j)$ (b) Eroded original images $\varepsilon_B(X_j)$

(c) NMF-based max-approximation to dilation $D_B(X_j)$ (d) NMF-based max-approximation to erosion $E_B(X_j)$

(e) Sparse NMF-based max-approximation to dilation $D_B(X_j)$ (f) Sparse NMF-based max-approximation to erosion $E_B(X_j)$

Figure 5: Comparison of dilation/erosion (a)/(b) vs. sparse pseudo operators for three examples of the Fish shapes. It is compared in particular the sparse max-approximation to dilation/erosion for the standard NMF (c)/(d) and for the Sparse-NMF (e)/(f), with ($S_w = 0$,$S_h = 0.6$). The structuring element B is a square of 5×5 pixels.

3.3 Sparse max-approximation to binary opening and closing

Composition of dilation and erosion produces two other operators, called opening and closing, which are the fundamental bricks for morphological filtering. More precisely, opening of a binary image X by B and the dual closing of a binary image X by B are defined as (Serra, 1982; Heijmans, 1994):

$$\gamma_B(X) = \delta_B(\varepsilon_B(X)) = \bigcup \{B(p_i) |\, p_i \in E \text{ and } B(p_i) \subseteq X\}, \tag{17}$$

$$\varphi_B(X) = \varepsilon_B(\delta_B(X)) = (\gamma_B(X^c))^c. \tag{18}$$

Similarly for the case of numerical functions $f \in \mathcal{F}(E, \mathcal{T})$, opening and closing of a grey-level image by B and its dual grey-level closing are respectively given by (Serra, 1982; Heijmans, 1994):

$$\gamma_B(f)(p_i) = \delta_B(\varepsilon_B(f))(p_i) = \sup_{p_m \in \breve{B}(p_n)} \inf_{p_n \in B(p_i)} \{f(p_m)\},$$

$$\varphi_B(f)(p_i) = \varepsilon_B(\delta_B(f))(p_i) = \inf_{p_m \in B(p_n)} \sup_{p_n \in \breve{B}(p_i)} \{f(p_m)\}.$$

In order to apply the non-negative decomposition underlaying the NMF coding, we first introduce sufficient conditions for a compatibility of opening with set decomposition, which are based on the geometric formulation of binary opening provided by (17). Let $X = \bigcup_{k \in K} X_k$ such that

- either $\partial X_i \cap \partial X_j = \emptyset$, $\forall i, j \in K$,

- or $X_i \subseteq X_j$ or $X_j \subseteq X_i$, $\forall i, j \in K$

then

$$\gamma_B(X) = \gamma_B\left(\bigcup_{k\in K} X_k\right) = \bigcup_{k\in K} \gamma_B(X_k). \tag{19}$$

Hence, opening commutes with the union only in the cases where the subsets are either totally disjoint or totally contained. We call these cases as separable subsets.

Under the assumption of separability of the atoms of the dictionary $\{\phi_k\}_{1\leq k\leq R}$, by combining (17) and (19) (equivalence only valid on the separable case), i.e.,

$$\gamma_B(X) = \bigcup_{k\in K}\bigcup \{B(p_i)|\, p_i \in E \text{ and } B(p_i) \subseteq X_k\} = \bigcup \left\{ B(p_i)|\, p_i \in E \text{ and } B(p_i) \subseteq \bigcup_{k\in K} X_k \right\},$$

and similarly to the dilation and erosion, we introduce the *sparse max-approximation to binary opening and to binary closing* respectively as

$$G_B(X_j) = \varpi_\alpha\left(\sum_{k=1}^{R} \gamma_B\left(\phi_k\right)(x_i)\mathbf{H}_{k,j}\right), \tag{20}$$

$$F_B(X_j) = \varpi_\alpha\left(\sum_{k=1}^{R} \complement\left[\gamma_B\left(\complement[\phi_k]\right)(p_i)\right]\mathbf{H}_{k,j}\right). \tag{21}$$

As discussed below, in Section 4, even in the case when the separability is not always fulfilled, operators $G_B(X_j)$ and $F_B(X_j)$ are interesting to approximate the effect of the opening $\gamma_B(X_j)$ and the closing $\varphi_B(X_j)$.

3.4 Consistency and noise robustness of sparse morphological operators

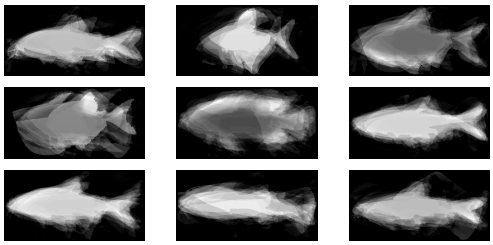
Let us discuss the two following important properties which have been empirically observed. From our viewpoint, they prove the pertinence of the sparse max-approximation to binary dilation based on Sparse-NMF representations.

Consistency. Behind this notion of consistency between Sparse-NMF and morphological binary dilation, we mean the fact that the Sparse-NMF basis from an image collection $\mathcal{X}$ should be stable to dilation of $\mathcal{X}$, i.e., $\delta_B(\mathcal{X}) = \{\delta_B(X_1), \cdots, \delta_B(X_M)\}$. Fig. 6 illustrates the consistency: it is compared in (a) the Sparse-NMF basis from the original image data set $\mathcal{X}$ to the Sparse-NMF basis obtained using the same parameters from two dilated image data set $\delta_B(\mathcal{X})$, in (b) for B a square of 5×5 and in (c) for a square 7×7. We have noted that, for a symmetric structuring element B of size relatively small with respect to the size of the objects, i.e., $|X_j \setminus \delta_B(X_j)|/|X_j| \ll 1$, $\forall j$, the consistency involves that $\sum_i \|\phi_k^{\oplus B} - \delta_B(\phi_k)\|(p_i)/\sum_i \phi_k(p_i) \approx 0$, $\forall k$, where

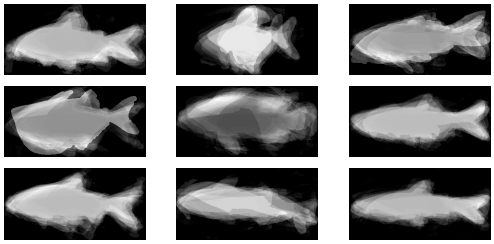
$$\mathcal{X} \xrightarrow{Sparse-NMF} \{\{\phi_k\}_{1\leq k\leq R};\, \mathbf{H}\} \Rightarrow \delta_B(\mathcal{X}) \xrightarrow{Sparse-NMF} \left\{\{\phi_k^{\oplus B}\}_{1\leq k\leq R};\, \mathbf{H}\right\}.$$

Using the semi-group property of multi-scale dilation δ_{nB} by homothetic convex structuring elements, i.e., $\delta_{nB} \circ \delta_{nB} = \delta_{(n+m)B}$, a consequence of such consistency is depicted in the example of Fig. 6(d): the sparse max-approximation to dilation of size 7 from $\mathcal{X}$ is close the sparse max-approximation to dilation of size 5 from a Sparse-NMF representation on $\delta_{2B}(\mathcal{X})$, i.e., $D_{5B}(\delta_{2B}(X_j)) \approx D_{7B}(X_j)$. More generally, given a convex symmetric structuring element B, if one has $|X_j \setminus \delta_{mB}(X_j)|/|X_j| \ll 1$ for a scale m, then we have

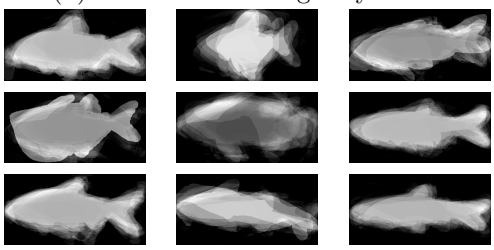
$$D_{(n+m)B}(X_j) \approx \varpi_\alpha \left(\sum_{k=1}^{R} \delta_{nB} \left(\phi_k^{\oplus mB} \right) (p_i) \mathbf{H}_{k,j} \right).$$



(a) From original images



(b) From dilated images by size 5



(c) From dilated images by size 9

Sparse NMF-based max-approximation to dilation $D_{7B}(X_j)$

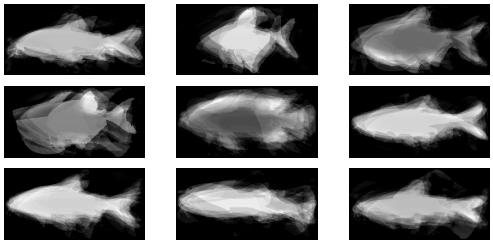
Sparse NMF-based max-approximation to dilation $D_{5B}(\delta_{2B}(X_j))$

(d)

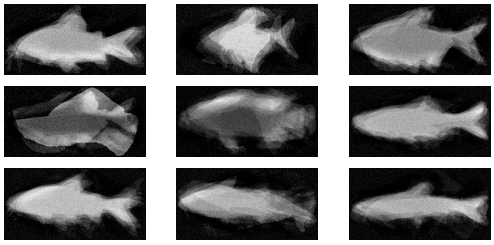
Figure 6: Consistency of Sparse-NMF representation for binary dilation: (a)-(c) Sparse-NMF basis images ($S_w = 0$,$S_h = 0.6$), $M = 100$ and $R = 10$; (d) max-approximation to dilation on original image collection vs. on dilated image collection.

Robustness against noise. The perturbation associated to $\mathcal{X}$ is now related to the

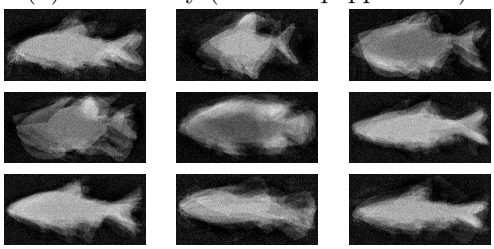
presence of noise; in particular, since we are dealing with binary images, salt-and-pepper noise is considered. Accordingly, a given percentage p of pixels is corrupted in our dataset to obtain $\mathcal{X}^{\text{noise}-p}$. Concerning the noise, it is well known that morphological operators are very sensitive to noise and even a small amount of salt-and-pepper provides a strong perturbation of dilation/erosion. As it is illustrated in Fig. 7, Sparse-NMF dictionary learning produces atoms or basis which are quite robust against this kind of noise. If we denote by $\phi_k^{\text{noise}-p}$ the basis obtained from noisy data $\mathcal{X}'_p$, we have observed that for p up to $20\% - 30\%$, we get $\sum_i \|\phi_k^{\text{noise}-p} - \phi_k\|(p_i)/\sum_i \phi_k(p_i) \approx 0$, $\forall k$, and therefore we obtain $D_B\left(X_j^{\text{noise}-p}\right) \approx D_B(X_j)$. Fig. 7(d) shows examples of the robustness of sparse max-approximation to dilation for $p = 10\%$ and $p = 20\%$. In conclusion, Sparse-NMF representation allows us (pseudo-)morphological processing noisy images.



(a) From original images



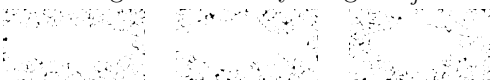
(b) From noisy (salt-and-pepper 10%)



(c) From noisy (salt-and-pepper 20%)

Original 10% noisy images X_j

Original 20% noisy images X_j



Dilated 10% noisy images $\delta_B(X_j)$

Dilated 20% noisy images $\delta_B(X_j)$

Sparse NMF-based max-approximation to dilation $D_B(X_j)$

Sparse NMF-based max-approximation to dilation $D_B(X_j)$

(d)

Figure 7: Robustness against noise of Sparse-NMF representation for binary dilation: (a)-(c) Sparse-NMF basis images ($S_w = 0$,$S_h = 0.6$), $M = 100$ and $R = 10$; (d) noisy binary images, dilated noisy images and max-approximation to dilation of noisy images.

4 Sparse approximation to numerical morphological operators

Approaches introduced in Section 3 are extended here to gray-levels images. In this case, we deal with families of discrete gray-level images, i.e., $\mathcal{F} = \{f_1(p_i), \cdots, f_M(p_i)\}$, with $f_j(p_i) \in \mathcal{F}(E, \mathcal{T})$, $\mathcal{T} = \{t_1, t_2, \cdots, t_L\}$ with $(t_{l+1} - t_l) = \Delta t$. Fig. 8 provides some examples of ORL face database (Guo et al., 2000) which is used to illustrate and compare our techniques. We consider in particular two alternative paradigms: i) each gray-scale function f_j is represented as a stack of upper level sets and Sparse-NMF processing is applied on the upper level sets, followed by image recomposition from processed upper level sets; ii) straightforward Sparse-NMF representation and processing of gray-scale images

Figure 8: Examples of original numerical images from the ORL face database.

4.1 Sparse-NMF processing of upper level sets

The thresholded set of f_j at each t_l, i.e., $X_j^{t_l} = \varpi_{t_l}(f_j)$, is called the upper level-set at t_l of f_j. The set of upper level sets constitutes a family of decreasing sets (Serra, 1982):

$$t_\lambda \geq t_\mu \Rightarrow X^{t_\lambda} \subseteq X^{t_\mu} \quad \text{and} \quad X^{t_\lambda} = \cap\{X^{t_\mu}, \mu < \lambda\}.$$

Moreover, any (semi-continuous) image f_j can be viewed as an unique stack of its upper level sets, which leads to the following reconstruction property (Serra, 1982):

$$f_j(p_i) = \sup\{t_l |\, p_i \in X_j^{t_l}\}, \quad t_l \in \mathcal{T}.$$

We prefer here to consider the alternative reconstruction using a numerical sum of the indicator function of upper level sets (Wendt et al., 1986; Ronse, 2009):

$$f_j(p_i) = \Delta t \sum_{l=1}^{L} \mathbb{1}_{X_j^{t_l}}(p_i). \tag{22}$$

It is well known in mathematical morphology that any binary increasing operator, such as the dilation and erosion, can be generalized to gray-level images by applying the binary operator to each cross-section, and then by reconstructing the corresponding gray-level image (Serra, 1982; Ronse, 2009), i.e.,

$$\delta_B(f_j)(p_i) = \Delta t \sum_{l=1}^{L} \mathbb{1}_{\delta_B\left(X_j^{t_l}\right)}(p_i), \tag{23}$$

$$\varepsilon_B(f_j)(p_i) = \Delta t \sum_{l=1}^{L} \mathbb{1}_{\varepsilon_B\left(X_j^{t_l}\right)}(p_i). \tag{24}$$

Consider now that each image of the initial gray-level family $\mathcal{F}$ of M images is decomposed into its L upper level set. Hence, we have

$$\mathcal{F} = \{f_1, \cdots, f_M\} \mapsto \mathcal{X} = \{X_1^{t_1}, X_1^{t_2} \cdots, X_1^{t_L}, \cdots, X_{M-1}^{t_L}, X_M^{t_1}, \cdots, X_M^{t_L}\},$$

where $\mathcal{X}$ is a family of $M' = M \times L$ binary images. Therefore, we can use NMF algorithms, for a given dimension R, to approximate each set $X_j^{t_l}$ and then approximate the corresponding function $f_j(p_i)$. Thus, using the results of the previous Section, we are able now to introduce the following definition for the *sparse max-approximation to grey-level dilation and erosion* given respectively by:

$$D_B(f_j)(p_i) = \Delta t \sum_{l=1}^{L} \mathbb{1}_{D_B\left(X_j^{t_l}\right)}(p_i), \tag{25}$$

and

$$E_B(f_j)(p_i) = \Delta t \sum_{l=1}^{L} \mathbb{1}_{E_B\left(X_j^{t_l}\right)}(p_i). \tag{26}$$

with

$$D_B\left(X_j^{t_l}\right) = \varpi_\alpha\left(\sum_{k=1}^{R} \delta\left(\phi_k\right)(p_i)\mathbf{H}_{k,j+l}\right),$$

$$E_B\left(X_j^{t_l}\right) = \varpi_\alpha\left(\sum_{k=1}^{R} \mathbb{C}\left[\delta_B\left(\mathbb{C}[\phi_k]\right)(p_i)\right]\mathbf{H}_{k,j+l}\right).$$

Similarly, a sparse max-approximation to gray-scale opening and closing can be achieved by means of the expressions (20) and (21) for respectively $G_B(X_j)$ and $F_B(X_j)$ applied to the upper level sets of f_j.

Behavior of sparse max-approximation to gray-level dilation and erosion based on upper level set decomposition is shown in Fig. 9. Note that, for this example, the number of initial images $M = 20$, and each of them has been quantized in $L = 10$ gray-levels, i.e., $M' = 20 \times 10 = 200$ dimensions and them for Sparse-NMF the number of atoms is fixed to $R = 75$. Hence a reduction factor which is not significant. Some of the corresponding atoms are given in Fig. 9(a). As expected the quality of the sparse max-approximation to dilation and to erosion depends on the quality of the initial Sparse-NMF reconstruction of the image. For instance, the first face (man with glasses) is not well approximated with the learned NMF basis and hence, its approximated dilation and erosion are also unsatisfactory. On the contrary, in the case of the last image (woman), the results are more relevant. Nevertheless, at this point, we can conclude that the sparse approximation using upper level set stack, which is theoretically consistent with the binary framework discussed in Section 3 is not very useful in practice.

4.2 Sparse-NMF representation and processing of gray-scale images

As an alternative to the previous formulation, we propose a straightforward use of Sparse-NMF representation from the image gray-scale image collection $\mathcal{F} = \{f_1(p_i), \cdots, f_M(p_i)\}$. That is, each column of the data matrix $\mathbf{V}$ corresponds to an image, i.e., $V_{i,j} = f_j(p_i)$. Then, after Sparse-NMF dictionary learning of dimension R, each image is approximated as

$$f_j(p_i) \xrightarrow{Sparse-NMF} \widetilde{f}_j(p_i) = \sum_{k=1}^{R} \phi_k(p_i)\mathbf{H}_{k,j}.$$

According to our principle, for any image f_j of the family $\mathcal{F}$ and given a structuring element B, we can introduce the following operators.

The *sparse max-approximation to gray-level dilation and to gray-level erosion* are respectively given by

$$D_B(f_j)(x_i) = \sum_{k=1}^{R} \delta_B(\phi_k)(x_i)\mathbf{H}_{k,j}, \tag{27}$$

$$E_B(f_j)(x_i) = \sum_{k=1}^{R} \mathsf{C}\left[\delta_B\left(\mathsf{C}[\phi_k]\right)(x_i)\right]\mathbf{H}_{k,j}. \tag{28}$$

In addition, the *sparse max-approximation to gray-level opening and gray-level closing* are respectively defined as

$$G_B(f_j)(x_i) = \sum_{k=1}^{R} \gamma_B(\phi_k)(x_i)\mathbf{H}_{k,j}, \tag{29}$$

$$F_B(f_j)(x_i) = \sum_{k=1}^{R} \mathsf{C}\left[\gamma_B\left(\mathsf{C}[\phi_k]\right)(x_i)\right]\mathbf{H}_{k,j}. \tag{30}$$

We must point out again that these approximate nonlinear operators do not satisfy the standard properties of grey-level dilation and erosion.

Similarly to the binary case, let us discuss the conditions under which these max-approximations to dilation and opening are more relevant. First, let us rewrite (27) as supremum of cylinders (Heijmans, 1994; Bloch et al., 2007):

$$\delta_B(f) = \bigvee \left\{ C_{B(p_i),\, f(p_i)} \ : \ p_i \in E \right\},$$

where $C_{B,t}$ is the cylinder of base B at the origin and height t. From this expression, it is easy to see that dilation commutes with supremum of functions, i.e.,

$$\begin{aligned}
\delta_B \left(\bigvee_{k \in K} f_k \right) &= \bigvee \left\{ C_{B(p_i),\, \bigvee_{k \in K} f_k(p_i)} \ : \ p_i \in E \right\} \\
&= \bigvee \left\{ \bigvee_{k \in K} C_{B(p_i),\, f_k(p_i)} \ : \ p_i \in E \right\} \\
&= \bigvee_{k \in K} \bigvee \left\{ C_{B(p_i),\, f_k(p_i)} \ : \ p_i \in E \right\} = \bigvee_{k \in K} \delta_B \left(f_k \right)
\end{aligned}$$

For the case of the opening (29), one has (Heijmans, 1994; Bloch et al., 2007):

$$\gamma_B(f) = \bigvee \left\{ C_{B(p_i),\, t} \ : \ p_i \in E,\, C_{B(p_i),\, t} \leq f \right\}.$$

Thus, a sufficient condition for the commutation of opening with the supremum is the "disjointness" of functions, i.e., $\bigwedge_{k \in K} f_k = 0$. It is obvious also that for "disjoint functions", one has: $\bigvee_{k \in K} f_k = \sum_{k \in K} f_k$. Therefore, operators (27) and (29) will provide a good approximation to dilation and opening in the case where the Sparse-NMF representation is composed of disjoint atoms, i.e.,

$$\bigwedge \phi_k(p_i) \approx 0,$$

which turns out to be a condition of non-redundancy of the Sparse-NMF basis.

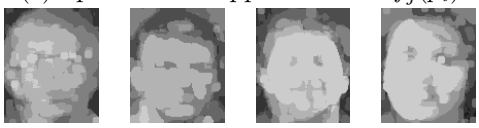
Fig. 10(a) gives the Sparse-NMF representation of the faces, where a collection of $M = 80$ images has been used in this experiment and dimension of dictionary has been fixed to $R = 20$ atoms. As we can observe, the effectivity of the max-approximation to dilation and opening is consistent with the approximation to the corresponding original image. Thus, such approximations can be then used to compute evolved morphological operators. We have in particular illustrated in Fig. 10 two operators which can be applied in face image feature extraction. The first one is the morphological gradient, i.e., $\delta_(B)(f) - \varepsilon_(B)(f)$, typically used for contour detection. The second one, the black top-hat, i.e., $\varphi_(B)(f) - f$, extracts the dark structures, which in the present case of faces, correspond just to the eyes and lips.

(a) Examples from $\{\phi\}_{1\leq k\leq 75}$

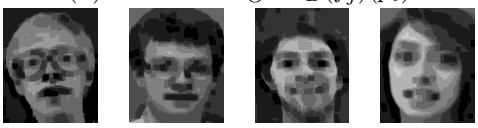
(b) Original images $f_j(p_i)$

(c) Sparse-NMF approximation $\widetilde{f}_j(p_i)$

(d) Dilated images $\delta_B(f_j)(p_i)$



(e) Sparse max-approximation dilat. $D_B(f_j)(p_i)$



(f) Eroded images $\varepsilon_B(f_j)(p_i)$

(g) Sparse max-approximation erod. $E_B(f_j)(p_i)$

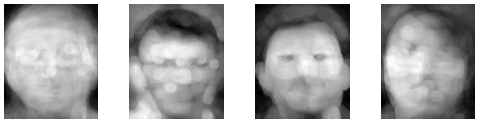
Figure 9: A collection of $M = 20$ faces has been used in this experiment, where the number of reduced dimensions for the binary matrix $\mathbf{V}$ has been fixed to $R = 75$ (note that $M' = 20 \times 10 = 200$ dimensions). Top, some examples of the Sparse-NMF atoms obtained from the 200 upper level sets. Bottom, four examples of the ORL face database (b) (quantized in $L = 10$ gray-levels) and (c) corresponding approximated images using Sparse-NMF. Comparison of dilation/erosion (d)/(f) vs. sparse max-approximation to dilation/erosion for the Sparse-NMF (e)/(g). The structuring element B is a square of size 3×3 pixels.

(a) Sparse-NMF basis $\{\phi\}_{1\leq k\leq 20}$ ($S_w = 0$, $S_h = 0.6$)

(b) Original images $f_j(p_i)$

(c) Sparse-NMF approximation $\widetilde{f}_j(p_i)$

(d) Dilated images $\delta_B(f_j)(p_i)$

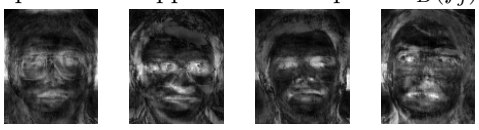


(e) Sparse max-approximation dilat. $D_B(f_j)(p_i)$

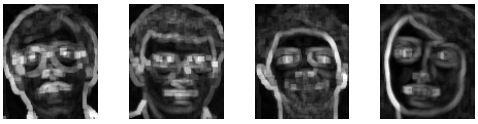
(f) Opened images $\gamma_B(f_j)(p_i)$

(g) Sparse max-approximation open. $G_B(f_j)(p_i)$

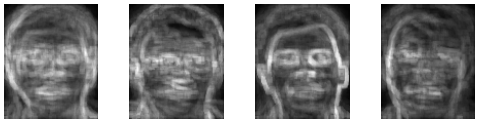
(h) Black top-hat images $\varphi_B(f_j)(p_i) - f_j(p_i)$



(i) Sparse max-approximation black top-hat $F_B(f_j)(p_i) - \widetilde{f}_j(p_i)$



(j) Morphological gradient images $\delta_B(f_j)(x_i) - \varepsilon_B(f_j)(p_i)$



(k) Sparse max-approximation morpho. gradient $D_B(f_j)(x_i) - E_B(f_j)(p_i)$

Figure 10: Sparse-NMF representation and processing of gray-scale images $\mathcal{F}$: A collection of $M = 80$ faces has been used in this experiment and dimension of dictionary has been fixed to $R = 20$. Top, the obtained Sparse-NMF atoms. Bottom, four examples of the ORL face database (b) and (c) corresponding approximated image using Sparse-NMF. Comparison of dilation, opening, black top-hat and gradient (d)/(f)/(h)/(j) vs. sparse max-approximation to this operators. The structuring element B is a square of size 5×5 pixels for the dilation and opening, size 9×9 for the black top-hat.

5 Applications

We consider in this Section to potential applications of Sparse-NMF morphological processing to problems arising from multi/hyperspectral image processing.

5.1 Sparse processing of multivariate Boolean textures

The motivation of this study is to deal with random image models on multivalued images, and more precisely with the case of multivariate Boolean random set model (Jeulin, 1991). This approach is an extension of the classical Boolean random model (Matheron, 1975). Multivariate texture images which fit with this model appear in different fields: multi/hyperspectral images, energy dispersive spectrometry, electron energy-loss spectrometry, etc.

Let us consider a Poisson point process in a set $E \subset \mathbb{R}^n$ with intensity (average number of point per unit of volume) θ. In each point $x_k \in E$ of the process, an independent realization of a multivariate random compact set $\mathbf{A} = (A_1, A_2, \cdots, A_d)$ is implanted: a realization of the primary grain sets A_i is placed at x_k. Each component X_i of the multivariate Boolean random set $\mathbf{X} = (X_1, X_2, \cdots, X_d)$, $\mathbf{X} \sim (\theta, \mathbf{A})$ is a Boolean random closed set obtained as $X_i = \bigcup_k A_i(x_k)$. An example of 2D multivariate image of 20 components $\mathbf{X} = (X_1, X_2, \cdots, X_{20})$, $X_i \in \mathcal{P}(E)$, $E = [0, 255] \times [0, 255]$, $\theta|E| = 30$, and where A_i are random disks is given in Fig. 11. In practical situations, this kind of binary textures can be for instance obtained after segmentation of multi/hyperspectral images.

In the framework of the theory of random sets, these structures and their models are fully characterized by a functional called Choquet capacity (Matheron, 1975). In the present model, the multivariate Choquet capacity $T_{\mathbf{X}}(\mathbf{K})$, with $\mathbf{K} = (K_1, K_2, \cdots, K_d)$, $K_i \in \mathcal{K}$ (compact set), is defined as (Jeulin, 1991):

$$T_{\mathbf{X}}(\mathbf{K}) = 1 - Q_{\mathbf{X}}(\mathbf{K}) = 1 - \exp\left(-\theta \overline{\mu}_n \left(\mathbf{A} \oplus_{\cup} \check{\mathbf{K}}\right)\right),$$

with

$$\mathbf{A} \oplus_{\cup} \check{\mathbf{K}} = \left\{A_1 \oplus \check{K}_1 \cup A_2 \oplus \check{K}_2 \cup \cdots A_d \oplus \check{K}_d\right\},$$

$\overline{\mu}_n$ being the Lebesgue measure in $\mathbb{R}^n$ and where $X \oplus \check{K}$ is the dilation of set $X \in \mathcal{P}(E)$ by structuring element K, i.e.,

$$\delta_K(X) = X \oplus \check{K} = \bigcup_{p_i \in \check{K}} X(p_i),$$

where $\check{K} = \{-p_i : p_i \in K\}$. Hence, an experimental estimation of $T_{\mathbf{X}}(\mathbf{K})$ can be obtained by morphological dilations on realizations of components of $\mathbf{K}$. If a single grain is used for each Poisson point, the components are independent Boolean random sets (Matheron, 1975; Jeulin, 1991); otherwise some correlations between components are present.

As a fundamental property, this model is stable by the union; such that any number of components (even correlated) of a Boolean multivariate random set is still a Boolean

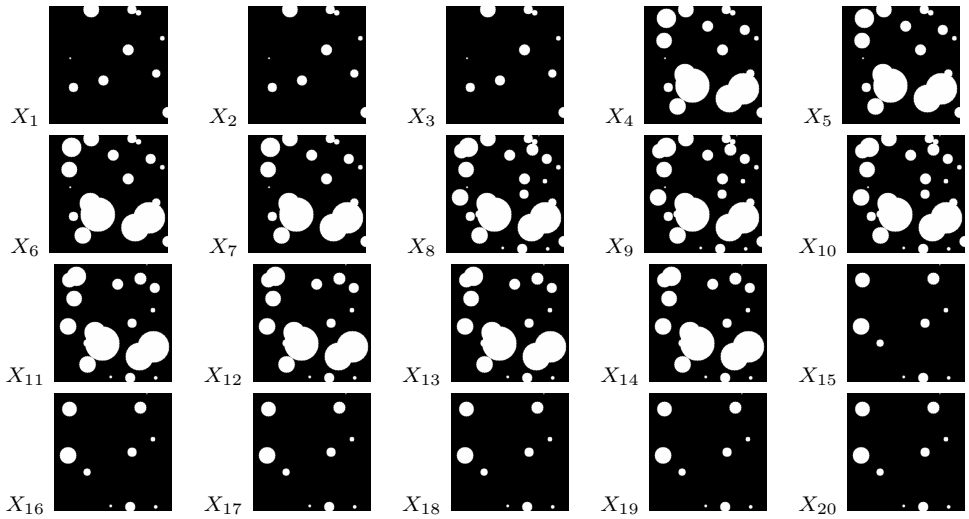


Figure 11: Example of 2D multivariate image of 20 components $\mathbf{X} = (X_1, X_2, \cdots, X_{20})$, $X_i \in \mathcal{P}(E)$, $E = [0, 255] \times [0, 255]$, $\theta|E| = 30$, A_i are random disks.

random set. Hence a notion of dimensionally reduction by supremum of primary schemes is compatible with the model. In our current example of Fig. 11, the A_i random disks have been simulated using three vector prototypes, see Fig. 12(b) which consequently involves correlations between the components.

Using our theory of Sparse-NMF representation and max-approximation to dilation we note that, on the one hand, the Sparse-NMF basis $\{\phi_k(p_j)\}$ gives just a dimensionality reduction of the initial multivariate Boolean model. In other words, thresholding the four components of Fig. 12(a) produces a Boolean model of dimension 4 instead of dimension 20 of the original model. We observe in the coding vectors $\mathbf{H}$ associated to basis, Fig. 12(c), that the four components are consistent with the decomposition of the original spectra Fig. 12(d). On the other hand, approximated dilations to the initial components of the model can be naturally obtained using expression (15), Fig. 12(d)-(f).

In summary, in order to model a multispectral Boolean set $\mathbf{X}$, we need to compute the characteristic curves related to Choquet capacity $T_{\mathbf{X}}(\mathbf{K})$, for different $\mathbf{K}$, and these computations can be efficiently approximated in a consistent theoretical way with our Sparse-NMF dilations.

5.2 Sparse processing of hyperspectral images

High-spatial resolution hyperspectral images are used nowadays in remote sensing and other application domains. They constitute high dimensional datasets: for instance, the Pavia image, used in this case-study, has dimensions $N = 340 \times 610$ and $M = 103$. As we have discussed in the introduction, NMF is widely used in hyperspectral imaging since the non-

negative representation is compatible with linear physical model of endmembers/abundances.

Fig. 13 shows some of the spectral bands of Pavia image in (a) as well as the classical NMF representation compared to the Sparse-NMF counterpart (with $S_w = 0$ and $S_h = 0.5$), where for both cases $R = 5$. We have assessed the effect of the sparseness parameter S_h on the approximation error (in average percentage) to morphological operators: max-approximation to dilation D_B (27), to erosion E_B (28) and to opening G_B (29), for the the $M = 103$ spectral bands. From this example we clearly observe that with $S_h = 0.5$ the approximation error is much better than without any sparseness. We note also that a low value of S_h can eventually produces worst results than unconstrained NMF.

Mathematical morphology is widely used in remote sensing hyperspectral imaging. One of the most popular applications in the state-of-the-art is the spectral-spatial classification based on the notion of morphological profile (Pesaresi and Benediktsson, 2001), and its various extensions to hyperspectral images (Fauvel et al., 2008; Velasco-Forero and Angulo, 2013). Morphological profiles are founded on the granulometry (or opening-based scale-space). We remind that a granulometry (Serra, 1982) is a family of openings $\{\gamma_{B_n}\}_{0 \leq n \leq L}$, depending on a (discrete) positive parameter n, representing scale of homothetic structuring element, i.e., $B_n = nB$. Then, given an image f, it produces a multi-scale decomposition of bright structures:

$$s_n(f)(p_j) = f(p_j) - \gamma_{B_n}(f)(p_j),$$

or using a differential representation:

$$r_n(f)(p_j) = \gamma_{B_{n-1}}(f)(p_j) - \gamma_{B_n}(f)(p_j)$$

such that:

$$f(p_j) = \gamma_{B_L}(f)(p_j) + \sum_{n=1}^{L} r_n(f)(p_j).$$

Then, this multi-scale decomposition can be used either for tensor-based spatial-spectral dimensionality reduction (Velasco-Forero and Angulo, 2013) (4D tensor = 2D space $\times$ 1D spectrum $\times$ 1D morphology):

$$\left\{ \{f_\lambda\}_{1 \leq \lambda \leq M}, \{\gamma_{B_1}(f_\lambda)\}_{1 \leq \lambda \leq M} \cdots, \{\gamma_{B_L}(f_\lambda)\}_{1 \leq \lambda \leq M} \right\};$$

or for spatial-spectral classification, where the feature vector per pixel p_j given by:

$$\left(f_1(p_j), f_2(p_j), \cdots, f_M(p_j), \gamma_{B_1}(f_1)(p_j), \gamma_{B_1}(f_2)(p_j), \cdots, \gamma_{B_L}(f_M)(p_j)\right),$$

is the so-called morphological profile at p_j. As it is shown in Fig. 15 for two examples of spectral bands, the max-approximation to opening-based scale-space using Sparse-NMF representation is quite satisfactory and can be naturally used in morphological profile-based spectral spatial classification.

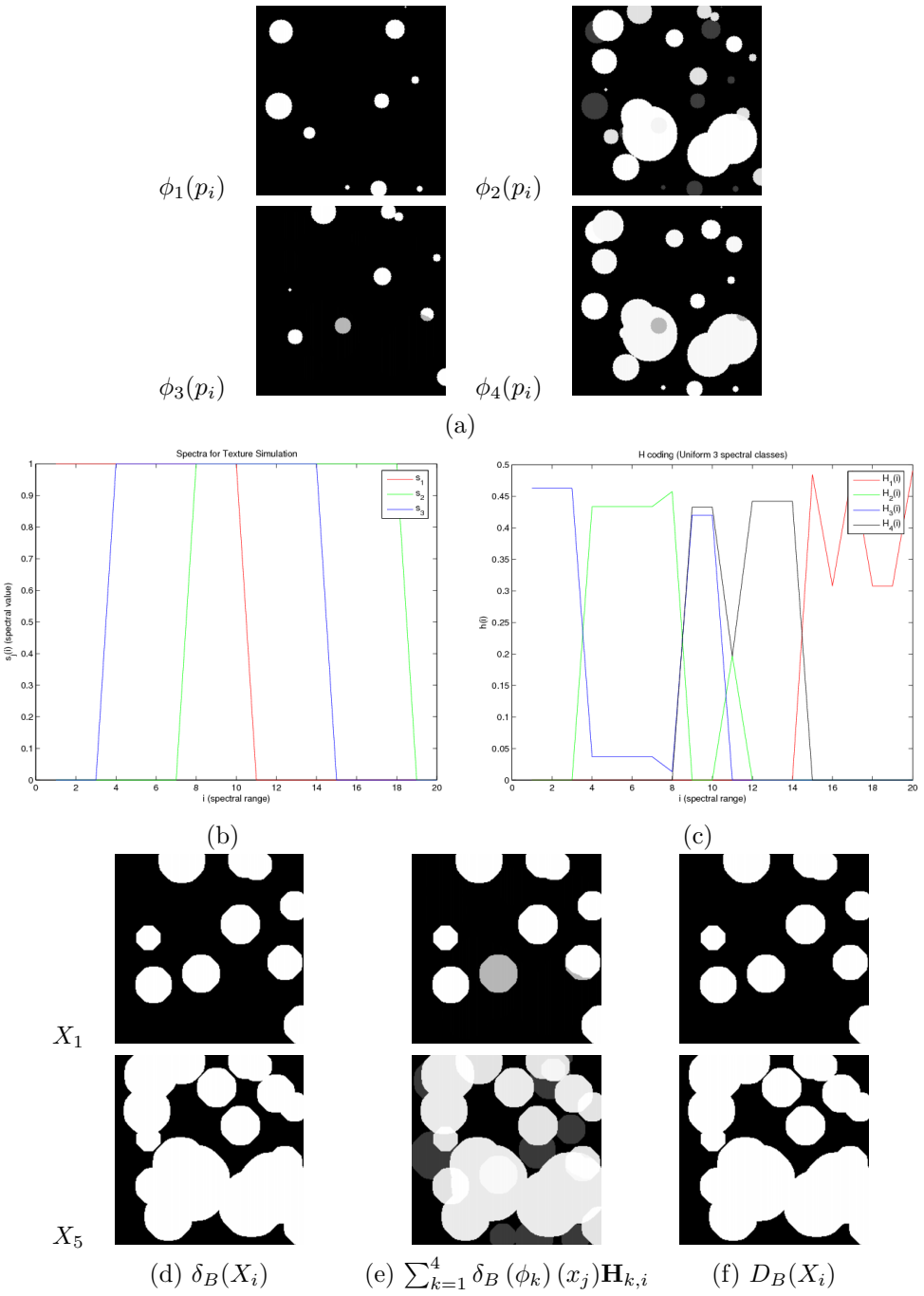
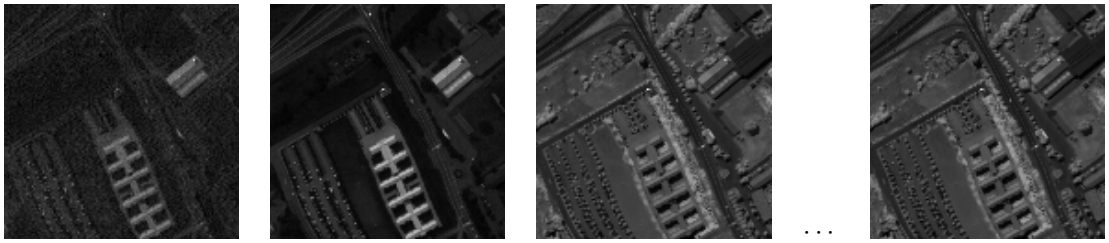
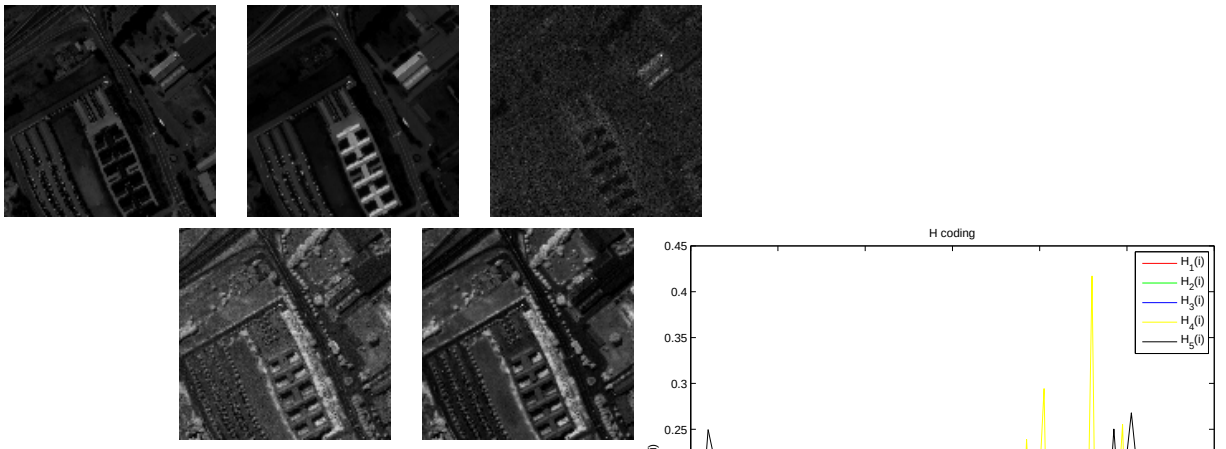


Figure 12: Dimensionality reduction of multivariate Boolean set $\mathbf{X}$ from Fig. 11: (a) Sparse-NMF basis with $R = 4$; (b) three spectra used to simulate multivariate set $\mathbf{X}$; (c) coding vectors $\mathbf{H}$ associated to basis of (a); (d) two dilated components of $\mathbf{X}$; (e) approximated dilation using the Sparse-NMF basis; (d) thresholded images.

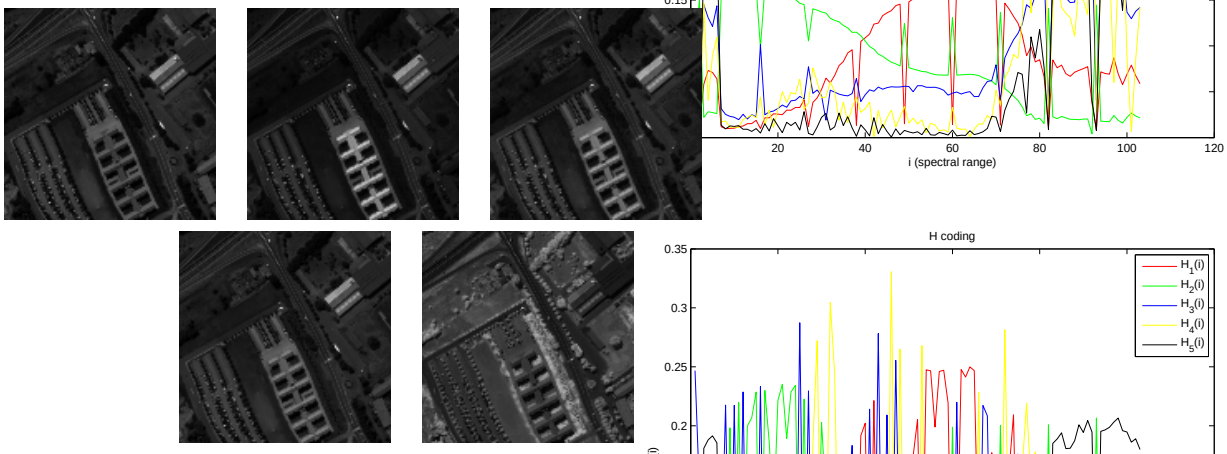


$f_1(p_j)$ $f_2(p_j)$ $f_3(p_j)$ ... $f_{103}(p_j)$

(a)



(b) NMF representation



(c) Sparse-NMF representation ($S_w = 0$, $S_h = 0.5$)

Figure 13: Pavia hypersectral images (crop of a part of the whole image): (a) some of the spectral bands, (b) classical NMF representation, (c) Sparse-NMF representation ($S_w = 0$, $S_h = 0.5$). In both cases, left, basis $\{\phi_k(p_j)\}_{1 \leq k \leq 5}$ and right, coding $\mathbf{H}$.

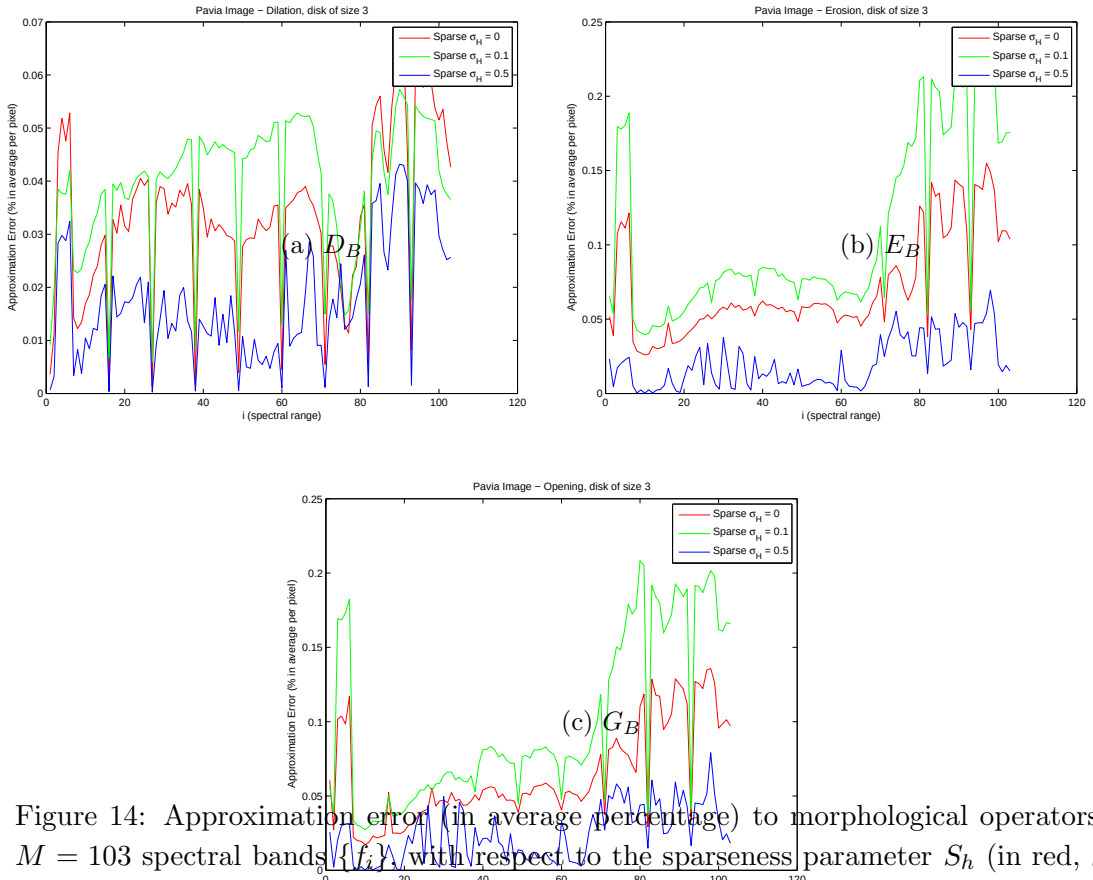
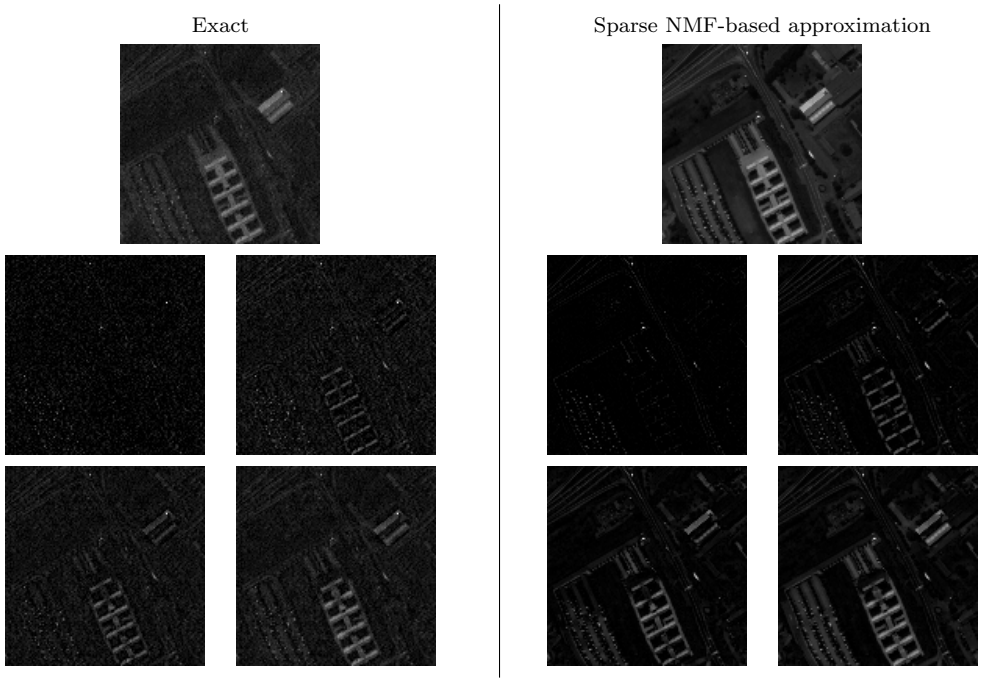
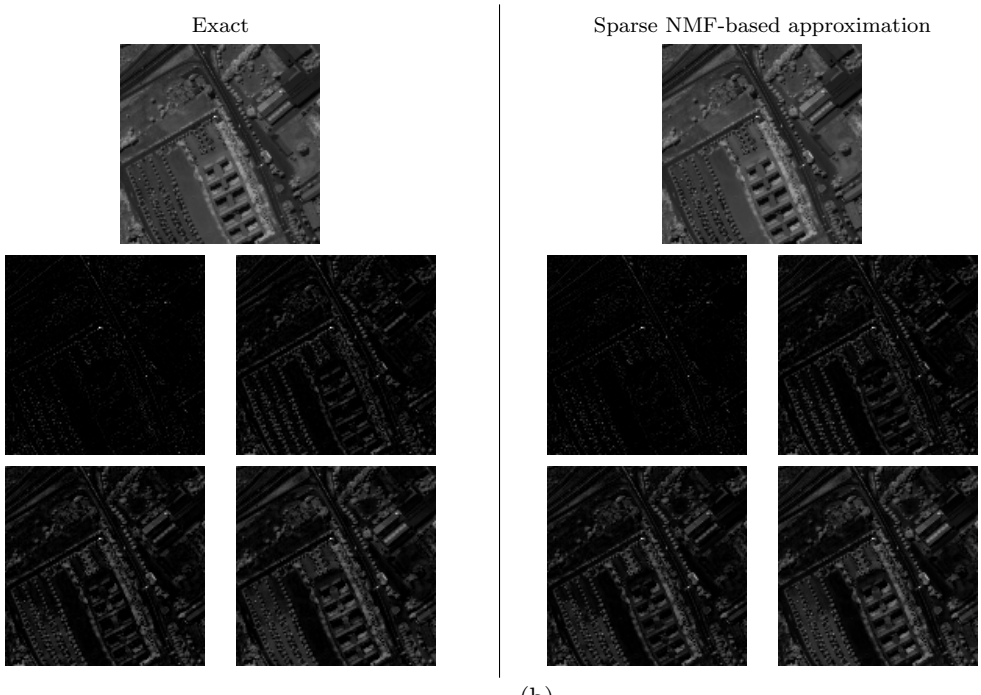


Figure 14: Approximation error (in average percentage) to morphological operators of the $M = 103$ spectral bands $\{f_i\}$ with respect to the sparseness parameter S_h (in red, $S_h = 0$; in green, $S_h = 0.1$ and in blue, $S_h = 0.5$): (a) max-approximation to dilation D_B, (b) max-approximation to erosion E_B, (b) max-approximation to opening G_B. For all the cases B is a square 3×3.



(a)



(b)

Figure 15: Comparison of opening-based scale-space $f_i(x)$, $s_1(f_i)(x)$, $s_3(f_i)(x)$, $s_5(f_i)(x)$, $s_7(f_i)(x)$ for exact openings and for Sparse-NMF max-approximations ($R = 5$, $S_h = 0.5$). In (a) for spectral band $f_1(p_j)$ and in (b) for $f_3(p_j)$.

6 Conclusions and perspectives

We have introduced the notion of sparse binary and gray-level max-approximation to morphological operators based on Sparse-NMF representation.

We have linked the role of non-negative decompositions of sets of functions, viewed as vectors in a high dimensional space, to the commutation by supremum of the dilation operator, which is just the counterpart of convolution in max-plus algebra. We have also seen that in the case of the opening operator, additional conditions of separability (or disjointness) are required for an exact compatibility. Empirical examples showed that this assumption can be relaxed since the obtained approximations are satisfactory in practical cases.

We have illustrated the practical interest of our approach for morphological processing of multivariate vector images, namely multispectral/hyperspectral images. Our results are relatively encouraging and they open a new avenue to study how the current paradigm of sparse modeling (based mainly on linear operations) in computer vision can be extended to the nonlinear morphological framework.

As we have discussed, NMF produces linear non-negative decompositions which are well indicated for morphological operators in the case of separable functions, i.e., decomposition by sum is equivalent to decomposition by max.

It is well known that morphological operators are linear in the max-plus algebra. Consequently, the intrinsic linear decomposition in such algebra is the more appropriate one to introduced sparse morphological processing, without separability conditions. Some recent work has addressed the problem of matrix factorization in max-plus algebra, see for instance (Hook, 2014), which is based on some classical results on max-plus spectral theory and max-plus eigenvalues as the asymptotic behavior of standard eigenvalues by means of a nonlinearization of the matrix (Gaubert et al., 2001; Schutter and De Moor, 2002). Other recent results on matrix factorization over max-times algebra (Karaev, 2013) can also be used as a representation approach for morphological operators. To explore the interest of decompositions on max-plus and max-times algebras for morphological processing will be the object of future research.

References

J. Angulo, S. Velasco-Forero. Sparse Mathematical Morphology using Non-Negative Matrix Factorization. In *Proc. of 2011 International Symposium on Mathematical Morphology (ISMM'11)*, LNCS 6671, Springer, p. 1–12, 2011.

S. Arora, R. Ge, R. Kannan, A. Moitra. Computing a nonnegative matrix factorization - provably. In *Proc. of the 44th Symposium on Theory of Computing,* 145–162, 2012.

I. Bloch, H. Heijmans, C. Ronse. Mathematical Morphology. *In (M. Aiello, I. Pratt-Hartmann, J. van Benthem, eds.) Handbook of Spatial Logics,* Springer, Chapter 14, pp. 857–944, 2007.

D. Cai, X. He, J. Han, T.S. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 33(8): 1548–1560, 2011.

A. Cichocki, R. Zdunek, A.H. Phan, S.I. Amari. Nonnegative matrix and tensor factorizations. John Wiley & Sons, 2009.

A. Damle, Y. Sun. Random projections for non-negative matrix factorization. *arXiv preprint arXiv:1405.4275,* 2014.

D. Donoho, V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts? In *Advances in Neural Information Processing 16 (Proc. NIPS'03),* MIT Press, 2004.

D. Donoho. Compressed sensing, *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

C. Ding, X. He, H.D. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proc. of SIAM Data Mining Conference (SDM'05),* 606–610, 2005.

M. Elad, M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. on Image Proc.,* 15(12):3736–3745, 2006.

E. Esser, M. Möller, S. Osher, G. Sapiro, J. Xin. A convex model for nonnegative matrix factorization and dimensionality reduction on physical space. *IEEE Trans. on Image Processing,* 21(7):3239–3252, 2012.

M. Fauvel, J.A. Benediktsson, J. Chanussot, J.R. Sveinsson. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Trans. Geoscience and Remote Sensing,* 46(11):3804–3814.

S. Gaubert, M. Akian, R. Bapat. Generic asymptotics of eigenvalues using min-plus algebra. In *Proc. of the Satellite Workshop on Max-Plus Algebras, IFAC SSSC01,* Elsevier, 2001.

G. Guo, Z. Li Stan, C. Kapluk. Face recognition by support vector machines. *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.

N. Guan, D. Tao, Z. Luo, B. Yuan. Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent. *IEEE Trans. on Image Processing,* 20(7): 2030–2048, 2011.

H.J.A.M. Heijmans, Ch. Ronse. The algebraic basis of mathematical morphology I. Dilations and erosions. *Computer Vision, Graphics, and Image Processing*, Vol 50(3), 245–295, 1990.

H.J.A.M. Heijmans. *Morphological Image Operators,* Academic Press, Boston, 1994.

J. Hook. Max-plus singular values. *MIMS EPrint 2014.7, The University of Manchester, UK,* 2014.

K. Huang, N. Sidiropoulos, A. Swami. Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE Trans. on Signal Processing,* 62(1):211–224, 2014.

P. Hoyer. Non-negative Matrix Factorization with Sparseness Constraints. *J. Mach. Learn. Res.,* 5: 1457-1469, 2004.

D. Jeulin. Multivariate Random Image Models. *Acta Stereologica,* 11: 59–66, 1991.

S. Karaev. Matrix Factorization over Max-Times Algebra for Data Mining. *Master's Thesis (supervisor P. Miettinen), Universität des Saarlandes,* Saarbrücken, Germany, 2013.

A. Kumar, V. Sindhwani, P. Kambadur. Fast conical hull algorithms for near-separable non-negative matrix factorization. In *Proc. of the 30th International Conference on Machine Learning,* 2013.

D.D. Lee, H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature,* 401(6755):788–791, 1999.

D.D. Lee, H.S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing 13 (Proc. NIPS'00),* MIT Press, 2001.

S.Z. Li, X. Hou, H. Zhang, Q. Cheng. Learning spatially localized parts-based representations. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR),* Vol. I, 207–212, Hawaii, USA, 2001.

J. Mairal, M. Elad, G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. on Image Processing,* 17(1):53–69, 2008.

J. Mairal, F. Bach, J. Ponce. Sparse Modeling for Image and Vision Processing. *Foundations and Trends in Computer Graphics and Vision,* 8(2-3):85–283, 2012.

G. Matheron. *Random Sets and Integral Geometry.* Wiley, New York, 1975.

P. Miettinen, T. Mielikainen, A. Gionis, G. Das, H. Mannila. The discrete basis problem. *IEEE Trans. on Knowledge and Data Engineering,* 20(10):1348–1362, 2008.

P. Miettinen. Sparse Boolean matrix factorizations. In *Proc. of IEEE 10th International Conference on Data Mining (ICDM'10),* 935–940, 2010.

M. Pesaresi, J.A. Benediktsson. A new approach for the morphological segmentation of high resolution satellite imagery. *IEEE Trans. Geoscience and Remote Sensing,* 39(2):309–320, 2001.

M.D. Plumbley. Algorithms for nonnegative independent component analysis. *IEEE Trans. on Neural Networks,* 14(3):534–543, 2003.

Ch. Ronse. Bounded variation in posets, with applications in morphological image processing. *Acta Universitatis Upsaliensis (M. Passare, ed., Proceedings of the Kiselmanfest-2006)*, Vol. 86, 249–281, 2009.

R. Schachtner, G. Pöppel, E.W. Lang. A nonnegative blind source separation model for binary test data. *IEEE Trans. on Circuits and Systems I,* 57(7):1439–1448, 2010.

B. De Schutter, B. De Moor. The QR decomposition and the singular value decomposition in the symmetrized max-plus algebra revisited. *SIAM Review,* 44(3):417–454, 2002.

J. Serra. *Image Analysis and Mathematical Morphology. Vol I, and Image Analysis and Mathematical Morphology. Vol II: Theoretical Advances,* London: Academic Press, 1982,1988.

P. Soille. *Morphological Image Analysis,* Springer-Verlag, Berlin, 1999.

F. Theis, K. Stadlthanner, and T. Tanaka. First results on uniqueness of sparse non-negative matrix factorization. In *Proc of the 13th European Signal Processing Conference (EUSIPCO '05),* 1–4, 2005.

A.M. Tomé, R. Schachtner, V. Vigneron, C.G. Puntonet, E.W. Lang. A logistic non-negative matrix factorization approach to binary data sets. *Multidimensional Systems and Signal Processing,* 26(1): 125–143, 2015.

S.A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization,* 20:1364–1377, 2009.

S. Velasco-Forero, J. Angulo. Classification of hyperspectral images by tensor modeling and additive morphological decomposition. *Pattern Recognition,* 46(2):566–577, 2013.

P.D. Wendt, E.J. Coyle, N.C. Gallagher. Stack Filters. *IEEE Trans. on Acoustics, Speech, and Signal Processing,* 34(4):898–911, 1986.

G.Yu, G. Sapiro, S. Mallat. Solving Inverse Problems with Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity. *IEEE Trans. on Image Processing,* 2011.

Y. Yuan, X. Li, Y. Pang, X. Lu, D. Tao. Binary Sparse Nonnegative Matrix Factorization. *IEEE Trans. on Circuits and Systems for Video Technology,* 19(5):772–777, 2009.

R. Zass, A. Shashua. Nonnegative sparse PCA. In *Advances in Neural Information Processing Systems,* 1561–1568, 2006.

S. Zafeiriou, M. Petrou. Nonlinear non-negative component analysis algorithms. *IEEE Trans. on Image Processing,* 19(4):1050–1066, 2010.

Z. Zhang, C. Ding, T. Li, X. Zhang. Binary matrix factorization with applications. In *Proc. of IEEE Seventh International Conference on Data Mining (ICDM'07),* 391–400, 2007.