

# Chapter 9

## High Spatial Resolution Hyperspectral Imaging with Machine-Learning Techniques

Motoki Shiga and Shunsuke Muto

**Abstract** Recent advances in scanning transmission electron microscopy (STEM) techniques have enabled us to obtain spectroscopic datasets such as those generated by electron energy-loss (EELS)/energy-dispersive X-ray (EDX) spectroscopy measurements in a PC-controlled way from a specified region of interest (ROI) even at atomic scale resolution, also known as hyperspectral imaging (HSI). Instead of conventional analytical procedures, in which the potential constituent chemical components are manually identified and the chemical state of each spectral component is successively determined, a statistical machine-learning approach, which is known to be more effective and efficient for the automatic resolution and extraction of the underlying chemical components stored in a huge three-dimensional array of an observed HSI dataset, is used. Among the statistical approaches suitable for processing HSI datasets, methods based on matrix factorization such as principal component analysis (PCA), multivariate curve resolution (MCR), and nonnegative matrix factorization (NMF) are useful to find an essential low-dimensional data subspace hidden in the HSI dataset. This chapter describes our developed NMF method, which has two additional terms in the objective function, and which is particularly effective for analyzing STEM-EELS/EDX HSI datasets: (i) a soft orthogonal penalty, which clearly resolves partially overlapped spectral components in their spatial distributions and (ii) an automatic relevance determination (ARD) prior, which optimizes the number of components involved in the observed

---

M. Shiga

Department of Electrical, Electronic and Computer Engineering,  
Gifu University, 1-1, Yanagido, Gifu 501-1193, Japan  
e-mail: shiga\_m@gifu-u.ac.jp

M. Shiga

Precursory Research for Embryonic Science and Technology, Japan Science  
and Technology Agency, 4-1-8, Honcho, Kawaguchi, Saitama 332-0012, Japan

S. Muto (✉)

Electron Nanoscopy Division, Advanced Measurement Technology Center,  
Institute of Materials and Systems for Sustainability, Nagoya University,  
Nagoya 464-8603, Japan  
e-mail: smuto@imass.nagoya-u.ac.jp

© The Author(s) 2018

I. Tanaka (ed.), *Nanoinformatics*, [https://doi.org/10.1007/978-981-10-7617-6\\_9](https://doi.org/10.1007/978-981-10-7617-6_9)

data. Our analysis of real STEM-EELS/EDX HSI datasets demonstrates that the soft orthogonal penalty is effective to obtain the correct decomposition and that the ARD prior successfully identifies the correct number of physically meaningful components.

**Keywords** Non-negative matrix factorization • Scanning transmission electron microscopy • Hyperspectral image analysis • Electron energy-loss spectroscopy Energy-dispersive X-ray spectroscopy

## 9.1 Introduction

Current scientific analytical instruments are mostly computer-controlled and based on digital circuits. This facilitates automated measurements because the experimental procedures can be specified by using program code. For instance, recent advances in scanning transmission electron microscopy (STEM) techniques, including the development of brighter electron sources, digitally controlled operation, detectors with higher sensitivity, and sophisticated online signal processing, have enabled us to obtain comprehensive information not only on the local structures but also on the chemistry of solids by concurrently applying spectroscopic methods such as electron energy-loss (EELS) and energy-dispersive X-ray (EDX) spectroscopy to a specified region of interest (ROI). The spectrometers collect a set of spectra, each from the subnanometer area of the sample, using subnanometric incident electron probe scanning over the two-dimensional ROI with a subnanometric step width. This method is known as hyperspectral imaging (HSI). The typical data acquisition time is now reduced to several minutes for an entire EELS dataset with 2,000 energy channels over  $10^4 = 100 \times 100$  pixels (sampling points). Accordingly, the associated volume of data to be analyzed has been drastically increasing. In this context, statistical analysis methods could be more effective to thoroughly extract information embedded in massive amounts of data without any preconception, rather than relying on conventional spectral analysis of sampling points detected manually based on the insight of experts.

Among the various statistical approaches, principal component analysis (PCA) [1–3] is one of the most fundamental and popular methods. PCA successively casts mutually orthogonal eigenvectors (basis vectors) and associated score images (spatial intensity distributions of the corresponding basis vectors) in the order of significance, that is, in the order of the magnitude of eigenvalues, by way of the singular value decomposition of the HSI data matrix consisting of the experimental spectra as its row vectors. Trebbia and Bonnet [2] and Bosman et al. [3] applied PCA to EELS-HSI datasets, and not only detected exotic chemical bonding states in the samples, but also effectively filtered statistical noise from the HSI data matrix by reconstructing this matrix with a few essential basis spectra and their spatial intensity distributions. Parish and Brewer [4] studied the validity of PCA in a quantitative composition analysis of the constituent phases in their

EDX-HSI data. Note that, in their treatment, the phase overlapping areas were masked for exclusion from the quantification process; otherwise, the derived phase compositions could be biased with respect to the actual ones. These reports on PCA assumed that each pixel contains a linear combination of principal components with the orthogonality condition intrinsic to PCA. Using simulated atomic resolution EELS-HSI data, Lichteret and Verbeeck [5] pointed out that, when the noise level exceeds the intensity of the signal of interest, the signal intensities are distributed over a number of principal components, and are thus usually considered as noise. This behavior seems statistically natural, but we would not notice this phenomenon in actual experimental data. On the other hand, Spiegelberg and Rusz recently reexamined the applicability of PCA to noisy EELS data [6]. In order to estimate the amount of bias present in each principal component, Lichteret and Verbeeck [5] proposed evaluation criteria which, however, do not exhibit the correct asymptotic behavior considering the size of the dataset. Spiegelberg and Rusz [6] proposed alternative evaluation criteria, taking the size of the dataset into account.

Dobigeon and Brun [7] compared the results obtained by applying PCA, independent component analysis (ICA) [8], vertex component analysis (VCA) [9], and Bayesian linear unmixing (BLU) [10] to experimental EELS-HSI data. They eventually found that BLU provided the most plausible spatial distributions for the constituent spectral components, presumably because of its more realistic modeling of the EELS-HSI data. Spiegelberg et al. [11] also discussed a set of such data decomposition methods. In particular, they established randomized VCA (RVCA), an extension of VCA for application to noisy data, and compared its efficiency with that of minimum volume simplex analysis (MVSA) and BLU.

Over a decade, our research group has been developing an alternative method to nonnegative matrix factorization (NMF), or multivariate curve resolution (MCR) for the analysis of EELS-HSI [12, 13]. We consider this approach to be successful because NMF naturally restricts both the spatial intensities and basis spectra to nonnegative values. Contrary to NMF, the methods mentioned above such as PCA allow the spatial intensities and spectra to have negative values, which hampers the direct physical interpretation of the resolved spectral profiles. We adopted the modified alternating least-square (MALS) fitting algorithm of NMF [14] to map the different phases in the degradation of Li battery cathodes [15–19] and the chemical states of nitrogen in nitrogen-doped  $\text{TiO}_2$  [20, 21]. We also successfully applied NMF to a series of EELS datasets for the extraction of atom site-specific core-loss spectra, where the relative excitation probabilities of the spectra varied with the diffraction condition because of the electron channeling effects [22–25]. In these applied data analyses, the nonnegative constraint of the elements of extracted basis spectra and spatial intensity distributions were effective, and the resulting spectra extracted by NMF were consistent with the computational results obtained by first principles calculations [15–19, 22–25].

In general, approaches such as PCA and NMF are known as matrix factorization because these methods factorize a HSI data matrix into the product of two thin matrices, i.e. matrices of the spatial intensity distribution and basis spectra, with some suitable constraints resulting from the designed model. The next section first

briefly formalizes the problem setting of matrix factorization with HSI data [26, 27]. We then present our proposed NMF [26], which presents two advantages with respect to HSI analysis against conventional NMFs: (i) spatially clear decomposition of overlapping intensity distributions achieved by introducing a spatially orthogonal penalty term and (ii) automatic selection of a number of essential chemical components by introducing a penalty term of an automatic relevance determination (ARD) prior distribution. Our analysis of real STEM-EDX/EELS HSI datasets demonstrates that the spatial orthogonal penalty is effective to obtain the correct decomposition and the ARD prior can successfully select the correct number of physically meaningful components.

## 9.2 Methodology

### 9.2.1 Mathematical Formulation of HSI Data

The observed HSI data are stored in a three-dimensional array termed a data cube  $D(x, y, E)$ , which is a function of the two-dimensional spatial position  $(x, y)$  on the specimen and the absorption/emission energy  $E$ . For the convenience of mathematical manipulation, the data cube is often transformed to a two-dimensional  $N_{xy} \times N_{ch}$  matrix  $X$ , where  $N_{xy} = N_x \times N_y$  is the number of pixels, i.e. the product of the number of scanning steps  $N_x$  and  $N_y$  along the spatial  $x$ - and  $y$ -axis, respectively, and  $N_{ch}$  is the number of detector channels. After the transformation, the observed spectrum at position  $(x, y)$  is stored in a row of matrix  $X$ . A basic statistical method to extract a few essential basis spectra and their spatial intensity distribution assumes that the spectral intensity at each sample pixel is represented by a linear combination of the basis spectra associated with the underlying chemical components (states or phases). Assuming that the number of essential chemical components in the observed spatial region is  $K$ , which is much smaller than the size of matrix  $X$ , this analysis can be formulated by matrix factorization, which factorizes HSI data matrix  $X$  into low rank (or thin) matrices of the spatial intensity distribution  $C$  and basis spectra  $S$ :

$$X \approx CS^T, \quad (9.1)$$

where the size of  $C$  is  $N_{xy} \times K$  and the size of  $S$  is  $N_{ch} \times K$ , and superscript  $\mathbf{T}$  denotes a matrix transpose. Each column vector of  $S$  (referred to as loading in multivariate analysis) is a basis spectrum of a chemical component. On the other hand, each column vector of  $C$  (referred to as score) is a spatial intensity distribution over the ROI positions. Hence, each row vector of  $C$  is the intensities of  $K$  chemical components at a spatial position. Using the  $i$ -th column of matrix  $C$ , a two-dimensional spatial distribution of the  $i$ -th chemical component can be reconstructed by rearranging the elements such that they are returned to the original two-dimensional position.

The matrix factorization can identify both spatial intensity matrix  $C$  and spectral matrix  $S$  by minimizing the reconstruction error, which is the distance between observation  $X$  and the reconstruction, i.e.  $CS^T$ . This identification is possible because matrix  $X$ , which consists of a huge number of elements with a relatively much smaller  $K$ , is highly redundant. Thus, the identification problem is equal to that intended to find the essential subspace where the original  $X$  occurs. This approach can identify plausible  $C$  and  $S$  with much higher signal-to-noise ratios (SNRs) than those manually selected from the small number of representative observed spatial points, i.e. point-to-point analysis.

Matrix factorization needs to assume a suitable restriction of  $C$  and  $S$  because the optimization problem results in many local minima. Principal component analysis (PCA) identifies  $C$  and  $S$  by minimizing the squared error  $\|X - CS^T\|^2$  with the orthogonal constraints in both  $C$  and  $S$ . Owing to the orthogonal constraint, PCA can easily find the global solution using a singular value decomposition (SVD) algorithm. However, PCA can generate unnatural  $C$  and  $S$ , in which the element can include negative spatial intensities and spectral values. Moreover, the strong orthogonal constraint cannot allow overlaps to exist among the chemical components in both spatial and spectral space. These problems require the outputs by PCA to be adjusted to obtain physically meaningful insights. We overcame these problems by using an approach involving non-negative matrix factorization, in which the elements of  $C$  and  $S$  are not allowed to be negative.

### 9.2.2 *Non-negative Matrix Factorization with a Gaussian Noise Model*

This section presents a formal mathematical description of our model and algorithms to provide the concept of our developed NMF. Let  $X \in R_+^{N_{xy} \times N_{ch}}$  be an HSI data matrix, where  $R_+$  is the set of all nonnegative real numbers. NMF factorizes  $X$  into two thin matrices  $C \in R_+^{N_{xy} \times K}$  and  $S \in R_+^{N_{ch} \times K}$ , where  $K$  is much smaller than both  $N_{xy}$  and  $N_{ch}$ . Hence, the factorization model is given by

$$X = CS^T + \boldsymbol{\varepsilon} \quad (9.2)$$

where  $\boldsymbol{\varepsilon} \in R^{N_{xy} \times N_{ch}}$  is a noise matrix of which the elements are generated statistically independent of each other. In our problem setting, only  $X$  is observed, whereas  $C$  and  $S$  are not observed. The goal of NMF is to identify the optimal  $C$  and  $S$  under a suitable noise model  $\boldsymbol{\varepsilon}$ . One of the most common models is a Gaussian noise model, in which an element of a noise matrix is generated from a Gaussian distribution:

$$p(\varepsilon_{ij}|0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\varepsilon_{ij}^2}{2\sigma^2}\right\}, \quad (9.3)$$

where  $\sigma^2$  is the noise variance. Since the statistically independent assumption of  $\boldsymbol{\varepsilon} = X - \bar{X}$ , where  $\bar{X} = CS^T$  is the noiseless data matrix, the log-likelihood function of matrix  $X$  is given by

$$\log p(X|\bar{X}, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2 - \frac{N_{xy}N_{ch}}{2} \log 2\pi\sigma^2 \quad (9.4)$$

Taking a common statistical estimation approach of the maximum likelihood estimation, i.e. maximizing  $\log p(X|\bar{X}, \sigma^2)$ ,  $C$  and  $S$  can be optimized using only data matrix  $X$ . By taking the negative value of the log-likelihood, i.e.  $-\log p(X|\bar{X}, \sigma^2)$ , and neglecting  $\sigma^2$  in the first term and the second term, the optimization problem is transformed into the minimization of the squared error function between observation  $X$  and reconstruction  $\bar{X}$ :

$$D_{EU}(X|\bar{X}) = \frac{1}{2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2. \quad (9.5)$$

Contrary to PCA, the minimization problem of Eq. (9.5) over both  $C$  and  $S$  is non-convex, and contains a number of local minima. The optimization algorithm for an NMF does not always converge to the global optimum of  $C$  and  $S$ . Hence, it is necessary to run the optimization algorithm multiple times from different initializations, resulting in considerable computational cost. The computational efficiency has been improved by developing fast optimization algorithms such as matrix multiplication (MM) [28], alternating least-squares (ALS) [29], and hierarchical alternating least-squares (HALS) [30]. In general, MM is sensitive to the initial configuration, whereas the other algorithms are not. Among these approaches, HALS offers the best convergence to local minima. Hence, we adopted the HALS framework for the optimization of our new NMF model. Another problem presented by NMFs is that the number of chemical components needs to be manually selected in advance, which inevitably introduces a problem similar to that of PCA if the noise level is larger than the signal intensities. As the number of components increases, the reconstruction error naturally decreases. However, this decrease is not essential to identify  $C$  and  $S$  because it results in overfitting to observed data when the number of components is excessively large. Thus, relying on the reconstruction error only cannot be useful to identify the essential number of physically meaningful components.

To overcome the above difficulties in STEM-EELS/EDX HSI data analysis, we developed a new NMF model that imposes the following penalty terms on the spatial intensity matrix  $C$ : (i) a spatial orthogonal penalty [31] and (ii) a sparse

penalty to optimize the number of components, termed an automatic relevance determination (ARD) prior [32]. For the optimization of low-rank matrices  $C$  and  $S$ , we further developed an algorithm based on hierarchical alternating least-squares (HALS) [30], which is more efficient than the matrix multiplication (MM) [28] used before [32]. The following section describes these extended models and their optimization algorithms.

### 9.2.3 Optimization Algorithms with Soft Spatial Orthogonal Constraint

A goal of HSI data analysis is to identify the pure spectra and spatial intensity distributions of each chemical component from the spectra and distribution of a mixture of chemical components, i.e. observed matrix  $X$ . The basic NMF model, e.g. the minimization of Eq. (9.5), often generates unresolved spectra and spatial distributions that still contain spatially overlapped or unnaturally unresolved spectra because the basic NMF induces sparse decomposition of  $C$  and  $S$ . However, the EELS spectrum of a pure chemical component is not sparse, meaning that the intensities of an EELS spectrum are more than zero over all energy bands, whereas the intensities of an EDX spectrum are almost zero except for the peak positions. Hence, poor resolution is more problematic in STEM-EELS analysis than in EDX analysis.

Our approach to solve the above problem entails introducing the spatial orthogonal constraint [31]. This constraint ensures that spectral matrix  $S$  is relatively more relaxed than  $C$  and then  $S$  can be non-sparse. Because the exact orthogonal constraint is too strict, we used weight parameter  $w$  to relax this constraint, which is known as a soft spatial orthogonal constraint. Then our objective function of  $C_{\cdot k}$  to be minimized is formulated as follows:

$$\frac{1}{2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} \left( [X^{(k)}]_{ij} - [C_{\cdot k} S_{\cdot k}^T]_{ij} \right)^2 + w \cdot \xi_k C_{\cdot k}^T c^{(k)} \quad \text{s.t.} \quad \|C_{\cdot k}\|_2 = 1, \quad (9.6)$$

where

$$X^{(k)} = X - CS + C_{\cdot k} S_{\cdot k}^T, \quad k = 1, \dots, K, \quad (9.7)$$

$$c^{(k)} = \sum_{m \neq k} C_{\cdot m}, \quad k = 1, \dots, K. \quad (9.8)$$

Parameter  $w$ ,  $0 \leq w \leq 1$ , is important to adjust the orthogonal penalty and  $\xi_k$  is the Lagrange multiplier for the exact orthogonal constraint of  $C$ . When  $w=1$  the optimized  $C$  is an exact orthogonal matrix in which any chemical components do NOT overlap. When  $w=0$ , among all the components, the optimized components

in  $C$  may extensively overlap. The optimal value of  $w$  depends on the situation, such as the spatial resolution of the data (step width of STEM-HSI) and localization of chemical components. Thus, the optimal value of  $w$  must be chosen according to the measurement level.

Applying some algebra to Eq. (9.6) enables us to obtain an analytical solution in terms of the Lagrange multiplier  $\xi_k$ . Substituting the obtained  $\xi_k$ , the update rule of  $C_{\cdot k}$  is given by

$$C_{\cdot k} = \left[ X^{(k)} S_{\cdot k} - w \frac{c^{(k)\text{T}} X^{(k)} S_{\cdot k}}{c^{(k)\text{T}} c^{(k)}} C_{\cdot k} \right]_+, \quad k = 1, \dots, K \quad (9.9)$$

where the operator  $[A]_+$  replaces all negative values in matrix  $A$  with zeros. Hence, it can be calculated by  $[A]_+ = \{A + \text{abs}(A)\}/2$ , where function  $\text{abs}$  outputs a matrix consisting of the absolute value of the elements in  $A$ . The second term weighted by  $w$  is due to the orthogonal penalty term. After applying Eq. (9.9), each column of  $C$  should be normalized by

$$C_{\cdot k} \leftarrow C_{\cdot k} / \|C_{\cdot k}\|_2, \quad k = 1, \dots, K. \quad (9.10)$$

Thus, we omit the normalization of  $S$ , and the update is given by

$$S_{\cdot k} = \left[ \left( X^{(k)} \right)^{\text{T}} C_{\cdot k} \right]_+, \quad k = 1, \dots, K. \quad (9.11)$$

Figure 9.1 provides the pseudo-code of this NMF, which we named SO-NMF.

### 9.2.4 Probabilistic View of a NMF Model with an Automatic Relevance Determination Prior

Optimizing the number of components using only the observed HSI data is practically important. Maximum likelihood estimation (or an estimation based on minimizing errors) cannot be effective for the optimization because it causes overfitting of the HSI data when the number of components is large. This overfitting problem is avoided by using a Bayes estimation (or a maximum a posteriori (MAP) estimation) with a prior distribution of scale parameters (relevance weights) [32]. The process of choosing only the important components is known as automatic relevance determination (ARD).

To perform ARD in NMF, we assume a prior distribution for  $C$  using an exponential distribution with a scale parameter  $\lambda_k$  for the probability density of column  $k$  of  $C$ , i.e.  $C_{\cdot k}$ :



	<p><b>Input:</b> Data matrix <math>\mathbf{X}</math>, the weight of orthogonal constraint <math>w</math>, the number of components <math>K</math>, the maximum number of iterations <math>T_{\max}</math>, the number of initializations <math>R_{\max}</math></p> <p><b>Output:</b> Spatial intensities of components <math>\mathbf{C}</math> and basis spectra <math>\mathbf{S}</math></p>
1:	For $r$ from 1 to $R_{\max}$ :
2:	$t = 0$
3:	While $t < T_{\max}$ and $L^{(t,r)} \neq L^{(t-1,r)}$ (not converged)
4:	$t = t + 1$
5:	For $k$ from 1 to $K$ :
6:	Update $\mathbf{C}_{\cdot k}$ by Eq. (9.9)
7:	Normalize $\mathbf{C}_{\cdot k}$ by Eq. (9.10)
8:	end
9:	For $k$ from 1 to $K$ :
10:	Update $\mathbf{S}_{\cdot k}$ by Eq. (9.11)
11:	end
12:	Compute $L^{(t,r)} = D_{EU}(\mathbf{X} \mathbf{C}\mathbf{S}^T)$ by Eq. (9.5)
13:	end
14:	end
15:	Choose the best optimization results by $r_{best} = \operatorname{argmin}_r L^{(T_{\max},r)}$

**Fig. 9.1** Pseudocode of our NMF with the soft orthogonal constraint (SO-NMF)

$$p(C_{nk}|\lambda_k) = \frac{1}{\lambda_k} \exp\left(-\frac{C_{nk}}{\lambda_k}\right), \quad n = 1, \dots, N_{xy}, \quad k = 1, \dots, K. \quad (9.12)$$

The above density distribution generates nonnegative random values with a large probability density around zero, resulting in a sparse matrix of  $\mathbf{C}$ . For the prior distribution of  $\lambda_k$ , we assume an inverse-Gamma distribution:

$$p(\lambda_k|a, b) = \frac{b^a}{\Gamma(a)} \lambda_k^{-(a+1)} \exp\left(-\frac{b}{\lambda_k}\right), \quad k = 1, \dots, K, \quad (9.13)$$

where  $a$  and  $b$  are hyper-parameters to adjust the sparseness of  $\lambda_k$ . On the other hand, the probability density distribution of column  $k$  of  $\mathbf{S}$ , i.e.  $p(S_{\cdot k})$ , is assumed to be uniformly distributed on the unit hyper-sphere in  $\mathbf{R}_+^{N_{ch}}$ . Using Eqs. (9.4), (9.12), and (9.13), the negative log-likelihood function of an NMF model with ARD priors is given by

$$\begin{aligned}
L(C, S, \lambda, \sigma^2) &= -\log p(X|\bar{X}, \sigma^2) - \sum_{i=1}^{N_{xy}} \sum_{k=1}^K \log p(C_{ik}|\lambda_k) \\
&\quad - \sum_{k=1}^K \log p(S_{\cdot k}) - \sum_{k=1}^K \log p(\lambda_k|a, b) \\
&= \frac{N_{xy}N_{ch}}{2} \log 2\pi\sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2 \\
&\quad + \sum_{k=1}^K \frac{1}{\lambda_k} \left( b + \sum_{i=1}^{N_{xy}} C_{ik} \right) + (N_{xy} + a + 1) \sum_{k=1}^K \log \lambda_k \\
&\quad + K(a \log b - \log \Gamma(a)),
\end{aligned} \tag{9.14}$$

With regard to the optimization of  $C$ ,  $L(C, S, \lambda, \sigma^2)$  is a penalized likelihood function with the  $L_1$  norm of  $C$ , resulting in a sparse matrix  $C$ . The NMF minimizing  $L(C, S, \lambda, \sigma^2)$  is referred to as ARD-NMF.

Because the simultaneous optimization of  $L(C, S, \lambda, \sigma^2)$  over all  $C, S, \lambda$ , and  $\sigma^2$  is non-convex, multiple optimizations from different initial configurations are required. To update  $C$  and  $S$ , we use HALS [30], which updates each column  $C_{\cdot k}$  and  $S_{\cdot k}$  alternately. Applying some algebra to the minimization of  $L(C, S, \lambda, \sigma^2)$ , we obtain the following update rule for  $C_{\cdot k}$ :

$$C_{\cdot k} = \left[ X^{(k)} S_{\cdot k} - \frac{\sigma^2}{\lambda_k} \right]_+, \quad k = 1, \dots, K \tag{9.15}$$

The second term in Eq. (9.15) is attributable to the ARD prior, which induces the sparse matrix of  $C$ . The update rule for  $S_{\cdot k}$  by HALS is given by

$$S_{\cdot k} = \frac{\tilde{s}_k}{\|\tilde{s}_k\|_2}, \quad k = 1, \dots, K \tag{9.16}$$

where  $\|x\|_2$  is the  $L_2$  norm of vector  $x$  and

$$\tilde{s}_k = \left[ \left( X^{(k)} \right)^T C_{\cdot k} \right]_+, \quad k = 1, \dots, K \tag{9.17}$$

Similarly, the update rules for the relevance weight  $\lambda$  and  $\sigma^2$  to minimize  $L(C, S, \lambda, \sigma^2)$  with all other quantities fixed are given by

$$\lambda_k = \frac{b + \sum_{i=1}^{N_{xy}} C_{ik}}{N_{xy} + a + 1}, \quad k = 1, \dots, K \tag{9.18}$$

$$\sigma^2 = \frac{1}{N_{xy}N_{ch}} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2 \quad (9.19)$$

The hyper-parameter  $b$  can be set using an approximate empirical estimator [32] as follows:

$$b = \frac{(a-1)\sqrt{N_{ch}}}{K} \frac{1}{N_{xy}N_{ch}} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} X_{ij}, \quad (9.20)$$

In our experiments, the hyper-parameter  $a$  was set to  $a = 1 + \delta$ , where  $\delta = 10^{-16}$ , to choose the minimum number of components with the minimum  $L(C, S, \lambda, \sigma^2)$ . After the optimization of ARD-NMF, the relevance (or importance) values of components are given by  $\lambda_k, k = 1, \dots, K$ . Because the values of redundant components cannot be exactly zero, we empirically set a threshold value to remove such components.

### 9.2.5 Optimization Algorithm for $C$ with Both ARD and Spatial Orthogonal Constraint

When we simply combine the soft orthogonal constraint and the ARD effect using both penalty terms, then the update rule of  $C_{.k}$  can be obtained as follows:

$$C_{.k} = \left[ X^{(k)} S_{.k} - \frac{\sigma^2}{\lambda_k} - w \frac{c^{(k)T} (X^{(k)} S_{.k} - \sigma^2 / \lambda_k)}{c^{(k)T} c^{(k)}} C_{.k} \right]_+, \quad k = 1, \dots, K \quad (9.21)$$

In this update,  $C_{.k}$  should not be renormalized to reduce the effect of the orthogonal constraint for irrelevant components. We propose Eq. (9.21) as an update rule for  $C_{.k}$  when the orthogonal constraint is necessary. Figure 9.2 shows the pseudo-code of our proposed NMF algorithm, which we named ARD-SO-NMF. In the special case without the orthogonal constraint, i.e.  $w = 0$ , the algorithm is simply ARD-NMF. Line 12–20 has the purpose of merging the components when the spectra are similar. In this procedure, the similarity is evaluated by using the cosine similarity and the spectra are considered to be the same when the value exceeds 0.99. This operation is necessary to choose the correct number of components because the orthogonality condition with  $w > 0$  enforces splitting of the components even when the spectra are exactly the same. Our MATLAB and Python codes are available at <https://github.com/MotokiShiga>.

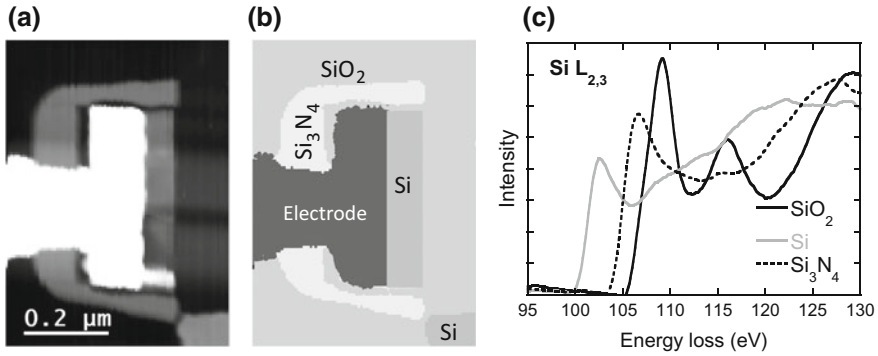
	<p><b>Input:</b> Data matrix <math>\mathbf{X}</math>, the weight of orthogonal constraint <math>w</math>, the maximum number of components <math>K</math>, the maximum number of iterations <math>T_{\max}</math>, the number of initializations <math>R_{\max}</math></p> <p><b>Output:</b> Spatial intensities of components <math>\mathbf{C}</math>, basis spectra <math>\mathbf{S}</math> and relativeness <math>\lambda</math></p>
1:	Set hyper-parameter $a = 1 + \delta$ and compute $b$ by Eq. (9.20)
2:	For $r$ from 1 to $R_{\max}$ :
3:	$t = 0$
4:	While $t < T_{\max}$ and $L^{(t,r)} \neq L^{(t-1,r)}$ (not converged)
5:	$t = t + 1$
6:	For $k$ from 1 to $K$ :
7:	Update $\mathbf{C}_{\cdot k}$ by Eq. (9.21)
8:	end
9:	For $k$ from 1 to $K$ :
10:	Update $\mathbf{S}_{\cdot k}$ by Eq. (9.16)
11:	end
12:	For $k$ from 1 to $K$ :
13:	For $m$ from 1 to $(k-1)$ :
14:	If $\mathbf{S}_{\cdot k}^T \mathbf{S}_{\cdot m} > 0.99$ :
15:	$\mathbf{S}_{\cdot m} = \mathbf{1} / \sqrt{N_{ch}}$
16:	$\mathbf{C}_{\cdot k} = \mathbf{C}_{\cdot k} + \mathbf{C}_{\cdot m}$
17:	$\mathbf{C}_{\cdot m} = 0$
18:	end
19:	end
20:	end
21:	Update $\lambda$ by Eq. (9.18)
22:	Update $\sigma^2$ by Eq. (9.19)
23:	Compute $L^{(t,r)} = L(\mathbf{C}, \mathbf{S}, \lambda, \sigma^2)$ by Eq. (9.14)
24:	end
25:	end
26:	Choose the best optimization results by $r_{best} = \operatorname{argmin}_r L^{(T_{\max}, r)}$

**Fig. 9.2** Pseudocode of our NMF with ARD and the soft orthogonal constraint (ARD-SO-NMF)

## 9.3 Application

### 9.3.1 Experimental Procedures

A real dataset was acquired from a cross-sectional TEM (XTEM) sample of a Si diode, prepared by a focused ion beam (FIB) technique. We recorded the HSI data for Si- $L_{2,3}$  including zero-loss peak (ZLP) using a JEOL JEM-1000 K RS ultra-high voltage S/TEM of Nagoya University, operated at 1000 kV, with a Gatan Quantum equivalent EEL spectrometer of which the energy dispersion was set to 0.2 eV/channel.

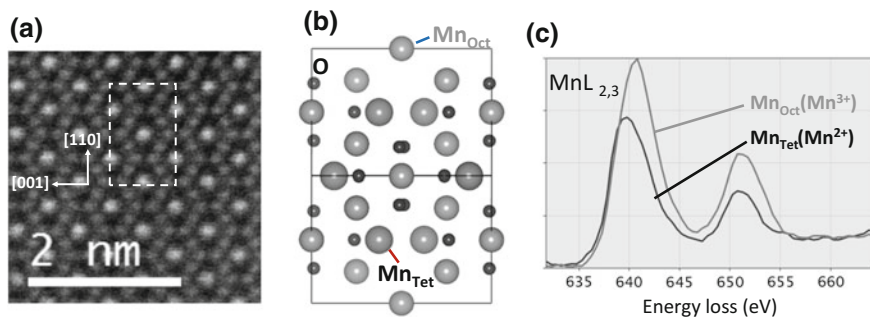


**Fig. 9.3** **a** Cross-sectional ADF-STEM image, **b** spatial distribution of three components, and **c** their reference Si- $L_{2,3}$  spectra of silicon diode test sample

The sample thickness of the measured area was estimated at  $0.1 \mu\text{m}$  from the low-loss spectrum. The energy drift of the spectra during the acquisition was corrected by ZLP alignment and calibration. After the energy calibration, the pre-edge background modeled by a power law was subtracted to isolate the Si- $L_{2,3}$  spectrum. Figure 9.3 shows an annular dark-field STEM (ADF-STEM) image of the ROI of the Si diode, a manually validated component map and spectra.

Another experimental STEM-EELS HSI dataset was prepared by measuring the atomic resolution EELS-HSI of  $\text{Mn}_3\text{O}_4$ . Polycrystalline  $\text{Mn}_3\text{O}_4$  with a spinel crystal structure was obtained, and a TEM sample was prepared by conventional ion milling as previously described [24]. We measured the Mn- $L_{2,3}$  HSI using a JEOL ARM-200F aberration-corrected STEM, operated at 200 kV, with the Gatan Quantum EELS having an energy dispersion of 0.5 eV/channel. The average full width at half maximum (FWHM) of ZLP collected simultaneously (Dual EELS mode) with Mn  $L_{2,3}$  was approximately 2 eV. The thickness of the measured area was approximately 40 nm, estimated from the low-loss spectra. Prior to applying NMF to the data, the energy drift of the spectra during the acquisition was corrected using the dual EELS mode synchronized with the ZLP alignment and calibration. After the energy calibration, the pre-edge background intensities were subtracted by modeling them with a power law.

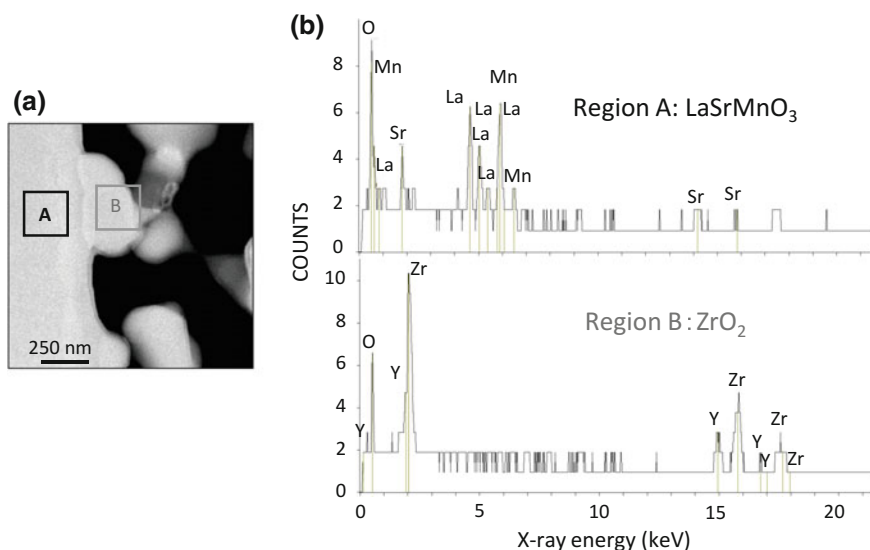
Figure 9.4a–c show the ADF-STEM image, schematic projected structure of the  $\text{Mn}_{\text{Tet}}$  (divalent Mn occupying the tetrahedral site,  $\text{Mn}_{\text{Oct}}$  (trivalent Mn at the octahedral site) and O (oxygen) columns along the present incident beam direction and the extracted site-specific spectra, respectively. In (a) the heavier element (Mn) alone appears bright. These data are more difficult to analyze, because the inner shell excitation is delocalized by a certain distance and the neighboring atomic columns simultaneously contribute to the spectrum intensity at a sampling point due to electron channeling effects [25] and orbital hybridization between the elements.



**Fig. 9.4** **a** ADF-STEM image, **b** corresponding atom-site positions in the framed area of **(a)**, and **c** Mn- $L_{2,3}$  reference spectra for STEM-EELS-HSI data from Mn<sub>3</sub>O<sub>4</sub>

For all datasets, even after the above pre-processing, a few elements in  $X$  had small negative values due to background removal. Thus, we replaced these values with zeros. To normalize the scale of  $X$ , all elements were divided by the average of the elements in  $X$ .

An STEM-EDX-HSI dataset was acquired from a sintered ceramic composite of Y-doped ZrO<sub>2</sub>-LaSrMnO<sub>3</sub> (supplied by courtesy of Dr. T. Mori of the National Institute of Materials Science), which exhibits a distinct composition variation across the electron transparent sample area. A thin film was prepared for TEM by using an FIB technique. We measured the EDX-HSI using a JEOL 2100F S/TEM



**Fig. 9.5** ADF-STEM image of LaSrMnO<sub>3</sub>-Y doped ZrO<sub>2</sub> ceramic composite sample **(a)** and typical EDX counts per pixel from framed areas **(b)**

operated at 200 kV, equipped with a JEOL EDX silicon drift detector, Dry SD60GV. Figure 9.5 shows the ADF-STEM image and typical counts (spectra) in representative points, corresponding to the two different phases, where the maximum net peak counts per pixel do not exceed 10 counts, and have a typical sparse feature that is suitable for testing the relevance of the proposed method.

### 9.3.2 Spatial Orthogonal Constraint on STEM-EELS Data

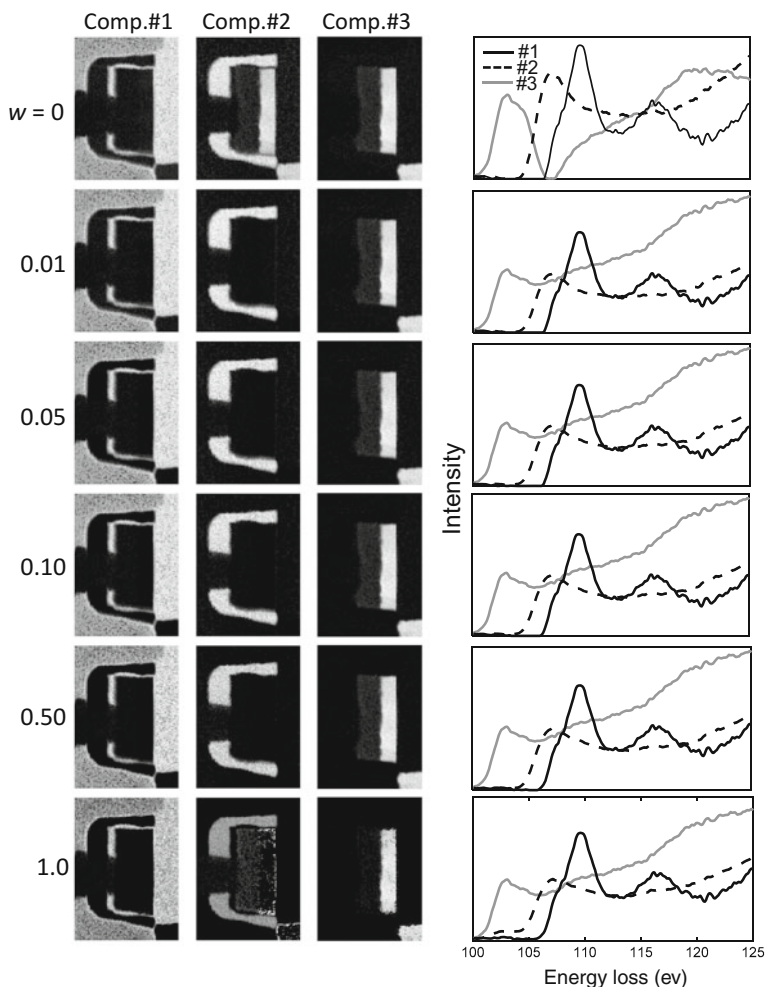
We evaluated the effect of the orthogonal constraint by changing the value of  $w$  with a fixed number of components, i.e. SO-NMF. We used the two STEM-EELS-HSI datasets described in Sect. 9.3.1. Because neither the spatial distribution maps nor the spectra in the datasets are sparse, the conventional NMF optimization has multiple local minima. Thus, reaching the global minima (or a good local minimum) is difficult. Our aim in this experiment was to verify that the orthogonal constraint reduces the search space and that SO-NMF generates a reasonable decomposition of NMF.

#### 9.3.2.1 XSTEM-EELS Data from a Silicon Device

The method was first applied to the dataset from the Si diode sample, as shown in Fig. 9.3 in Sect. 9.3.1. The number of components in SO-NMF was set to  $K = 3$ , which is the number of reference components. In the result with  $w = 0$  (no orthogonal constraint: first row in Fig. 9.6), the third spectral components exhibit unnatural intensity decreases at 110 eV, where a sharp peak from the first spectral component is overlaid. This can happen in EELS-HSI under certain conditions [17]. This sudden lowering in intensity disappears when spatial orthogonality ( $w \geq 0.01$ ) is included, as seen in Fig. 9.6. Slight cross-talk between the second and third components remains in both the spatial distribution maps and spectra for  $w = 0.01$ . The resolved spectral profiles and their spatial distributions are almost the same for  $w \geq 0.05$ , which effectively reproduces the spectra and expected spatial distributions, although the spatial phase separation seems (unnaturally) overly emphasized for  $w = 1.0$ .

#### 9.3.2.2 Atomic Resolution STEM-EELS of $\text{Mn}_3\text{O}_4$

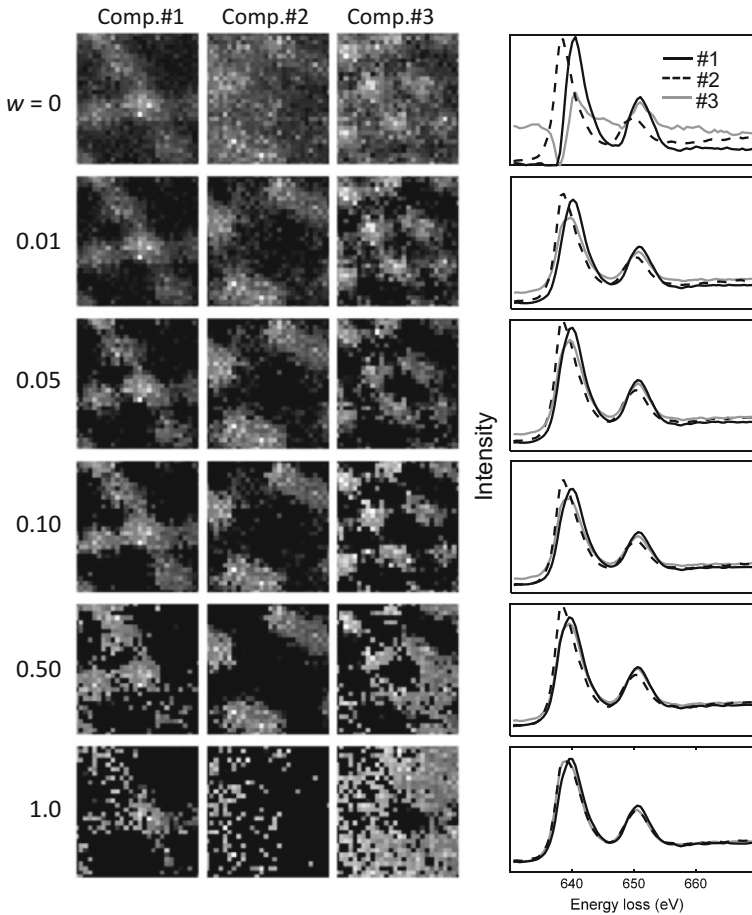
Next, we validated the method using the atomic resolution Mn- $L_{2,3}$  SI data from the  $\text{Mn}_3\text{O}_4$  spinel sample (cf. Fig. 9.4 in Sect. 9.3.1). The number of components in SO-NMF was set to  $K = 3$ , which is the number of components determined by ARD-SO-NMF in Sect. 9.3.3.3. The SO-NMF results for  $0 \leq w \leq 1$  are shown in Fig. 9.7, with the score images in the first, second, and third columns and the resolved spectral profiles in the fourth column. In the case without spatial



**Fig. 9.6** Results of SO-NMF with various weights of spatial orthogonality constraint for Si- $L_{2,3}$  STEM-EELS-HSI data

orthogonality ( $w = 0$ ), the resolved spectral profiles are inconsistent with the expected reference profiles (Fig. 9.4c), the peak at around 640 eV of component 2 shifted to the left. Further, component 3 exhibits a physically unnatural intensity drop at the distinct peak positions of component 1, as for the case of no orthogonality (Fig. 9.7, top-right figure). With a small spatial orthogonality ( $w = 0.01$ ) included, the spectral shapes converged to those consistent with the reference spectra and the additional component localized on the oxygen columns. It can be seen that there is an optimum value of  $w$  for reproducing good spectral profiles and plausible spatial distributions. As  $w$  increases, the spatial distributions become more orthogonal to each other, whereas the resolved spectral shapes converge to one





**Fig. 9.7** Results of SO-NMF with various weights of the spatial orthogonality constraint for Mn- $L_{2,3}$  STEM-EELS-HSI data

form. For  $w \geq 0.5$  the spatial distributions are far from the actual projected structures, even though the resolved spectral shapes are essentially identical.

We subsequently focus on the additional third component, the spatial distribution of which was found to be localized on the projected oxygen atom positions. This localization was attributed to the electron channeling effect [25], which is responsible for propagating the incident electron wave function along the neighboring Mn columns for a sample exceeding a certain thickness when the electron probe is placed on the oxygen column. The resolved spectrum of the third component actually exhibits a spectral profile characteristic of the weighted average of the other two components, because each oxygen atom is coordinated with trivalent  $Mn_{Oct}$  and divalent  $Mn_{Tet}$  atoms.

### 9.3.3 Results of Optimizing the Number of Components by ARD-NMF

#### 9.3.3.1 STEM-EDX Data

To examine whether ARD can select the correct number of components, our ARD-NMF (i.e. without the orthogonal constraint imposed,  $w = 0$ ) was applied to the STEM-EDX-HSI data of a Y-doped  $\text{ZrO}_2$ (YSZ)– $\text{LaSrMnO}_3$ (LSM) ceramic composite material. The conventional elemental distributions are shown in Fig. 9.8 for reference purposes. Starting with 10 components, only two survived after the optimization algorithm terminated, as shown in Fig. 9.9. The distribution of each identified component shown in Fig. 9.9a is consistent with the elemental distributions of: (1) the union of La, Mn, and Sr and (2) the union of Zr and Y in Figs. 9.5 and 9.8. The identified spectra shown in Fig. 9.9c consist of sets of peaks, each reflecting the correct composition of YSZ or LSM in Fig. 9.5. This indicates that our ARD-NMF identified the constituent phases correctly for this STEM-EDX data.

These results indicate that the present method effectively removed the statistical noise in the resolved spectra (Fig. 9.9c), and the score images (Fig. 9.9a) exhibit no artificial mixing of the two spectral components. Note that a 10-nm layer of LSM (Fig. 9.9a: Comp.#1) covers the YSZ substrate surface; this can be seen more clearly here than in the elemental maps.

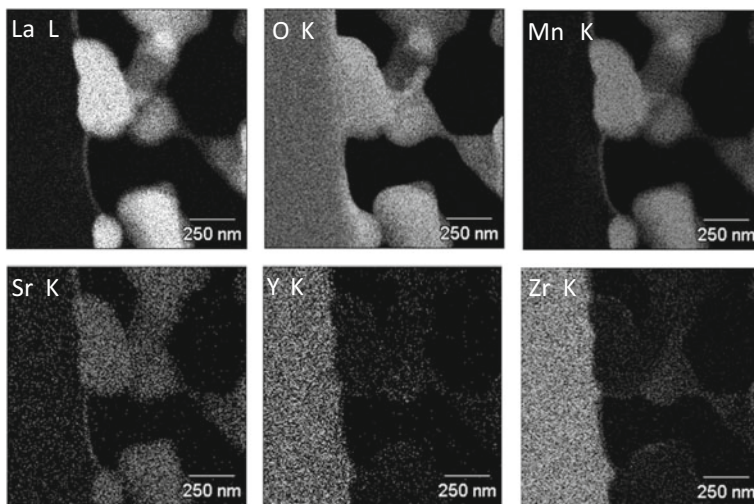


Fig. 9.8 EDX elemental maps of  $\text{LaSrMnO}_3$ -Y-doped  $\text{ZrO}_2$  ceramic composite sample

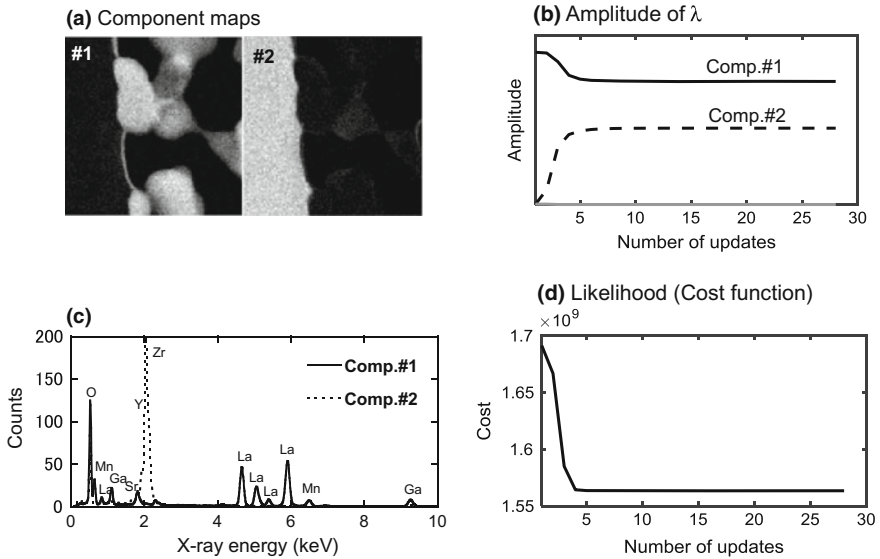


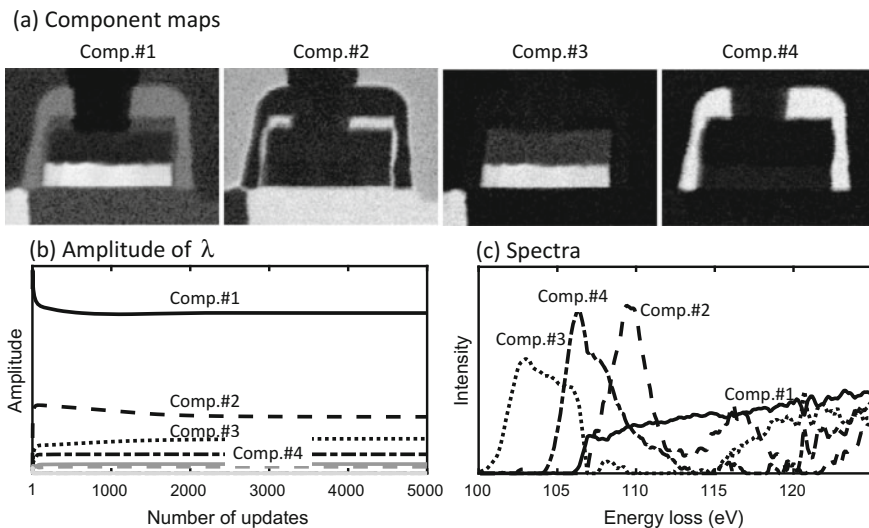
Fig. 9.9 Result of ARD-NMF for STEM-EDX-HSI data

### 9.3.3.2 XSTEM-EELS Data from a Silicon Device

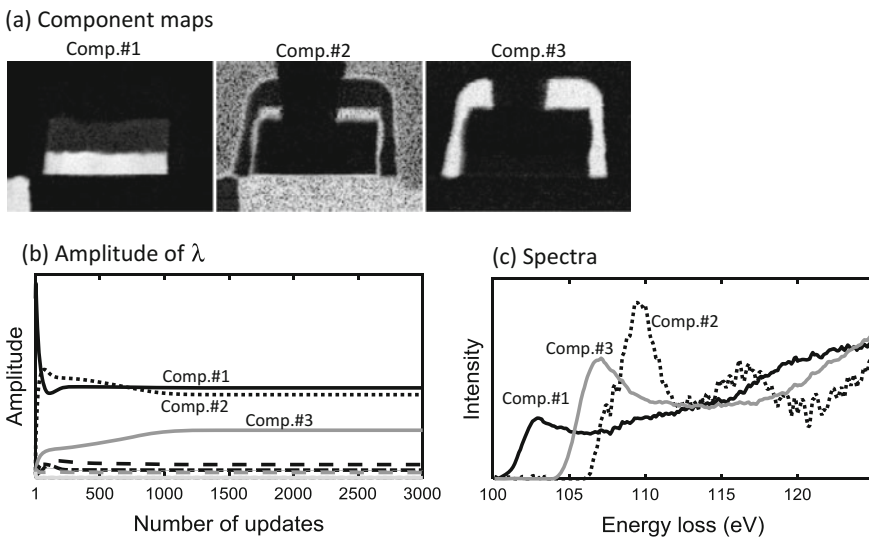
We then applied both ARD-NMF and ARD-SO-NMF with  $K = 10$  to the EELS-HSI data of Si- $L_{2,3}$  energy-loss near edge structure (ELNES) obtained from a cross-sectional Si diode sample. The reference component spectra (Fig. 9.3c) are not sparse, that is, nonzero values range over the energy-loss axis. We compared the results of ARD-NMF with those from ARD-SO-NMF to verify that the orthogonal constraint produces a clearer decomposition for non-sparse data. The results are shown in Figs. 9.10 and 9.11.

Figure 9.10b shows that ARD-NMF selected four components whereas the reference contains three. In Fig. 9.10a, the generated component distribution of components 1 and 3 exhibit extensive overlap, whereas the actual spectra of these components are not overlapped, as also seen in the case of  $w = 0$  in Fig. 9.6. This result was attributed to a property of basic NMF, which induces a sparse decomposition on both spatial and spectra matrices.

The ARD-SO-NMF results (with  $w = 0.1$ ) are shown in Fig. 9.11b. Figure 9.11c shows that the identified spectra are consistent with the reference spectra shown in Fig. 9.3c. The spatial distributions of components obtained by ARD-SO-NMF (Fig. 9.11a) are clearly separated, whereas those resulting from ARD-NMF (Fig. 9.10a) overlap extensively. This difference demonstrates the effect of the orthogonal constraint. In this case, ARD-SO-NMF selected three components and their spectra are consistent with their reference spectra, whereas the spectra by ARD-NMF ( $w = 0$ ) display unnatural reductions in intensity. Thus, these results



**Fig. 9.10** Result of ARF-NMF ( $w = 0$ ) for Si- $L_{2,3}$  STEM-EELS-HSI data from silicon diode sample



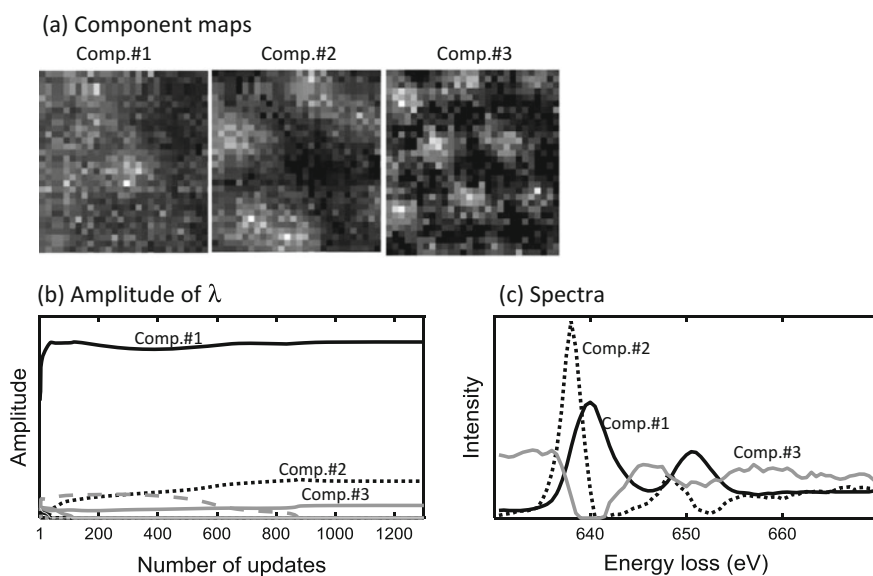
**Fig. 9.11** Result of ARF-SO-NMF ( $w = 0.01$ ) for Si- $L_{2,3}$  STEM-EELS-HSI data from silicon diode sample

suggest that the method can effectively detect subtle spectral differences by introducing the orthogonal constraint.

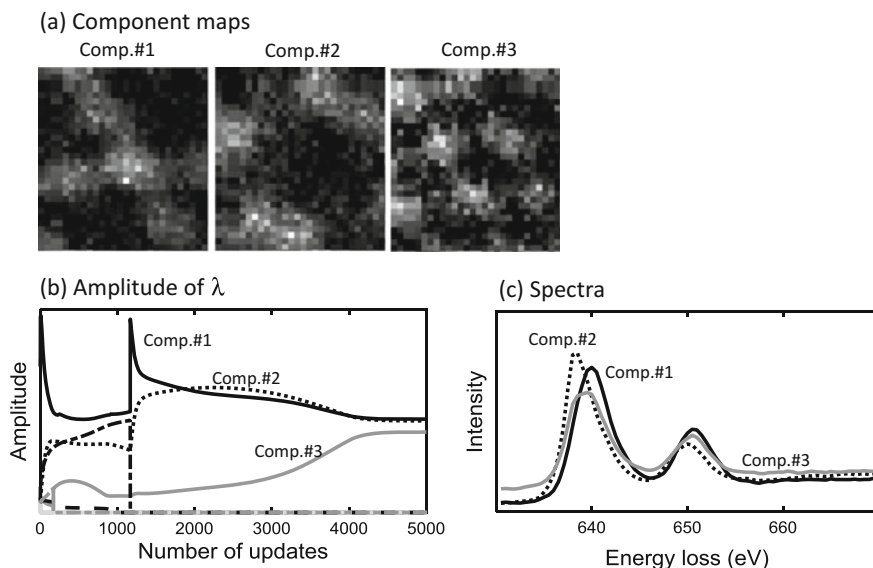
### 9.3.3.3 Atomic Resolution STEM-EELS of $\text{Mn}_3\text{O}_4$

The ARD-NMF technique was also applied to the experimental atomic resolution Mn- $L_{2,3}$  SI data from the  $\text{Mn}_3\text{O}_4$  spinel sample. Figures 9.12a–c show the ARD-NMF results with  $K = 10$ . As shown in Fig. 9.12b, ARD-NMF selected three components and eliminated the other seven components during the optimization. Thus, ARD-NMF detected an additional component other than those related to the two Mn sites, as discussed in Sect. 9.3.2.2. The spatial distributions of the resolved components shown in Fig. 9.12a and c are basically consistent with the projected  $\text{Mn}_{\text{Oct}}$  and a  $\text{Mn}_{\text{Tet}}$  atom positions in Figs. 9.4b and 9.4c, respectively, the relative chemical shifts of which are also consistent with their valence states. However, the boundary between the components is less clear because of the delocalization of the chemical bonding states, and the resolved spectral profiles (Fig. 9.12c) are inconsistent with the expected theoretical profiles, with component 2 exhibiting physically unnatural intensity decreases at the distinct peak positions of component 1. Because an ARD prior induces sparseness, this problem often occurs when ARD-NMF is applied to EELS in which different spectral profiles largely overlap.

We overcame this problem by applying ARD-SO-NMF with  $w = 0.01$ . Figure 9.13a demonstrates that, because of the orthogonal condition, the



**Fig. 9.12** Result of ARF-NMF ( $w = 0$ ) for Mn- $L_{2,3}$  STEM-EELS-HSI data from  $\text{Mn}_3\text{O}_4$



**Fig. 9.13** Result of ARF-SO = NMF ( $w = 0.01$ ) for Mn- $L_{2,3}$  STEM-EELS-HSI data from  $\text{Mn}_3\text{O}_4$

components are separated more clearly. Especially, overlaps between the first component and the others are resolved with greater clarity, as shown in Fig. 9.13a, and detection of the third component was improved.

## 9.4 Discussion

The ARD-SO-NMF and ARD-NMF algorithms proposed in this study were able to optimize the number of spectral components for both the EDX and EELS datasets. An additional orthogonal constraint was required when neither the spatial distribution nor the spectra were sparse. Such a constraint is offered by the proposed ARD-SO-NMF. Our NMF realistically extracts spectral components from the EELS-HSI data when the spatial orthogonality penalty is appropriate, implying that different spectral features are spatially well separated.

Because of the differing complexities intrinsic to EELS and EDX spectra, NMF processes these datasets differently. An EDX spectrum can be characterized by a set of Gaussian-like peaks, generally separated in energy, whereas an elemental core EELS includes various spectral components overlapped in the same energy range, where the corresponding electronic energy levels in solids are approximately continuously distributed. Furthermore, the spectral components of EDX are mostly sparse and orthogonal along the energy axis, contrary to those in EELS. NMF with only an error function as the objective function prefers the orthogonal basis spectra

in the energy axis because of their completeness. On the other hand, NMF with spatial orthogonality models the practical situation, in which an EELS-HSI dataset is assumed to be more orthogonal in space than in energy, more accurately.

There are several local minima in the likelihood functions. The type of NMF algorithm appears to eventually achieve a more appropriate minimum, although it is not mathematically possible to prove the dependence. Moreover, because of the computational cost, it is difficult to obtain all of the local minima, even when we apply the spatial orthogonality constraint and sparse priors for the ARD effect, i.e. by using a small value of  $K$ , in the data matrix. The present NMF method is capable of minimizing and extracting the objective function of particular solutions by systematically varying the weight of the spatial orthogonality in the object function. In both of the EELS examples presented herein, an increase in  $w$  caused the resolved components to be distributed more widely over the sample space and their spectral shapes to become less sparse (or orthogonal). This change in spectral shape, which is prone to be sparse under the basic NMF, is clearly resolved and exhibits the composition more accurately when the orthogonal constraint is applied. Hence, the proposed NMF can identify chemical states from the resolved spectra more accurately than existing methods that do not use spatial orthogonality. In the case of the atomic resolution HSI of  $Mn_3O_4$ , the method resulted in physically meaningless solutions when we overestimated the spatial orthogonality. In general, we can reach solutions that are physically more realistic/interpretable, comparable to the theoretical spectra predicted by first principles calculations or reference experimental spectra, by changing the value of  $w$  systematically and understanding the extent to which the spectral shapes and spatial distributions of the resolved components vary. This scheme seems much more effective and pragmatic than estimating the solution bounds by repeating the decomposition routines with many different initial random numbers in the loading or score matrices.

In this respect the proposed SO constraint may fail when the spatial distributions of the component states strongly overlap with each other. Spiegelberg et al. proposed an alternative scheme to extract nonnegative source signals of strongly mixed data [33], instead of imposing the present SO constraint. By randomly drawing samples from the space of positive spectra in the signal subspace spanned by the prominent principal components, a sampled dataset of which the spectral components can be conveniently extracted using, e.g., VCA or NMF, is obtained with a large probability. These components typically correspond well to the pure spectra of the original data assuming that the spectral components are orthogonal to each other in at least one channel.

Existing processing schemes can produce controversial results should the spectral background be subtracted in advance before the statistical processing, and this probably depends on the type of spectral data being processed. Although background subtraction is generally considered to lose important spectral information, we believe background subtraction to be necessary in the present framework because our NMF assumes that no background structure is incorporated. As demonstrated in the supplementary material in a previous paper [26], our proposed NMF was unable to provide the expected correct results for STEM-EELS SI

without background subtraction, which thus presents further work for the future, i.e. incorporating a background structure in our model.

## 9.5 Summary

We proposed a new multivariate curve resolution method based on NMF with two penalty terms: (i) a soft orthogonal constraint to effectively resolve overlapping spectra, and (ii) an ARD prior to optimize the number of components. Validations using experimental STEM-EDX/EELS SI data demonstrated that the ARD prior successfully resolved the correct number of physically interpretable spectral components. The soft orthogonal constraint was effective for STEM-EELS HSI data that were neither sparse in the spatial nor the spectral regions. The proposed SO-NMF and ARD-SO-NMF schemes can successfully resolve physically meaningful components by reducing the search space for low-rank matrices, even in cases where conventional NMF is unable to correctly resolve the components. These advantages reduce the costs of HSI data analysis and of extracting hidden spectral information from experimental data using objective and statistical measures rather than empirical knowledge. The proposed method is applicable to any type of HSI dataset, such as that generated by Raman spectroscopy, infrared absorption, and time-of-flight mass spectroscopy. Future prospects would include investigating the ability of the present ARD-NMF scheme to correctly detect small amounts of significant phases.

**Acknowledgements** This work was in part supported by Grants-in-Aid for Scientific Research on Innovative Areas “Nano Informatics” (Grant No. 25106004, 26106510 and 16H00736), KIBAN-KENKYU A (Grant No. 26249096) and KIBAN-KENKYU B (Grant No. JP16H02866) from the Japan Society for the Promotion of Science (JSPS) and by Precursory Research for Embryonic Science and Technology (Grant No. JPMJPR16N6) from Japan Science and Technology Agency (JST).

## References

1. N. Bonnet, N. Brun, C. Colliex, *Ultramicroscopy* **77**, 97 (1999)
2. P. Trebbia, N. Bonnet, *Ultramicroscopy* **34**, 165 (1990)
3. M. Bosman, M. Watanabe, D.T.L. Alexander, V.J. Keast, *Ultramicroscopy* **106**, 1024 (2006)
4. C.M. Parish, L.N. Brewer, *Ultramicroscopy* **110**, 134 (2010)
5. S. Lichtert, J. Verbeeck, *Ultramicroscopy* **125**, 35 (2013)
6. J. Spiegelberg, J. Rusz, *Ultramicroscopy* **175**, 40 (2017)
7. N. Dobigeon, N. Brun, *Ultramicrosc.* **120**, 25 (2012)
8. N. Bonnet, D. Nuzillard, *Ultramicroscopy* **102**, 327 (2005)
9. J. Nascimento, J. Bioucas-Dias, I.E.E.E. Trans, *Geosci. Remote Sens.* **43**, 898 (2005)
10. N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tournet, A.O. Hero, I.E.E.E. Trans, *Signal Process.* **57**, 4355 (2009)
11. J. Spiegelberg, J. Rusz, T. Thersleff, K. Pelckmans, *Ultramicroscopy* **174**, 14 (2017)



12. S. Muto, T. Yoshida, K. Tatsumi, *Mater. Trans.* **50**, 964 (2009)
13. P.G. Kotula, M.R. Keenan, J.R. Michael, *Microsc. Microanal.* **9**, 1 (2003)
14. J.H. Wang, P.K. Hopke, T.M. Hanczewicz, S.L. Zang, *Anal. Chim. Acta* **476**, 93 (2003)
15. S. Muto, Y. Sasano, K. Tatsumi, T. Sasaki, K. Horibuchi, Y. Takeuchi, Y. Ukyo, *J. Electrochem. Soc.* **156**, A371 (2009)
16. S. Muto, K. Tatsumi, T. Sasaki, H. Kondo, T. Ohsuna, K. Horibuchi, Y. Takeuchi, *Electrochem. Solid State Lett.* **13**, A115 (2010)
17. Y. Kojima, S. Muto, K. Tatsumi, H. Oka, H. Kondo, K. Horibuchi, Y. Ukyo, *J. Power Sources* **196**, 7721 (2011)
18. S. Muto, K. Tatsumi, Y. Kojima, H. Oka, H. Kondo, K. Horibuchi, Y. Ukyo, *J. Power Sources* **205**, 449 (2012)
19. Y. Honda, S. Muto, K. Tatsumi, H. Kondo, K. Horibuchi, T. Sasaki, *J. Power Sources* **291**, 85 (2015)
20. T. Yoshida, S. Muto, J. Wakabayashi, *Mater. Trans.* **48**, 2580 (2007)
21. J. Senga, K. Tatsumi, S. Muto, Y. Yoshida, *J. Appl. Phys.* **118**, 115702 (2015)
22. K. Tatsumi, S. Muto, *J. Phys. Condens. Matter* **21**, 104213 (2009)
23. Y. Yamamoto, K. Tatsumi, S. Muto, *Mater. Trans.* **48**, 2590 (2007)
24. K. Tatsumi, S. Muto, Y. Yamamoto, H. Ikeno, S. Yoshioka, I. Tanaka, *Ultramicroscopy* **106**, 1019 (2006)
25. N.R. Lugg, G. Kothleitner, N. Shibata, Y. Ikuhara, *Ultramicroscopy* **151**, 150 (2015)
26. M. Shiga, K. Tatsumi, S. Muto, K. Tsuda, Y. Yamamoto, T. Mori, T. Tanji, *Ultramicroscopy* **170**, 43 (2016)
27. M. Shiga, S. Muto, K. Tatsumi, T. Tsuda, *Trans. Mat. Res. Soc. Jpn.* **41**, 333 (2016)
28. D.D. Lee, H.S. Seung, *Adv. Neural Inform. Process. Syst.* **13**, 556 (2001)
29. M.W. Berry, M. Browne, A.N. Langville, V.P. Pauca, R.J. Plemmons, *Comput. Stat. Data Anal.* **52**, 155 (2007)
30. A. Cichocki, *IEICE Trans. Fund. Electron. Commun. Comput.* **92**, 708 (2009)
31. K. Kimura, Y. Tanaka, M. Kudo, In *Proceedings of the 6th Asian Conference on Machine Learning*, Nha Trang, Vietnam, 26–28 Nov. 2014
32. V.Y.F. Tan, C. Fevotte, *I.E.E.E. Trans. Pat. Anal. Mach. Intell.* **35**, 1592 (2013)
33. J. Spiegelberg, S. Muto, M. Ohtsuka, K. Pelckmans, J. Rusz, *Ultramicrosc.* **182**, 205 (2017)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

