Sparse Modeling of EELS and EDX Spectral Imaging Data by Nonnegative Matrix Factorization

Motoki Shiga^{a*}, Kazuyoshi Tatsumi^b, Shunsuke Muto^b, Koji Tsuda^{c,g,h}, Yuta Yamamoto^d, Toshiyuki Mori^e and Takayoshi Tanji^f

^a Department of Electrical, Electronic and Computer Engineering, Gifu University, 1-1, Yanagido, Gifu 501-1193, Japan

^b Advanced Measurement Technology Center, Institute of Materials and Systems for Sustainability, Nagoya University, Chikusa-ku, Nagoya 464-8603, Japan

^c Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa 277-8561, Japan

^d High-Voltage Electron Microscope Laboratory, Institute of Materials and Systems for Sustainability, Nagoya University, Chikusa-ku, Nagoya 464-8603, Japan

^e Environment and Energy Materials Division, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan

^f Division of Materials Research, Institute of Materials and Systems for Sustainability, Nagoya University, Chikusa-ku, Nagoya 464-8603, Japan

^g Center for Materials Research by Information Integration, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba 305-0047, Japan

^h Biotechnology Research Institute for Drug Discovery, National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi Koto-ku, Tokyo 135-0064, Japan

* Corresponding author.

E-mail address: shiga_m@gifu-u.ac.jp (M. Shiga).

Keyword: spectral imaging, nonnegative matrix factorization, sparse modeling, automatic relevance determination, spatial orthogonality

Abstract:

Advances in scanning transmission electron microscopy (STEM) techniques have enabled us to automatically obtain electron energy-loss (EELS)/energy-dispersive X-ray (EDX) spectral datasets from a specified region of interest (ROI) at an arbitrary step width, called spectral imaging (SI). Instead of manually identifying the potential constituent chemical components from the ROI and determining the chemical state of each spectral component from the SI data stored in a huge three-dimensional matrix, it is more effective and efficient to use a statistical approach for the automatic resolution and extraction of the underlying chemical components. Among many different

statistical approaches, we adopt a non-negative matrix factorization (NMF) technique, mainly because of the natural assumption of non-negative values in the spectra and cardinalities of chemical components, which are always positive in actual data. This paper proposes a new NMF model with two penalty terms: i) an automatic relevance determination (ARD) prior, which optimizes the number of components, and ii) a soft orthogonal constraint, which clearly resolves each spectrum component. For the factorization, we further propose a fast optimization algorithm based on hierarchical alternating least-squares. Numerical experiments using both phantom and real STEM-EDX/EELS SI datasets demonstrate that the ARD prior successfully identifies the correct number of physically meaningful components. The soft orthogonal constraint is also shown to be effective, particularly for STEM-EELS SI data, where neither the spatial nor spectral entries in the matrices are sparse.

1. Introduction

Advances in scanning transmission electron microscopy (STEM) techniques, including the development of brighter electron sources, digitally controlled operation, detectors with higher sensitivity, and sophisticated on-line signal processing, have enabled us to obtain comprehensive electron energy-loss (EELS)/energy-dispersive X-ray (EDX) spectral datasets from a specified region of interest (ROI). The spectrometers collect a set of spectra, each from the sub-nanometre area of the sample, using sub-nanometric incident electron probe scanning over the two-dimensional ROI with a sub-nanometric step width. This method is known as spectral imaging (SI). Recent technical progress has reduced the typical data acquisition time 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 analysed has increased exponentially. The spectral intensity in the collected data is a function of the two-dimensional position (x, y) on the sample and absorption/emission energy E , expressed in a three-dimensional matrix form (data cube), $\mathbf{D}(x, y, E)$. A number of attempts to statistically extract a few basis spectra and their contributions at individual positions have been reported, with the assumption that the spectral intensity at each sample pixel is represented by a linear combination of the basis spectra associated with the underlying chemical states or phases. For the convenience of mathematical manipulation, a data cube is often transformed to a two-dimensional $N_{xy} \times N_{ch}$ matrix $\mathbf{X}$, where $N_{xy} = N_x \times N_y$ is the number of pixels, *i.e.* the number of scanning steps N_x (N_y) along the x (y) direction, and N_{ch} is the number of detector channels used for analysis. In this case, the experimental spectrum at position (x, y) is stored in a row of matrix $\mathbf{X}$. When there exist K essential spectra in the experimental data, where K is much smaller than the matrix size, *i.e.* $N_{xy} \times N_{ch} \gg N_{xy} \times K + N_{ch} \times K$, then the basis spectra and concentration coefficients can be stored in the $N_{xy} \times K$ matrix $\mathbf{C}$ and the $N_{ch} \times K$ matrix $\mathbf{S}$, respectively. The original data matrix $\mathbf{X}$ can then be approximated by the matrix product of $\mathbf{C}$ and $\mathbf{S}$ in the following form:

$$\mathbf{X} \approx \mathbf{C}\mathbf{S}^T, \quad (1)$$

where the superscript T denotes the transpose. Each column vector of $\mathbf{S}$ (referred to as the ‘loading’ in multivariate analysis) is a basis spectrum; each column vector of $\mathbf{C}$ (referred to as the ‘score’) is the cardinality of a chemical state or phase (referred to as a ‘component’) over the ROI positions, and each row vector of $\mathbf{C}$ is the contribution of the loading vectors at one sampling position. The spatial distribution of the i -th spectral component can then be reconstructed by selecting the i -th column of $\mathbf{C}$ and rearranging the elements back to the original two-dimensional order. There are a number of statistical approaches that aim to determine a solution to Eq. (1) using only the data matrix $\mathbf{X}$, which consists of a huge number of elements with a relatively small number of K . In the problem setting, a statistical approach is expected to extract plausible $\mathbf{C}$ and $\mathbf{S}$ with much higher signal-to-noise ratios (SNRs) than those picked from the small number of representative sampling points in a conventional manner (i.e., point-to-point analysis).

Among the various statistical approaches to solve Eq. (1) and obtain the unknown $\mathbf{C}$ and $\mathbf{S}$, 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 distributions of the corresponding basis vectors) in the order of significance, that is, in the order of the magnitude of eigenvalues through the singular value decomposition of $\mathbf{X}$. Trebbia et al. [2] and Bosman et al. [3] applied PCA to EELS-SI datasets, and not only detected exotic chemical bonding states in the samples, but also effectively filtered statistical noise from $\mathbf{X}$ by reconstructing the data matrix with a few basis spectra and their distributions. Parish and Brewer [4] studied the validity of PCA in a quantitative composition analysis of the constituent phases in their EDX-SI data. Note that, in their treatment, the phase overlapping areas were masked so as to be excluded in the quantification process; otherwise, the derived phase compositions could be biased with respect to the actual ones. These reports on PCA assume that each pixel contains a linear combination of principal components, because of the orthogonality condition intrinsic to PCA.

Using simulated atomic resolution EELS-SI data, Lichteret and Verbeeck [5] pointed out that, when the noise level is larger than the intensity of the signal of interest, the signal intensities spread out over a number of principal components, and are thus usually considered as noise. This behaviour seems statistically natural, but we would not notice this phenomenon in actual experimental data. Dobigeon and Brun [6] compared the results obtained by applying PCA, independent component analysis (ICA) [7], vertex component analysis (VCA) [8], and Bayesian linear unmixing (BLU) [9] to experimental EELS-SI 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-SI data.

Over more than a decade, our research group and Kotula’s group have applied the alternative method of non-negative matrix factorization (NMF), or multivariate curve resolution, to solve Eq. (1)

for the analysis of EELS-SI [10, 11]. We have taken this approach because NMF naturally restricts the score images and basis spectra to being non-negative. The methods mentioned above allow the images and spectra to take negative values, which hampers the direct physical interpretation of the resolved spectral profiles. We adopted the modified alternating least-square (MALS) fitting algorithm [12] to solve a number of problems such as the mapping of different phases in the degradation of Li battery cathodes [13-17] and the nitrogen chemical states in nitrogen-doped TiO₂ [18, 19]. 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 channelling effects [20-22]. The non-negative constraint imposed on the elements of S and C was significant in these applications, and the resolved spectral profiles were consistent with the theoretical ones predicted by first principles calculations [13-17, 20-22].

Matrix factorization by NMF with MALS is performed by simply minimizing the squared error, $\|X - CS^T\|^2$, which is the Euclidean distance between the experimental and reconstructed data. Our previous NMF suffers from the following intrinsic problems: First, unlike PCA, the optimization algorithm does not always converge to the optimal C and S , because the optimization problem has multiple local minima, *i.e.* non-convex. Hence, the best solution must be empirically selected by repeated optimizations of S and C from different initial configurations. Second, our previous NMF sets the number of spectral components in advance, which inevitably brings about a similar problem to the PCA case if the noise level is larger than the signal intensities. As the number of components increases, the error function naturally decreases. Thus, it is not the error function alone that identifies the essential number of physically meaningful components.

To overcome the above difficulties in STEM-EELS/EDX SI data analysis, this paper describes an extended NMF model that imposes the following penalty terms on C : i) a sparse penalty to optimize the number of components, called automatic relevance determination (ARD)[23], and ii) a spatial orthogonal constraint [24]. For the optimization of low rank matrices C and S , we further propose an algorithm based on hierarchical alternating least-squares (HALS) [25], which is more efficient than the matrix manipulation (MM) [26] used in [23]. Experimental results demonstrate that our new NMF successfully determines the number of relevant spectral components in experimental EDX-SI and EELS-SI datasets. We also demonstrate that the new penalty for spatial orthogonality casts the correct resolution for EELS-SI datasets in which chemical states are intricately mixed in each position.

2. Methodology

This section first describes our NMF model, some basic concepts, and update rules for low rank matrices and parameters. The detailed derivation of all update rules is described in the Supplementary Material. The experimental procedures for STEM-EELS/EDX-SI are described in detail in later subsections.

2.1 Nonnegative matrix factorization with Gaussian noise models

Let $\mathbf{X} \in \mathbf{R}_+^{N_{xy} \times N_{ch}}$ be a data matrix, where $\mathbf{R}_+$ is the set of all non-negative real numbers. NMF factorizes $\mathbf{X}$ into two thin matrices $\mathbf{C} \in \mathbf{R}_+^{N_{xy} \times K}$ and $\mathbf{S} \in \mathbf{R}_+^{N_{ch} \times K}$, where K is much smaller than both N_{xy} and N_{ch} . Thus, the factorization model is

$$\mathbf{X} = \mathbf{CS}^T + \boldsymbol{\varepsilon} \quad (2)$$

where $\boldsymbol{\varepsilon} \in \mathbf{R}^{N_{xy} \times N_{ch}}$ is a matrix of observation noise whose elements are statistically independent of each other. In our problem setting, $\mathbf{X}$ is observed, whereas $\mathbf{C}$ and $\mathbf{S}$ are not observed. The goal of NMF is to identify the optimal $\mathbf{C}$ and $\mathbf{S}$ under a suitable noise model $\boldsymbol{\varepsilon}$. One of the most common noise models is the Gaussian noise model. The negative log-likelihood function under the Gaussian model is given by

$$-\log p(\mathbf{X}|\bar{\mathbf{X}}, \sigma^2) = \frac{1}{2\sigma^2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2 + \frac{N_{xy}N_{ch}}{2} \log 2\pi\sigma^2 \quad (3)$$

where σ^2 is a variance parameter and $\bar{\mathbf{X}} = \mathbf{CS}^T$ is an estimate of the noiseless data matrix. To optimize $\mathbf{C}$ and $\mathbf{S}$, and thus minimize $\log p(\mathbf{X}|\bar{\mathbf{X}}, \sigma^2)$, we need to consider only the first term by neglecting σ^2 . Thus, the term to be optimized is a squared error function between observation $\mathbf{X}$ and estimation $\bar{\mathbf{X}}$:

$$D_{EU}(\mathbf{X}|\bar{\mathbf{X}}) = \frac{1}{2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2 \quad (4)$$

The minimization problem of Eq. (4) over both $\mathbf{C}$ and $\mathbf{S}$ is non-convex. Thus, the optimization algorithm must be run multiple times from different initializations, resulting in a heavy computational cost. To simplify this computational problem, optimization algorithms such as MM [26], ALS [27], and HALS [25] have been used. 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 apply the HALS framework to an optimization of our new NMF model, which is suitable for SI.

2.2 Probabilistic view of an NMF model with an automatic relevance determination prior

Optimizing the number of components using only the given data is an important practical problem. Maximum likelihood estimations (or estimations based on minimizing errors) cannot work in such cases, because they tend to over-fit the data. To avoid this over-fitting problem, a Bayesian estimation (or maximum a posteriori (MAP) estimation) with a prior distribution of scale parameters (relevance weights) is useful [23]. The process of choosing only the important components is called automatic relevance determination (ARD).

To perform ARD in NMF, we assume a prior distribution for $\mathbf{C}$ using an exponential distribution with a scale parameter λ_k for the probability density of column k of $\mathbf{C}_{\cdot k}$:

$$p(C_{ik}|\lambda_k) = \frac{1}{\lambda_k} \exp\left(-\frac{C_{ik}}{\lambda_k}\right), \quad i = 1, \dots, N_{xy}, \quad k = 1, \dots, K \quad (5)$$

The density distribution generates non-negative random values with a large probability density around zero, resulting in the sparseness of $\mathbf{C}$. For the prior distribution of λ_k , we assume an inverse-Gamma distribution:

$$p(\lambda_k|a, b) = \frac{b^a}{\Gamma(a)} \lambda_k^{-(a+1)} \exp\left(-\frac{b}{\lambda_k}\right), \quad k = 1, \dots, K, \quad (6)$$

where a and b are hyper-parameters to adjust the sparseness of λ_k . The density distribution of each column in $\mathbf{S}$, $p(\mathbf{S}_{\cdot k})$, is uniformly distributed on the unit hyper-sphere in $\mathbf{R}_+^{N_{ch}}$. Using Eqs. (3), (5), and (6), the negative log-likelihood function of our NMF model is given by:

$$\begin{aligned} L(\mathbf{C}, \mathbf{S}, \boldsymbol{\lambda}, \sigma^2) &= -\log p(\mathbf{X}|\bar{\mathbf{X}}, \sigma^2) - \sum_{i=1}^{N_{xy}} \sum_{k=1}^K \log p(C_{ik}|\lambda_k) - \sum_{k=1}^K \log p(\mathbf{S}_{\cdot k}) - \sum_{k=1}^K \log p(\lambda_k|a, b) \\ &= \frac{N_{xy}N_{ch}}{2} \log 2\pi\sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2 + \sum_{k=1}^K \frac{1}{\lambda_k} \left(b + \sum_{i=1}^{N_{xy}} C_{ik} \right) \\ &\quad + (N_{xy} + a + 1) \sum_{k=1}^K \log \lambda_k + K(\log \Gamma(a) - a \log b), \end{aligned} \quad (7)$$

For the optimization of $\mathbf{C}$, $L(\mathbf{C}, \mathbf{S}, \boldsymbol{\lambda}, \sigma^2)$ is a penalized likelihood function with L_1 norms, resulting in a sparse matrix $\mathbf{C}$. The NMF minimizing $L(\mathbf{C}, \mathbf{S}, \boldsymbol{\lambda}, \sigma^2)$ is referred to as ARD-NMF.

Because the simultaneous optimization of $L(\mathbf{C}, \mathbf{S}, \boldsymbol{\lambda}, \sigma^2)$ over all $\mathbf{C}$, $\mathbf{S}$, $\boldsymbol{\lambda}$, and σ^2 is non-convex, multiple optimizations from different initial configurations are required. To update $\mathbf{C}$ and $\mathbf{S}$, we use HALS [25], which updates each column $\mathbf{C}_{\cdot k}$ and $\mathbf{S}_{\cdot k}$ alternately. Applying some algebra to the minimization of $L(\mathbf{C}, \mathbf{S}, \boldsymbol{\lambda}, \sigma^2)$, we obtain the following update rule for $\mathbf{C}_{\cdot k}$:

$$\mathbf{C}_{\cdot k} = \left[\mathbf{X}^{(k)} \mathbf{S}_{\cdot k} - \frac{\sigma^2}{\lambda_k} \right]_+, \quad k = 1, \dots, K \quad (8)$$

where the operator $[\mathbf{A}]_+$ replaces all negative values in $\mathbf{A}$ with zeros. This can be calculated by $[\mathbf{A}]_+ = \{\mathbf{A} + \text{abs}(\mathbf{A})\}/2$, where the abs function outputs a matrix consisting of the absolute values of the matrix elements in $\mathbf{A}$. Matrix $\mathbf{X}^{(k)}$ is defined by

$$\mathbf{X}^{(k)} = \mathbf{X} - \mathbf{C} \mathbf{S}^T + \mathbf{C}_{\cdot k} \mathbf{S}_{\cdot k}^T. \quad (9)$$

The second term in Eq. (8) comes from the ARD prior, while the optimization for the basic NMF [25] has only the first term.

An update rule for $\mathbf{S}_{\cdot k}$ is given by

$$\mathbf{S}_{\cdot k} = \frac{\tilde{\mathbf{s}}_k}{\|\tilde{\mathbf{s}}_k\|_2}, \quad k = 1, \dots, K \quad (10)$$

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

$$\tilde{\mathbf{s}}_k = \left[(\mathbf{X}^{(k)})^T \mathbf{C}_{\cdot k} \right]_+, \quad k = 1, \dots, K \quad (11)$$

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

$$\lambda_k = \frac{b + \sum_{i=1}^{N_{xy}} C_{ik}}{N_{xy} + a + 1}, \quad k = 1, \dots, K \quad (12)$$

$$\sigma^2 = \frac{1}{N_{xy} N_{ch}} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2 \quad (13)$$

The hyper-parameter b can be set using an approximate empirical estimator [23] as follows:

$$b = \frac{(a-1)\sqrt{N_{ch}}}{K} \frac{1}{N_{xy} N_{ch}} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} X_{ij}, \quad (14)$$

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 $L(\mathbf{C}, \mathbf{S}, \lambda, \sigma^2)$. After the optimization of ARD-NMF, the relevance of components $\lambda_k, k = 1, \dots, K$ is given. Because the values of redundant components cannot be exactly zero, we estimate the relevance of each component by

$$M_k = \frac{\lambda_k - \min_k \lambda_k}{\max_k \lambda_k}, \quad k = 1, \dots, K \quad (15)$$

This leaves only components with $M_k \geq \theta$ as relevant components. The threshold θ should be adjusted manually by checking the distribution of $\lambda_k, k = 1, \dots, K$.

2.3 Optimization algorithms with spatial orthogonal constraints

In SI data analysis, one objective is to automatically decompose spectra mixed from different chemical components into those of individual components. However, decomposition by a basic NMF model, *e.g.* minimization of Eq. (3) or (7), often outputs unresolved spectra, which are still a mixture of spatially close or overlapped chemical components. To handle these unresolved problems for the optimization of each column in $\mathbf{C}$, we introduce the orthogonal constraint [24]. Because the orthogonality constraint is too strict, we relax this constraint with a weight parameter, called a soft-orthogonal constraint, which is formulated as follows:

$$\frac{1}{2\sigma^2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} \left([\mathbf{X}^{(k)}]_{ij} - [\mathbf{C}_{\cdot k} \mathbf{S}_{\cdot k}^T]_{ij} \right)^2 + w \cdot \xi_k \mathbf{C}_{\cdot k}^T \mathbf{C}^{(k)} \quad \text{s.t.} \quad \|\mathbf{C}_{\cdot k}\|_2 = 1, \quad (16)$$

with

$$\mathbf{c}^{(k)} = \sum_{m \neq k} \mathbf{c}_{\cdot m}, \quad k = 1, \dots, K \quad (17)$$

where w , $0 \leq w \leq 1$, is a parameter to adjust the orthogonal constraint and ξ_k is the Lagrange multiplier for the exact orthogonal constraint of $\mathbf{C}$. When $w = 1$, the optimized $\mathbf{C}$ is an orthogonal matrix in which no chemical components overlap. When $w = 0$, the optimized components in $\mathbf{C}$ may be heavily overlapped. The optimal value of w depends on the situation, such as the spatial resolution of the data (step width of STEM-SI) and localization of chemical states. Thus, the optimal value of w must be chosen according to the measurement level. Applying some algebra to Eq. (16), we can derive the Lagrange multiplier ξ_k . Substituting this ξ_k , we obtain the update rule:

$$\mathbf{c}_{\cdot k} = \left[\mathbf{X}^{(k)} \mathbf{S}_{\cdot k} - w \frac{\mathbf{c}^{(k)T} \mathbf{X}^{(k)} \mathbf{S}_{\cdot k}}{\mathbf{c}^{(k)T} \mathbf{c}^{(k)}} \mathbf{c}^{(k)} \right]_+, \quad k = 1, \dots, K \quad (18)$$

The second term weighted by w comes from the orthogonal constraint. Unlike our original setting, each column of $\mathbf{C}$ should be normalized after applying Eq. (18) [24]. Thus, we skip the normalization of $\mathbf{S}$, and the update is given by

$$\mathbf{S}_{\cdot k} = \left[(\mathbf{X}^{(k)})^T \mathbf{c}_{\cdot k} \right]_+, \quad k = 1, \dots, K \quad (19)$$

Figure 1 gives the pseudo-code of this NMF, which we call SO-NMF.

2.4 Optimization algorithm for $\mathbf{C}$ with both ARD and spatial orthogonal constraint

We introduce both the soft orthogonal constraint and the ARD effect using both penalty terms as follows:

$$\mathbf{c}_{\cdot k} = \left[\mathbf{X}^{(k)} \mathbf{S}_{\cdot k} - \frac{\sigma^2}{\lambda_k} - w \frac{\mathbf{c}^{(k)T} (\mathbf{X}^{(k)} \mathbf{S}_{\cdot k} - \sigma^2 / \lambda_k)}{\mathbf{c}^{(k)T} \mathbf{c}^{(k)}} \mathbf{c}^{(k)} \right]_+, \quad k = 1, \dots, K \quad (20)$$

In this update, $\mathbf{c}_{\cdot k}$ is not renormalized to reduce the effect of the orthogonal constraint for irrelevant components. We propose Eq. (20) as an update rule for $\mathbf{c}_{\cdot k}$ when the orthogonal constraint is necessary. Figure 2 shows the pseudo-code of our proposed NMF algorithm, which we call ARD-SO-NMF. In the special case without the orthogonal constraint, *i.e.* $w = 0$, this is simply ARD-NMF. Lines 12–20 merge components whose spectra are almost the same. This operation is necessary to ensure the correct number of components, because the orthogonality condition with $w > 0$ tends to split components, even when the spectra are exactly the same. This side-effect of the orthogonality condition was investigated using both phantom and real data. These results are shown in the Supplementary Material. (After the acceptance of this manuscript, our MATLAB codes will be available.)

2.5 Another variant of our NMF with another ARD prior

Similar to [24], the half-normal distribution

$$p(C_{ik}|\lambda_k) = \left(\frac{2}{\pi\lambda_k}\right)^{\frac{1}{2}} \exp\left(-\frac{C_{ik}^2}{2\lambda_k}\right), \quad i = 1, \dots, N_{xy}, \quad k = 1, \dots, K \quad (21)$$

can be used as an ARD prior. When a half-normal distribution is used in Eq. (5), the negative log-likelihood becomes

$$\begin{aligned} L(\mathbf{C}, \mathbf{S}, \boldsymbol{\lambda}, \sigma^2) = & \frac{N_{xy}N_{ch}}{2} \log 2\pi\sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \bar{X}_{ij})^2 + \sum_{k=1}^K \frac{1}{\lambda_k} \left(b + \sum_{i=1}^{N_{xy}} \frac{C_{ik}^2}{2} \right) \\ & + \left(\frac{N_{xy}}{2} + a + 1 \right) \sum_{k=1}^K \log \lambda_k + \frac{KN_{xy}}{2} \log \frac{2}{\pi} + K(\log \Gamma(a) - a \log b), \end{aligned} \quad (22)$$

Minimizing the likelihood, we derive the following update rules for column k in matrix $\mathbf{C}$ and λ_k :

$$\mathbf{C}_{\cdot k} = (\sigma^{-2} + \lambda_k^{-1})^{-1} \left[\sigma^{-2} \mathbf{X}^{(k)} \mathbf{S}_{\cdot k} - w \frac{\mathbf{c}^{(k)T} \mathbf{X}^{(k)} \mathbf{S}_{\cdot k}}{\sigma^2 \mathbf{c}^{(k)T} \mathbf{c}^{(k)}} \mathbf{c}^{(k)} \right]_+, \quad k = 1, \dots, K \quad (23)$$

$$\lambda_k = \frac{b + \frac{1}{2} \sum_{i=1}^{N_{xy}} C_{ik}^2}{N_{xy}/2 + a + 1}, \quad k = 1, \dots, K \quad (24)$$

Replacing Eq. (20) with Eq. (23) and Eq. (12) with Eq. (24) in Figure 2, a further NMF algorithm with a half normal prior can be implemented.

2.6 Experimental procedures

A phantom dataset of an atomic resolution EELS spectrum image showing Mn- $L_{2,3}$ of Mn_3O_4 was prepared according to the method described in [5], where phantom data of an atomic resolution EELS spectrum image of h-BN was analyzed by a fine PCA approach. In our system of Mn_3O_4 , where trivalent manganese (Mn^{3+}) and divalent manganese (Mn^{2+}) occupy the octahedral (Mn_O) and tetrahedral (Mn_T) sites, respectively, the two site-specific spectra of Mn- $L_{2,3}$ were calculated [22] by an ab-initio configuration interaction method based on fully relativistic molecular orbitals [28-30]. The spatial distribution of the two chemical components was represented by the superposition of Gaussian distributions centered at the projected Mn atomic sites. The weight of the Gaussian distribution at each site was proportional to the number of Mn atoms in the column at each site. The present phantom data considered no dynamical electron scattering effects (particularly electron channeling effects), and the core-loss spectra are localized on the respective Mn columns alone. Figure 3(a) and (b) show the generated distributions and spectra of two components, respectively. We used these data as the true matrix $\mathbf{C} \in \mathbf{R}_+^{N_{xy} \times 2}$ and $\mathbf{S} \in \mathbf{R}_+^{N_{ch} \times 2}$ for a phantom dataset. Using these matrices, an SI data matrix with noise was generated by $\mathbf{X} = [\mathbf{C}\mathbf{S}^T + \boldsymbol{\varepsilon}]_+$, where $\boldsymbol{\varepsilon} \in \mathbf{R}^{N_{xy} \times N_{ch}}$ is a noise matrix whose elements were generated from a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = r_{\text{noise}} \cdot S_{\text{max}}$. The value S_{max} is the maximum value in $\mathbf{S}$, and the noise

level r_{noise} was set to various values to evaluate the robustness of our method.

An experimental STEM-EELS SI dataset was prepared by atomic resolution EELS-SI of Mn_3O_4 . A polycrystalline Mn_3O_4 with a spinel crystal structure was obtained, and a TEM sample was prepared as described in [22]. We measured the $Mn-L_{2,3}$ SI using a JEOL ARM-200F operated at 200 kV, with a Gatan Quantum EELS having an energy dispersion of 0.5 eV/channel. The average full width at half maximum (FWHM) of the zero-loss peak (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 4(a)–(b) shows the schematic projected structure of the Mn_T , Mn_O , and O (oxygen) columns along the present incident beam direction and the theoretical site-specific spectra, respectively. These data are more difficult to analyse, because the inner shell excitation is delocalized by a certain distance and the neighbouring atomic columns simultaneously contribute to the spectrum intensity at a sampling point due to electron channelling effects [31].

Another real dataset was also acquired from a cross-sectional TEM sample of a Si diode, prepared by a focused ion beam technique (FIB). We measured SI data for $Si-L_{2,3}$ including ZLP using a JEOL JEM-1000K RS S/TEM operated at 1000 kV, with a Gatan Quantum equivalent EEL spectrometer whose energy dispersion was set to 0.2 eV/channel. The sample thickness of the measured area was estimated at 0.1 μm 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 modelled by a power law was subtracted to isolate the $Si-L_{2,3}$ spectrum. Figure 5 shows a manually validated component map and spectra.

Another phantom dataset of a Si diode was generated using the output of NMF with $K = 3$ for the real Si diode dataset. A Savitzky–Golay smoothing filter using a third-order polynomial and frame size of 31 was then applied to the spectral components to remove noise. To generate a non-overlapping component distribution, intensities lower than the pre-set threshold (10% of the maximum intensity) were set to zero for each resolved image. The true matrixes $\mathbf{C}$ and $\mathbf{S}$ used as a phantom dataset were then generated. These data are essentially the same as the reference component distribution and spectra in Figure 5. Note that matrix $\mathbf{C}$ satisfies $\mathbf{C}^T \mathbf{C} = \mathbf{I}$, because the overlaps between components have been removed. Thus, our orthogonal penalty is expected to work well. Similar to the Mn_3O_4 phantom datasets, observation noise was added to generate a Si diode phantom dataset.

An EDX-SI dataset was acquired from a sintered ceramic composite of Y-doped ZrO_2 – $LaSrMnO_3$, which exhibits a distinct composition variation across the electron transparent sample area. A thin film for TEM was prepared by FIB. We measured the EDX-SI using a JEOL 2100F S/TEM operated at 200 kV, equipped with a JEOL EDX silicon drift detector, Dry SD60GV. Figure 6

shows an ADF image and typical counts (spectra) at representative points.

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

3. Results

3.1 Results of optimizing the number of components by ARD-NMF and ARD-SO-NMF

3.1.1 STEM-EELS phantom dataset of Mn_3O_4

We first evaluated ARD-NMF (*i.e.* where the orthogonal constraint is not imposed, $w = 0$) with $K = 10$ for the phantom data of Mn_3O_4 , which are atomic resolution data with slightly overlapping components. We used datasets with three noise levels $r_{\text{noise}} = 0.01, 0.1, 0.2$ to evaluate the robustness of ARD-NMF. These results are shown in Figure 7(a)–(c). Panel (4) in these figures shows that ARD-NMF can reduce the cost function (the negative log-likelihood in Eq. (7)) at each update step and converge to a local minimum. Panel (2) of Figure 7(a) and (b) demonstrates that our proposed NMF correctly selected the number of components for SI datasets with low noise levels ($r_{\text{noise}} = 0.01$ and 0.1). However, as the noise levels increase, the component distributions and their spectra begin to collapse, as shown in the case of $r_{\text{noise}} = 0.1$. A smaller component then disappears when the noise level is increased ($r_{\text{noise}} = 0.2$). These results are natural, because an ARD approach that evaluates whether each component is meaningful or noise can detect larger components than the noise level in SI data.

The ARD-SO-NMF results (with $w = 0.1$) for three noise levels are shown in Figure 8(a)–(c). These results demonstrate that the orthogonal penalty acts to separate components spatially and detect smaller components more accurately. In the results for $r_{\text{noise}} = 0.1$, two components were separated more clearly than with $w = 0$. As a result, the spectra produced by ARD-SO-NMF were also identified more accurately, whereas the spectra with $w = 0$, shown in Figure 7(b), are unnaturally caved.

3.1.2 Atomic resolution STEM-EELS of Mn_3O_4

The ARD-NMF technique was then applied to the experimental atomic resolution $\text{Mn-}L_{2,3}$ SI data from the Mn_3O_4 spinel sample. Figures 9(a-1)–(a-4) show the ARD-NMF results with $K = 10$. As shown in Figure 9(a-2), 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. The spatial distributions of the resolved component 1 and 2 shown in Figure 9(a-1) are basically consistent with the projected Mn_O and Mn_T atom positions in Figure 4(a), respectively, and the relative chemical shifts 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 (Figure 9(a-3)) are inconsistent with the expected theoretical profiles, with component 2 exhibiting physically unnatural intensity drops 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 are heavily overlapped. To overcome this problem, we applied ARD-SO-NMF with $w = 0.01$. Figure 9(b) demonstrates that, due to the orthogonal condition, components were separated more clearly. In particular, overlaps between the first components and others were better resolved, as shown in Figure 9(b-1), and the third component was detected more clearly.

We now focus on the additional third component. The spatial distribution of this component was found to be localized on the projected oxygen atom positions, which should be due to the electron channelling effect [31], where the incident electron wave function is also propagating along the neighbouring Mn columns in samples above 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 two Mn_O and one Mn_T atoms.

3.1.3 STEM-EELS phantom dataset of a silicon device

We further evaluated both ARD-NMF ($w = 0$) and ARD-SO-NMF ($w = 0.1$) with $K = 10$ using the Si phantom datasets with three noise levels $r_{\text{noise}} = 0.01, 0.1, 0.2$. Figures 10 and 11 show the results given by NMFs. Figure 10 demonstrates that, even when the noise level is small, the spatial distributions decomposed by ARD-NMF overlap and their spectra drop unnaturally. These results are characteristic of a basic NMF model, which generates a sparse decomposition in both spectra and spatial distributions. In such a situation with a non-sparse structure, the decomposition of NMF that minimizes the likelihood in Eq. (7) is not unique. Thus, NMF should reach one of the local optima. On the other hand, Figure 11 demonstrates that ARD-SO-NMF decomposed both the spatial distribution and their spectra more accurately, especially for the SI data with $r_{\text{noise}} = 0.01$ and 0.1 . Because these phantom data are sparse in the spatial distribution but non-sparse in the spectra, inducing sparseness in only the spatial distribution performs well for these datasets. However, when the noise level is much larger, *e.g.* noise level $r_{\text{noise}} = 0.2$ in Figure 11(c), even ARD-SO-NMF cannot decompose the smallest component, in which the intensity is less than the noise, from the other components. These results are natural in a statistical sense.

3.1.4 XSTEM-EELS data from a silicon device

We applied both ARD-NMF and ARD-SO-NMF with $K = 10$ to the EELS-SI data for Si- $L_{2,3}$ energy-loss near edge structure (ELNES) obtained from a cross-sectional Si diode sample. The structure of this sample is illustrated in Figure 5(a), and the reference spectra from standard samples are shown in Figure 5(b). These spectra are not sparse, that is, non-zero values range over the energy-loss axis similar to the Si phantom data. 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 Figure 12. Figure 12(a-2) shows that ARD-NMF selected four components, whereas the reference number is three. From Figure 12(a-1), we can see that the generated component distribution of components 1 and 3 are heavily overlapped, whereas the spectra of these components do not overlap. This result was due to a basic NMF property that induces a sparse decomposition on both spatial and spectral matrices.

The ARD-SO-NMF results (with $w = 0.1$) are shown in Figure 12(b). Figure 12(b-3) shows that the identified spectra are consistent with the reference spectra shown in Figure 5(a). The spatial distributions of components given by ARD-SO-NMF (Figure 12(b-1)) are clearly separated, whereas the distributions given by ARD-NMF (Figure 12(a-1)) are heavily overlapped. This difference demonstrates the effect of the orthogonal constraint. In this case, ARD-SO-NMF has chosen three components whose spectra are consistent with the reference spectra, whereas the spectra given by ARD-NMF have unnatural drops. Thus, these results suggest that the method can effectively detect subtle spectral differences by introducing the orthogonal constraint.

3.1.5 STEM-EDX data

We further examined our ARD-NMF (*i.e.* where the orthogonal constraint is not imposed, $w = 0$) using the EDX-SI data of a Y-doped $\text{ZrO}_2(\text{YSZ})\text{--LaSrMnO}_3(\text{LSM})$ ceramic composite material. Typical sets of EDX spectra extracted from the single phase areas (Areas A and B) are shown in Figure 6, where the maximum peak intensity is no more than 10 counts, having a typical sparse feature that is suitable for testing the relevance of the proposed method. Starting with 10 components, only two had survived after the optimization algorithm terminated, as shown in Figure 13(a-2). The distribution of each identified component shown in Figure 13(a-1) is consistent with the elemental distributions of: 1) the union of Zr and Y, and 2) the union of La, Mn, and Sr in Figure 13(b). The identified spectra shown in Figure 13(a-3) consist of sets of peaks, each reflecting the correct composition of YSZ or LSM in Figure 13(b). This indicates that our ARD-NMF identified the constituent phases correctly for this STEM-EDX data.

The present method effectively removed the statistical noise in the resolved spectra (Figure 13(a-3)), and the score images (Figure 13(a-1)) exhibit no artificial mixing of the two spectral components. Note that a 10 nm layer of LSM (Figure 13(b)) covers the YSZ substrate surface; this can be seen more clearly here than in the elemental maps.

3.2 Spatial orthogonal constraint on STEM-EELS data

We additionally 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 EELS datasets described in the previous

sections. Because neither the spatial distribution maps nor spectra in the EELS datasets are sparse, the basic 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.

3.2.1 XSTEM-EELS data from a silicon device

The method was first applied to the dataset from the Si diode sample described in Section 3.1.4. The number of components in SO-NMF was fixed to $K = 3$, which is the number of reference components. In the result with $w = 0$ (no orthogonal constraint) shown in Figure 14(c), the third spectral component exhibits unnatural intensity drops at 110 eV, where a sharp peak from the first spectral component is overlaid, as shown in the fourth column of the first row of Figure 14(c). This can happen in EELS-SI under certain conditions [15]. This intensity drop disappears when spatial orthogonality ($w \geq 0.01$) is included, as seen in the fourth column of Figure 14(c). In the cases of Figure 14(c), a slight cross-talk between the second and third components remains in both the spatial distribution maps and spectra for $w = 0$. The resolved spectral profiles and their spatial distributions are almost the same for $w \geq 0.01$, which effectively reproduces the spectra and expected spatial distributions.

3.2.2 Atomic resolution STEM-EELS of Mn_3O_4

Next, we validated the method using the atomic resolution Mn- $L_{2,3}$ SI data from the Mn_3O_4 spinel sample described in Section 3.1.2. The number of component in SO-NMF was fixed to $K = 3$, which is the number of components determined by ARD-SO-NMF. The SO-NMF results for $0 \leq w \leq 1$ are shown in Figure 15(c), with the score images in the first, second, and third columns and the resolved spectral profiles in the fourth column. In the case of no spatial orthogonality ($w = 0$), the resolved spectral profiles are inconsistent with the expected theoretical profiles, with the peak at around 655 eV of component 2 shifted to the left. Component 3 exhibits a physically unnatural intensity drop at the distinct peak positions of component 1, as for the case of no orthogonality (Figure 15(c), top-right figure). With a small spatial orthogonality ($w = 0.01$) included, the spectral shapes converged to those consistent with the theoretical spectra and the additional component localized on the oxygen columns, as discussed in Section 3.1.2. 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 increasingly orthogonal to one another, whereas the resolved spectral shapes converge to one form. For $w \geq 0.5$, the spatial distributions are far from the actual projected structures, though the resolved spectral shapes are essentially identical.

4. Discussion

The ARD-SO-NMF and ARD-NMF proposed in this study were able to optimize the number of spectral components for both EDX and EELS datasets. An additional orthogonal constraint was required when neither the spatial distribution nor spectra were sparse. Such a constraint is offered by the proposed ARD-SO-NMF. Our NMF extracts realistic spectral components from the EELS 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, the NMF works differently for these datasets. 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 electron energy levels in solids are rather continuously distributed. The spectral components of EDX are mostly sparse and orthogonal in the energy axis, whereas those in EELS are not. NMF with only an error function as the objective function will prefer the orthogonal basis spectra in the energy axis because of their completeness. On the other hand, NMF with spatial orthogonality better models the practical situation in which an EELS-SI dataset is assumed more orthogonal in space than in energy.

There are several local minima in the likelihood functions. The type of NMF algorithm appears to determine which minimum we eventually reach, although it is not mathematically possible to prove the dependence. Moreover, because of the computational cost, it is hard to obtain all of the local minima, even when we apply the spatial orthogonality constraint and sparse priors for the ARD effect, *i.e.* small value of K , in the data matrix. In the present NMF method, systematically varying the weight of the spatial orthogonality in the object function can minimize and extract the objective function of some particular solutions. In both of the EELS examples presented herein, an increase in w caused the resolved components to be distributed in more separate manners 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 more accurate composition 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 SI 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 theoretical spectra predicted by first principles calculations or reference experimental spectra, by changing the value of w systematically and understanding how 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.

It can be controversial to subtract the spectral background before the statistical processing, depending on what type of spectral data are being treated. Although it is generally considered that background subtraction results in the loss of important spectral information, we believe that

background subtraction is necessary in the present framework because our NMF assumes that no background structure is incorporated. As demonstrated in the supplementary material, our proposed NMF did not provide the expected correct results for STEM-EELS SI without background subtraction, which thus poses further questions as to how a background structure can be incorporated into our model.

5. Summary

We have proposed a new multivariate curve resolution method based on NMF with two penalty terms: i) an ARD prior to optimize the number of components and ii) a soft orthogonal constraint to effectively resolve overlapping spectra. Validations using experimental STEM-EDX/EELS SI data demonstrated that the ARD prior successfully chose the correct number of physically interpretable spectral components. The soft orthogonal constraint worked well for STEM-EELS SI data that were sparse in neither spatial nor 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 the conventional NMF is unable to correctly resolve the components. These advantages reduce the costs of SI data analysis and mining hidden spectral information in experimental data using objective and statistical measures rather than empirical knowledge. The proposed method is applicable to any type of SI dataset, such as Raman spectroscopy, infrared absorption, and time-of-flight mass spectroscopy.

Acknowledgments

This work is in part supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research) #26106510, #25870322 and Toyota Physical & Chemical Research Institute Scholar (MS), JSPS KAKENHI #25106004, #26120518, and #26249096 (KTa & SM), and by JST CREST, JST ERATO, RIKEN PostK, NIMS MI2I, and KAKENHI #15H05711 (KTs).

References

- [1] N. Bonnet, N. Brun, C. Colliex, Extracting information from sequences of spatially resolved EELS spectra using multivariate statistical analysis, *Ultramicroscopy*, 77 (1999) 97–112.
- [2] P. Trebbia, N. Bonnet, EELS elemental mapping with unconventional methods I. Theoretical basis: Image analysis with multivariate statistics and entropy concepts, *Ultramicroscopy* 34 (1990) 165–178.
- [3] M. Bosman, M. Watanabe, D.T.L. Alexander, V.J. Keast, Mapping chemical and bonding information using multivariate analysis of electron energy-loss spectrum images, *Ultramicroscopy*, 106 (2006) 1024–1032.
- [4] C.M. Parish, L.N. Brewer, Multivariate statistics applications in phase analysis of STEM-EDS spectrum images, *Ultramicroscopy*, 110 (2010) 134–143.

- [5] S. Lichtert, J. Verbeeck, Statistical consequences of applying a PCA noise filter on EELS spectrum images, *Ultramicroscopy*, 125 (2013) 35–42.
- [6] N. Dobigeon, N. Brun, Spectral mixture analysis of EELS spectrum-images, *Ultramicroscopy*, 120 (2012) 25–34.
- [7] N. Bonnet, D. Nuzillard, Independent component analysis: A new possibility for analysing series of electron energy loss spectra, *Ultramicroscopy*, 102 (2005) 327–337.
- [8] J. Nascimento, J. Bioucas-Dias, Vertex component analysis: A fast algorithm to unmix hyperspectral data, *IEEE Trans. on Geoscience and Remote Sensing* 43 (2005) 898–910.
- [9] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tournet, A.O. Hero, Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery, *IEEE Trans. on Signal Processing* 57 (2009) 4355–4368.
- [10] S. Muto, T. Yoshida, K. Tatsumi, Diagnostic Nano-Analysis of Materials Properties by Multivariate Curve Resolution Applied to Spectrum Images by S/TEM-EELS, *Mater. Trans.* 50 (2009) 964–969.
- [11] P.G. Kotula, M.R. Keenan, J.R. Michael, Automated Analysis of SEM X-Ray Spectral Images: A Powerful New Microanalysis Tool, *Microscopy and Microanalysis* 9 (2003) 1–17
- [12] J.H. Wang, P.K. Hopke, T.M. Hancewicz, S.L. Zang, Application of modified alternating least squares regression to spectroscopic image analysis, *Anal. Chim. Acta.* 476 (2003) 93–109.
- [13] S. Muto, Y. Sasano, K. Tatsumi, T. Sasaki, K. Horibuchi, Y. Takeuchi, Y. Ukyo, Capacity-Fading Mechanisms of LiNiO₂-Based Lithium-Ion Batteries II. Diagnostic Analysis by Electron Microscopy and Spectroscopy, *J. Electrochem. Soc.*, 156, (2009) A371–A377.
- [14] S. Muto, K. Tatsumi, T. Sasaki, H. Kondo, T. Ohsuna, K. Horibuchi, Y. Takeuchi, Mapping of Heterogeneous Chemical States of Lithium in a LiNiO₂-Based Active Material by Electron Energy-Loss Spectroscopy, *Electrochem. Solid State Lett.*, 13 (2010) A115–A117.
- [15] Y. Kojima, S. Muto, K. Tatsumi, H. Oka, H. Kondo, K. Horibuchi, Y. Ukyo, Degradation analysis of a Ni-based layered positive-electrode active material cycled at elevated temperatures studied by scanning transmission electron microscopy and electron energy-loss spectroscopy, *J. Power Sources*, 196 (2011) 7721–7727.
- [16] S. Muto, K. Tatsumi, Y. Kojima, H. Oka, H. Kondo, K. Horibuchi, Y. Ukyo, Effect of Mg-doping on the degradation of LiNiO₂-based cathode materials by combined spectroscopic methods, *J. Power Sources*, 205 (2012) 449–455.
- [17] Y. Honda, S. Muto, K. Tatsumi, H. Kondo, K. Horibuchi, T. Sasaki, Microscopic mechanism of path-dependence on charge–discharge history in lithium iron phosphate cathode analysis using scanning transmission electron microscopy and electron energy-loss spectroscopy spectral imaging, *J. Power Sources*, 291 (2015) 85–94.
- [18] T. Yoshida, S. Muto, J. Wakabayashi, Depth-Resolved EELS and Chemical State Mapping of

N⁺-Implanted TiO₂ Photocatalyst, Mater. Trans., 48 (2007) 2580–2584.

- [19] J. Senga, K. Tatsumi, S. Muto, Y. Yoshida, Effects of nitrogen-related defects on visible light photocatalytic response in N⁺ implanted TiO₂: A first-principles study, J. Appl. Phys., 118 (2015) 115702.
- [20] K. Tatsumi, S. Muto, Local electronic structure analysis by site-selective ELNES using electron channeling and first-principles calculations, J. Phys.: Condens. Matter 21(2009) 104213.
- [21] Y. Yamamoto, K. Tatsumi, S. Muto, Site-Selective Electronic Structure of Aluminum in Oxide Ceramics Obtained by TEM-EELS Analysis Using the Electron Standing-Wave Method, Mater. Trans. 48 (2007) 2590–2594.
- [22] K. Tatsumi, S. Muto, Y. Yamamoto, H. Ikeno, S. Yoshioka, I. Tanaka, Site-specific electronic structure analysis by channeling EELS and first-principles calculations, Ultramicroscopy, 106 (2006) 1019–1023.
- [23] V.Y.F. Tan, C. Fevotte, Automatic Relevance Determination in Nonnegative Matrix Factorization with the beta-Divergence, IEEE Trans. on Pat. Anal. Mach. Intell. 35, (2013) 1592–1605.
- [24] K. Kimura, Y. Tanaka, M. Kudo, A Fast Hierarchical Alternating Least Squares Algorithm for Orthogonal Nonnegative Matrix Factorization, Proc. of the 6th Asian Conference on Machine Learning (2014) 129–141.
- [25] A. Cichocki, Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations, IEICE Trans. Fund. Elec. Commun. Comput., 92 (2009) 708–721.
- [26] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, Adv. Neural Inform. Process. Syst. 13, (2001) 556–562.
- [27] M.W. Berry, M. Browne, A.N. Langville, V.P. Pauca, R.J. Plemmons, Algorithms and Applications for Approximate Nonnegative Matrix Factorization, Comput. Statist. Data Anal., 52 (2007) 155–173.
- [28] H. Ikeno, F.M.F. de Groot, E. Stavitski and I. Tanaka, J. Phys.: Condens. Matter 21, 104208 (2009).
- [29] H. Ikeno, T. Mizoguchi, Y. Koyama, Y. Kumagai and I. Tanaka, Ultramicroscopy **106**, 970 (2006).
- [30] H. Ikeno, I. Tanka, Y. Koyama, T. Mizoguchi and K. Ogasawara, Phys. Rev. B **72**, 075123 (2005)
- [31] N. R. Lugg, G. Kothleitner, N. Shibata and Y. Ikuhara, On the quantitiveness of EDX STEM, Ultramicroscopy, 151 (2015) 150–159.

Figure 1: Pseudo-code of our NMF with the orthogonal constraint (SO-NMF)

	<p>Input: Data matrix $\mathbf{X}$, the weight of orthogonal constraint w, the number of components K, the maximum number of iterations $T_{\max}$, the number of initializations $R_{\max}$</p> <p>Output: The cardinalities $\mathbf{C}$, spectra $\mathbf{S}$</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. (18)
7:	Normalize $\mathbf{C}_{\cdot k}$ by $\mathbf{C}_{\cdot k} \leftarrow \mathbf{C}_{\cdot k} / \ \mathbf{C}_{\cdot k}\ _2$
8:	end
9:	For k from 1 to K :
10:	Update $\mathbf{S}_{\cdot k}$ by Eq. (19)
11:	end
12:	Compute $L^{(t,r)} = D_{EU}(\mathbf{X} \mathbf{C}\mathbf{S}^T)$ by Eq. (4)
13:	end
14:	end
15:	Choose the best optimization results by $r_{best} = \operatorname{argmin}_r L^{(T_{\max},r)}$

Figure 2: Pseudo-code of our NMF with ARD and the soft orthogonal constraint (ARD-SO-NMF)

	<p>Input: Data matrix $\mathbf{X}$, the weight of orthogonal constraint w, the maximum number of components K, the maximum number of iterations $T_{\max}$, the number of initializations $R_{\max}$</p> <p>Output: The cardinalities $\mathbf{C}$, spectra $\mathbf{S}$ and relativeness $\boldsymbol{\lambda}$</p>
1:	Set hyper-parameter $a = 1 + \delta$ and compute b by Eq. (14)
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. (20)
8:	end
9:	For k from 1 to K :
10:	Update $\mathbf{S}_{\cdot k}$ by Eq. (10)
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 $\boldsymbol{\lambda}$ by Eq. (12)
22:	Update σ^2 by Eq. (13)
23:	Compute $L^{(t,r)} = L(\mathbf{C}, \mathbf{S}, \boldsymbol{\lambda}, \sigma^2)$ by Eq. (7)
24:	end
25:	end
26:	Choose the best optimization results by $r_{best} = \operatorname{argmin}_r L^{(T_{\max}, r)}$

Figure 3: Component distributions and spectra of the phantom Mn- $L_{2,3}$ EELS-SI data

Figure 4: Atom-site position and spectra for the theoretical Mn- $L_{2,3}$ EELS-SI data

Figure 5: Validated components of the Si- $L_{2,3}$. (a) Distribution of three components, and (b) their spectra.

Figure 6: ADF image of the LaSrMnO₃-Y doped ZrO₂ ceramic composite sample (a) and typical EDX counts per pixel (b).

a) Noise level $r_{\text{noise}}=0.01$

b) Noise level $r_{\text{noise}}=0.1$

c) Noise level $r_{\text{noise}}=0.2$

Figure 7: Result for Mn- $L_{2,3}$ phantom data by NMF-ARD ($w=0$)

a) Noise level $r_{\text{noise}}=0.01$

b) Noise level $r_{\text{noise}}=0.1$

c) Noise level $r_{\text{noise}}=0.2$

Figure 8: Result for Mn- $L_{2,3}$ phantom data by NMF-ARD-SO ($w=0.1$)

A) ARD-NMF ($w = 0$)

1) COMPONENT MAPS

2) AMPLITUDE OF λ

3) SPECTRA

4) LIKELIHOOD (COST FUNCTION)

B) ARD-SO-NMF ($w = 0.01$)

1) COMPONENT MAPS

2) AMPLITUDE OF λ

3) SPECTRA

4) LIKELIHOOD (COST FUNCTION)

Figure 9: Result for the real Mn- $L_{2,3}$ EELS-SI data

a) Noise level $r_{\text{noise}} = 0.01$

b) Noise level $r_{\text{noise}} = 0.1$

c) Noise level $r_{\text{noise}} = 0.2$

Figure 10: Result for Si diode phantom data by NMF-ARD ($w=0$)

a) Noise level $r_{\text{noise}} = 0.01$

b) Noise level $r_{\text{noise}} = 0.1$

c) Noise level $r_{\text{noise}} = 0.2$

Figure 11: Result for Si diode phantom data by NMF-ARD-SO ($w=0.1$)

a) ARD- NMF ($w = 0$)

1) COMPONENT MAPS

4) LIKELIHOOD (COST FUNCTION)

2) AMPLITUDE OF λ

3) SPECTRA

b) ARD-SO-NMF ($w = 0.1$)

1) COMPONENT MAPS

4) LIKELIHOOD (COST FUNCTION)

2) AMPLITUDE OF λ

3) SPECTRA

Figure 12: Result for the real Si- $L_{2,3}$ EELS-SI data

a) RESULT BY ARD-NMF

b) ELEMENTAL MAPS

Figure 13: Result of ARD-NMF for the real EDX-SI data

Figure 14: Results of SO-NMF with various weights of spatial orthogonality constraint for $\text{Si-L}_{2,3}$ EELS-SI data. (a) Reference component map, (b) reference spectra and (c) decomposition results.

Figure 15: Results of SO-NMF with various weights of spatial orthogonality constraint for Mn-L_{2,3} EELS-SI data. (a) Reference component map, (b) reference spectra and (c) decomposition results.