Non-negative Matrix Factorization and Its Extensions for Spectral Image Data Analysis

Motoki Shiga,a,b,c, † Shunsuke Muto d

a Faculty of Engineering, Gifu University, 1-1 Yanagido, Gifu, Gifu 501-1193, Japan
b PRESTO, Japan Science and Technology Agency, 4-1-8, Honcho, Kawaguchi-shi, Saitama 332-0012, Japan
c Center for Advanced Intelligence Project, RIKEN, Nihombashi 1-chome Mitsui Building, 15th floor, 1-4-1 Nihombashi, Chuo-ku, Tokyo 103-0027, Japan
d Electron Nanoscopy Division, Advanced Measurement Technology Center, Institute of Materials and Systems for Sustainability, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

† Corresponding author: shiga_m@gifu-u.ac.jp

Received: 30 July, 2019; Accepted: 5 September, 2019; Published: 21 September, 2019

Scanning transmission electron microscopy combined with electron energy-loss spectroscopy and energy-dispersive X-ray spectroscopy is useful for analyzing chemical states and elemental components in a new material. Using these instruments, the spectra over the spatial grid points in a region of interest can be observed. This measurement technique is called spectral imaging (SI). Because of the large size of SI data, the analysis cost is a bottleneck in the evaluation process of the material. To reduce the analysis cost, machine learning techniques can be applied, which can automatically extract essential information from the data. This paper reviews our developed machine learning method, which is based on non-negative matrix factorization and its extensions. A spatial orthogonality constraint and a generalized noise model, which includes Gaussian and Poisson noise models, are introduced. Numerical experiments demonstrate the effectiveness and characteristics of our developed methods.

Keywords Scanning transmission electron microscopy; Electron energy-loss spectroscopy; Spectral image analysis; Non-negative matrix factorization; Machine learning

I. INTRODUCTION

Scanning transmission electron microscopy (STEM), combined with electron energy-loss spectroscopy (EELS), and energy-dispersive X-ray spectroscopy are useful for analyzing the chemical states and elemental components of a material. Using these instruments, the spectra over the spatial grid points in a region of interest (ROI) can be observed. This measurement technique is called spectral imaging (SI). Because of the large size of SI data, the analysis cost is a bottleneck of this evaluation process.

Machine learning and data mining techniques are useful for analyzing massive datasets. In materials science, these techniques have been used to automatically identify novel rules, such as groups of examples or relationships between a set of descriptors and a target variable, for real measurement datasets as well as for theoretically computed datasets [1–5]. The advantages of these techniques are as follows: i) reduction of analysis cost; and ii) utilization of automatic computational procedures to avoid the subjectivity of experts (which might be an obstacle to novel findings).

This paper reviews our machine learning method developed to automatically identify chemical components and their spectra from SI datasets [4]. A goal of SI analysis is to identify chemical compositions (or chemical states) and their quantities at each measured spatial point solely from observed SI data. This is an unsupervised learning problem because supervised labels such as chemical components are unavailable. One useful approach to solve this problem is matrix factorization, such as singular value decomposition...
(SVD) or principal component analysis (PCA), which assumes that an observed spectrum at each spatial observed point is generated by linear mixing of the spectra of chemical components. Non-negative matrix factorization (NMF)—the main topic of this review paper—is an extension of matrix factorization in which matrix elements are constrained to non-negative values. This assumption is reasonable in many datasets, such as natural images [6], because such data consist of non-subtractive signals of several components. The seminal NMF study involved decomposition of face images into face parts such as eyes, nose, and mouth [6]. NMF can naturally decompose a face image into face parts correctly although the PCA of the subtractive approach, in which both non-negative and negative values are assumed, cannot identify face parts. Following the seminal study [6], NMF has been widely applied to other data such as audio signals [7], remotely sensed hyperspectral data [8], and microscopic SI data [3–5]. The non-negative assumption is also natural in SI data because the intensities of spectral signals and component densities are non-negative. Thus, NMF is expected to identify physically natural chemical components from observed SI data. Our developed NMF is an extension realized by introducing a spatial orthogonality constraint and a sparsity constraint to obtain physically meaningful results. This paper also introduces a generalization of a noise model, including a Gaussian noise model and a Poisson noise model. Numerical experiments are used to demonstrate the effectiveness of our developed methods.

The remainder of this paper is organized as follows: Section II explains the spectrum imaging dataset and the problem setting to automatically identify the spatial intensity distributions and spectra of the chemical components. Additionally, the basic machine learning methods and our proposed approach—which is based on non-negative matrix factorization—are also described. Section III presents the empirical demonstration of our developed methods using synthetically generated SI datasets. Finally, Section IV summarizes our developed methods and outlines possible future work.

II. NON-NEGATIVE MATRIX FACTORIZATION

A. Spectral image data and matrix factorization model

A spectral image dataset consists of a data cube (or a three-dimensional array) having two-dimensional spatial axes and a spectral channel axis, such as electron energy loss in an EELS spectrum. By stacking each observed spectrum along a row in a matrix, the data cube can be converted into a two-dimensional array called a data matrix.

When the essential chemical components in the ROI are limited to a small number, the data matrix can be described by a mixture of spectra and spatial intensity distributions of these chemical components. In particular, under a linear mixture assumption, the i-th spatial position and the j-th spectral channel in the data matrix \( X_{ij} \) is modeled by a linear combination

\[
X_{ij} = \sum_{k=1}^{K} C_{ik} S_{jk}, \quad \ell = 1, \ldots, N_{xy}, j = 1, \ldots, N_{ch},
\]

where \( K \) is the number of chemical components; \( N_{xy} \) is the number of spatial measured points; \( N_{ch} \) is the number of spectrum channels; \( C_{ik} \) is the intensity of the i-th spatial position of the k-th chemical component; and \( S_{jk} \) is the intensity of the j-th spectrum channel of the k-th component. The model can be expressed as the matrix representation

\[
\bar{X} = CS^T,
\]

where \( C \in \mathbb{R}^{N_{xy} \times K} \) is a factorized matrix of chemical component intensity distributions and \( S \in \mathbb{R}^{N_{ch} \times K} \) is that of chemical component spectra. The approximation model of Eq. (2) is called matrix factorization. When \( K \) is much smaller than both \( N_{xy} \) and \( N_{ch} \), the total number of elements in \( C \) and \( S \) is much lower than that in the original data matrix \( X \). This model is equivalent to a low rank approximation in linear algebra. In this problem setting, \( X \) is known, and \( C \) and \( S \) are unknown. \( C \) and \( S \) can be estimated by minimizing the reconstruction error between \( X \) and \( \bar{X} \). This error is usually measured by the mean squared error under an additive Gaussian noise model. For another noise model, another error measure could be used, which is discussed in a later section.

Several matrix factorization methods exist with various constraints of matrices \( C \) and \( S \) [1, 5]. A basic method is SVD, which decomposes a data matrix into two orthonormal matrices, \( U \in \mathbb{R}^{N_{xy} \times K} \) and \( V \in \mathbb{R}^{N_{ch} \times K} \), and a diagonal matrix of singular values, \( D \in \mathbb{R}^{K \times K} \). These matrices consist of the basis vectors corresponding to the \( K \) largest singular values, which is given by

\[
\hat{X} = UDV^T.
\]

After the decomposition, \( U \) is considered as a spatial intensity matrix (or a redundant matrix); \( C \) and \( VD \) are considered as spectrum matrices (or basis matrices); \( S \). Matrices \( C \) and \( S \) are orthogonal even when the true component spectral matrix and the true intensity distribution matrix are not orthogonal. This can cause an unnatural decomposition when the assumption is violated in the true chemical components.

Another well-known method is PCA. The main difference between SVD and PCA is that SVD finds basis vectors (basis spectra) from the origin, while PCA finds basis vectors from the mean vector (i.e., the averaged spectrum). Thus, the procedure in PCA is to first compute the averaged spectra and then subtract the spectra from each observed spectrum. The obtained matrix is used in the SVD procedure and then factorized matrices \( C \) and \( S \) are obtained. Similar to SVD, \( C \) and \( S \) by PCA are also orthogonal. This can also cause an unnatural decomposition when the assumption is violated.

For SI analysis, there is another useful method called ver-
text component analysis (VCA) [9]. In VCA, it is assumed that data is distributed on a convex cone whose vertexes correspond to spectra of pure chemical components that include only a chemical state. The VCA algorithm first projects the original data into a low-dimensional space by SVD or PCA and then finds vertexes corresponding to pure spectra. Next, VCA estimates the spatial intensity of each component included in an observed spectrum. VCA is powerful when the observed SI has pure spectra of all components to be extracted. However, it may drastically degrade when an observed SI does not include pure spectra. To mitigate this problem, signal subspace sampling (SSS), in which pure spectra are virtually generated and estimated by sampling procedures, was proposed [10]. However, the problem is not fully solved when the observed spectra are heavily overlapped in a SI dataset. This is a limitation of inductive approaches in statistical inference and machine learning, rather than in the SSS procedure. The solution requires an additional assumption (or an effective constraint) in the estimation model.

B. Non-negative matrix factorization

A natural assumption in SI data is non-negativity, for both the spectrum and the spatial intensity of the chemical component at an observed spatial position. NMF is a matrix factorization model with a non-negativity assumption for the data matrix \( X \in \mathbb{R}^{N_{xy} \times N_{ch}} \) and factorized matrices \( C \in \mathbb{R}^{N_{xy} \times K} \) and \( S \in \mathbb{R}^{N_{ch} \times K} \). The optimization problem is formulated as

\[
\min_{C \geq 0, S \geq 0} J(C, S),
\]

where \( J(C, S) \) is the reconstruction error, such as a mean squared error:

\[
J(C, S) = \| X - CS^T \|_F^2 = \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - (CS^T)_{ij})^2.
\]

The additional assumption is simple, but powerful for obtaining a physically meaningful decomposition. However, the assumption makes the optimization much more difficult. The optimization of both \( C \) and \( S \) is a non-convex optimization problem, and the global optimum by each optimization procedure with an initialization cannot be guaranteed. Thus, NMF should be implemented numerous times from different initial matrices \( C \) and \( S \), and the best result minimizing the reconstruction error is chosen.

To avoid incorrect local minima, in which the reconstruction error is much larger than the global minimum, and to converge to an optimum quickly, a number of optimization algorithms have been proposed. To the best of our knowledge, in all these algorithms, the optimizing variables in \((C, S)\) are not updated simultaneously, but a subset of the variables are updated while the others are fixed. One of the seminal optimization algorithms is the matrix multiplication (MM) algorithm [11]. It can be implemented by only matrix products, resulting in a small computation cost for an update. However, the convergence speed is not fast when the optimal matrices include a lot of zeros. Another efficient algorithm is the alternative least squares (ALS) algorithm [12]. This algorithm updates a matrix with the other matrix fixed by solving a linear equation to minimize the error without the non-negative constraint. The non-negative constraint is enforced after solving the equation. Empirically the ALS algorithm converges faster than the MM algorithm. However, it requires a high computation cost for an update to solve the linear equation. An extension of the ALS algorithm is the hierarchical ALS (HALS) algorithm [13], in which a column vector of \( C \) or \( S \), rather than a matrix, is updated. The optimal vector for the update is analytically solved without solving a linear equation, resulting in a low computation cost. Empirically, the HALS algorithm converges to a local optimum faster than the ALS algorithm.

C. Soft spatial orthogonality constraint

The abovementioned optimization algorithms do not impose any additional constraint for factorized matrices except for the non-negativity. Although novel optimization techniques can improve the convergence speed, they still suffer from the problem of incorrect local optima, which causes physically unnatural results. To mitigate this problem, an additional constraint in analyzing datasets is necessary. Our paper [4] proposed a constraint in which the spatial positions of different chemical components cannot be perfectly overlapped. When a pair of components are spatially non-overlapped and these intensities are non-negative, the inner product between these column vectors of \( C \) is zero, that is, \( C_i^T C_j = 0, i \neq j \). Thus, the spatial overlap can be measured by an inner product of different column vectors in matrix \( C \). The total overlap between all pairs of different components can be measured by the sum of the off-diagonal elements of \( C^T C \) [14].

This condition of perfectly orthogonality is too strict for real observation situations, where the components may be slightly overlapped. To relax the strict orthogonality constraint, our paper [4] introduced a weight parameter for the orthogonality penalty, called soft orthogonality. Using the weighted penalty, we proposed a new NMF called soft orthogonalized NMF (NMF-OSO). NMF-OSO can be used to resolve unnatural spatial overlaps by adjusting the orthogonality penalty.

The NMF-OSO optimization algorithm was derived based on the HALS approach. To update the \( k \)-th column vector \( C_k \), the cost function with the orthogonality penalty can be formulated by

\[
J(C_k) = \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} (X_{ij} - \hat{X}_{ij}) + w \cdot \lambda \cdot C_k^T C_k, \tag{6}
\]

where \( w \) is a weight used to adjust the orthogonality penalty; \( \lambda \) is the Lagrange multiplier for the perfect orthogonality.
constraint; and $c^{(k)} = \sum_{j \neq k} c_{j}$. The second term is the orthogonality penalty term, which evaluates the overlap between the $k$-th column vector and all other columns; it is zero when the $k$-th component is not spatially overlapped with all the other components. The cost function is a quadratic function of $C_{k}$. The update rule for minimizing Eq. (6) can be derived using the normal equation, in which the gradient of $C_{k}$ is zero, as follows:

$$C_{k} = [X^{(k)}S_{k} - w \cdot \frac{c^{(k)}T X^{(k)} S_{k} c^{(k)}}{c^{(k)}T c^{(k)}}, k = 1, \ldots, K], (7)$$

where the function $[\cdot]_{+}$ replaces negative values with zeros. This update rule is iteratively implemented for each column $k = 1, 2, \ldots, K$. The update rule of $S_{k}$ is the same as in the original HALS algorithm because any additional constraint is not imposed on matrix $S$.

### D. Generalized noise model

The reconstruction error measured by Eq. (5) assumes additive Gaussian noise. When the countable number (or the S/N ratio) of signals is limited, one should assume a different noise model, such as a Poisson noise model. A generalized noise model is a Tweedie distribution, which is a family of probability distributions, including a Gaussian (or normal) distribution, a gamma distribution, and a Poisson distribution. When the countable number (or the S/N ratio) of signals is limited, one should assume a different noise model, such as a Poisson noise model. A generalization of the Tweedie distribution is the normal equation, in which the gradient of $C_{k}$ can be derived using the normal equation, in which the gradient of $C_{k}$

$$d_{\beta=a}(x||y) = x \log \frac{x}{y} - x + y, \quad (11)$$

which is derived under the noise assumption of a Poisson distribution. When $\beta = -1$, the divergence is the Itakura–Saito divergence:

$$d_{\beta=-1}(x||y) = \frac{x}{y} - \log \frac{x}{y} - 1, \quad (12)$$

which is equivalent to assuming the noise model of a gamma distribution.

**Figure 1** shows the $\beta$-divergence with $\beta = -1, 0, 1$. It indicates the left-right asymmetrical shape at $x = y$ when $\beta < 1$. The divergence grows more rapidly for $x > y$ than $x < y$ when $\beta < 1$. Thus, an approximation at a small observed value of $x$ should be more precise than that at a large value. This characteristic agrees with a Poisson noise assumption, whose variance is larger with a large mean than with a small mean.

Using the $\beta$-divergence, the reconstruction error for NMF is given by

$$f(C, S; \beta) = \sum_{i=1}^{N_{xy}} \sum_{j=1}^{N_{ch}} d_{\beta}(X_{ij}||Y_{ij}). \quad (13)$$

Although the formulation of $\beta$-divergence, which has nonlinear terms, is more complicated than the Euclidean distance, the HALS algorithm can be derived by the Taylor expansion of the divergence [16] as follows:

$$C_{k} \leftarrow \left[\underbrace{\frac{\nabla^{2} \phi_{\beta}(C^{T}S_{k})}{\nabla^{2} \phi_{\beta}(C^{T}S_{k})}}_{\Theta}\right]^{+}, \quad (14)$$

where the $\Theta$ operator is the element-wise product between two matrices, known as the Hadamard product. Introducing the orthogonality constraint, the update rule based on HALS [14] is given by

![Figure 1: $\beta$-divergence $d_{\beta}(x||y)$ when (a) $x = 0.8$, and (b) $x = 2.5$. The red dotted line indicates the value of $x$. The blue, orange, and green lines are equivalent to Euclidian distance, KL divergence and Itakura–Saito divergence, respectively.](image-url)
\[ \mathbf{C}_k' \leftarrow \mathbf{A} \mathbf{S}_k' - \mathbf{w} \mathbf{c}^{(k)} \mathbf{c}^{(k)} + \nabla^2 \phi \mathbf{c}^{(k)} \mathbf{S}_k' \mathbf{S}_k' + \mathbf{S}_k' \mathbf{S}_k', \]  \tag{15}

where

\[ \mathbf{A} = \{ \nabla^2 \phi \mathbf{c}^{(k)} \} \otimes \mathbf{X}^{(k)}. \]  \tag{16}

For the soft orthogonality penalty, the update rule can be derived as

\[ \mathbf{C}_k' \leftarrow \mathbf{A} \mathbf{S}_k' - \mathbf{w} \mathbf{c}^{(k)} \mathbf{c}^{(k)} + \nabla^2 \phi \mathbf{c}^{(k)} \mathbf{S}_k' \mathbf{S}_k' + \mathbf{S}_k' \mathbf{S}_k', \]  \tag{17}

An experimental demonstration is provided in Section III.

E. Automatic relevance determination

There is another extension of NMF where the number of components is chosen from only observed data, which introduces sparse penalty terms in the cost function of NMF [4, 17]. This approach is important when the number is unknown and one needs to evaluate the statistically significant number of components from observed data without expert knowledge. This approach is called automatic relevance determination (ARD). The ARD term is designed based on a Bayesian approach, and the optimization of the factorized matrices \( \mathbf{C} \) and \( \mathbf{S} \) is performed based on a maximum a posteriori probability (MAP) estimation. The ARD approach can be combined with our soft spatial orthogonality penalty. Furthermore, the HALS-based algorithm with \( \beta \)-divergence can be derived by replacing the Euclidean distance term in...
with its Taylor expansion in [14, 16]. Because of space limitations, the details are omitted.

III. NUMERICAL EXPERIMENTS
A. Synthetic data

This numerical experiment assumes a limited number of signal counts, which is equivalent to a Poisson noise model. The probability of \( n \) event observations in an interval is given by

\[
P_{\text{poisson}}(n; \lambda) = \frac{\lambda^n}{n!} \exp(-\lambda),
\]

where \( \lambda \) is the average number of events per interval. The expectation and variance under this distribution are \( \lambda \). For a sufficiently large value of \( \lambda \), the distribution can be approximated by a Gaussian distribution. Thus, this distribution is assumed only for the case where the number of counts is limited.

Synthetic datasets were generated using i) a theoretical computation of an atomic resolution STEM-EELS SI of Mn-L\(_{2,3}\) of Mn\(_3\)O\(_4\), which includes two chemical components of trivalent manganese (Mn\(^{3+}\)) and divalent manganese (Mn\(^{2+}\)), i.e., \( K = 2 \); and ii) an ideal decomposition result by NMF with \( K = 3 \) for real STEM-EELS SI data measured in a Si diode. The details of the true chemical components are described in Ref. 18. These true chemical component maps and spectra are shown in Figures 2 and 3. Using the distribution matrix \( C \) and spectrum matrix \( S \), an element of an SI data matrix \( X \) was generated by

\[
X_{ij} \sim P_{\text{poisson}}(n; \lambda_{ij}), i = 1, \ldots, N_x, j = 1, \ldots, N_c,
\]

where \( \lambda_{ij} = \alpha \cdot [CS^T]_{ij} \) and \( \alpha = \mu / \max(j)[CS^T]_{ij} \); \( \mu \) is a parameter of the synthetic dataset generation to adjust the number of the averaged count in an SI dataset. Synthetic datasets were generated with \( \mu = 10, 100 \) for Mn\(_3\)O\(_4\) data and Si device data.

B. Experimental results

Using our developed Python package [19], \( \beta \)-NMF (\( \beta \)-NMF-SO with \( w = 0 \)) and \( \beta \)-NMF-SO with error parameters \( \beta = 1, 0 \) were implemented to analyze the STEM-EELS SI datasets. The soft orthogonality weight of \( \beta \)-NMF-SO was adjusted for each data, which was set to \( w = 0.1 \) for atomic resolution Mn\(_3\)O\(_4\) datasets and \( w = 0.2 \) for Si device datasets. For datasets with essentially overlapped components, the smaller the weight is, the better the decomposition result is.

The decomposition results of Mn\(_3\)O\(_4\) datasets are shown in Figures 4 and 5. Figure 4(a, b), which illustrates results for a low count case, shows that results by \( \beta \)-NMF-SO with

Figure 6: The decomposition results by \( \beta \)-NMF-SO with \( w = 0 \) for STEM-EELS-SI of a Si diode device. Top: decomposed spectra. Bottom: decomposed component maps (left: \( k = 1 \), middle: \( k = 2 \), right: \( k = 3 \)).

Figure 7: The decomposition results by \( \beta \)-NMF-SO with \( w = 0.2 \) for STEM-EELS-SI of a Si diode device. Top: decomposed spectra. Bottom: decomposed component maps (left: \( k = 1 \), middle: \( k = 2 \), right: \( k = 3 \)).
Signals could result in improved identification of the chemical component via machine learning.

Acknowledgments

Additional modeling of background disadvantages to our method caused by background signals. The orthogonality assumption could work by effectively removing the arbitrary space, which is equivalent to a set of physically unnatural decompositions, although the noise model could not remove the space.

Figures 6 and 7 show results for Si device datasets. For these datasets, a Poisson noise model does not improve decomposition performance, although the soft orthogonality penalty can drastically improve the performance. The optimum minimizing reconstruction error for this dataset may have a large arbitrary space due to non-sparse spectra. The orthogonality assumption could work by effectively removing the arbitrary space, which is equivalent to a set of physically unnatural decompositions, although the noise model could not remove the space.

IV. CONCLUSIONS

This paper reviewed automatic chemical component identification from spectral image data based on NMF. Extensions from the basic NMF consisting of i) a general noise model and ii) the soft spatial orthogonality penalty, were introduced and demonstrated using synthetic STEM-EELS-SI datasets. NMF with KL divergence, which is equivalent to a Poisson noise model, performed better than a Gaussian noise model when the number of observed signal counts was limited. However, the improvement was minor for these datasets. The main reason for this is that the S/N ratio is still incorrect even when a suitable noise model is used. On the other hand, another assumption of the spatial orthogonality constraint drastically improved the decomposition performance, by effectively restricting the solution space minimizing the reconstruction error of NMF. This analysis gives us the important message that the information quantity hidden in observed data (or the S/N ratio) does not increase even by using a machine learning method, but the method can efficiently extract all the information with a suitable assumption for the data.

A future study will address the treatment of background signals. Our previous study [4] summarized that there are disadvantages to our method caused by background signals. Thus, these signals need to be eliminated before implementation of our method. Additional modeling of background signals could result in improved identification of the chemical component via machine learning.

Acknowledgments

This work was supported in part by Grants-in-Aid for Scientific Research on Innovative Areas “Nano Informatics” (Grant Nos. 25106004, 26106510, and 16H00736). “Interface Ionics” (Grant No. 19H05815), KIBAN-KENKYU A (Grant No. 26240906) and B (Grant No. 16H02866) from the Japan Society for the Promotion of Science and by Precursory Research for Embryonic Science and Technology (Grant No. JPMJPR16N6) from the Japan Science and Technology Agency.

References

[1] H. F. Grahn and P. Geladi (Eds.), Techniques and Applications of Hyperspectral Image Analysis (John Wiley & Sons, 2007).
[2] K. Kiyohara, T. Miyata, K. Tsuda, and T. Mizoguchi, Sci. Rep. 8, 13548 (2018).
[3] M. Shiga and S. Muto, in: Nanoinformatics, edited by I. Tanaka (Springer, Singapore, 2018) pp. 179–203.
[4] M. Shiga, K. Tatsumi, S. Muto, K. Tsuda, Y. Yamamoto, T. Mori, and T. Tanji, Ultramicroscopy 170, 43 (2016).
[5] M. Shiga, S. Muto, K. Tatsumi, and K. Tsuda, Trans. Mater. Res. Soc. Jpn. 41, 333 (2016).
[6] D. D. Lee and H. S. Seung, Nature 401, 788 (1999).
[7] C. Févotte, N. Bertin, and J.-L. Durrieu, Neural Comput. 21, 793 (2009).
[8] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, IEEE Trans. Geosci. Remote Sens. 54, 6076 (2016).
[9] J. M. P. Nascimento and J. M. B. Dias, IEEE Trans. Geosci. Remote Sens. 43, 898 (2005).
[10] J. Spiegelberg, S. Muto, M. Ohtsuka, K. Peckmans, and J. Rusz, Ultramicroscopy 182, 205 (2017).
[11] D. D. Lee and H. S. Seung, in: Advances in Neural Information Processing Systems 13 (NIPS 2000) p. 556.
[12] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, Comput. Stat. Data Anal. 52, 155 (2007).
[13] A. Cichocki and A.-H. Phan, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. E92.A 708 (2009).
[14] K. Kimura, M. Kudo, and Y. Tanaka, Mach. Learn. 103, 285 (2016).
[15] A. Cichocki, R. Zdunek, A.-H. Phan, and S. Amari, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation (John Wiley & Sons, 2009) pp. 81–129.
[16] L. Li, G. Lebanon, and H. Park, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM Publications, New York, 2012) pp. 307–315.
[17] V. Y. F. Tan and C. Févotte, IEEE Trans. Pattern Anal. Mach. Intell. 35, 1592 (2013).
[18] K. Tatsumi, S. Muto, Y. Yamamoto, H. Ikeno, S. Yoshioka, and I. Tanaka, Ultramicroscopy 106, 1019 (2006).
[19] https://github.com/MotokiShiga/malspy

All articles published on e-J. Surf. Sci. Nanotechnol. are licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0). You are free to copy and redistribute articles in any medium or format and also free to remix, transform, and build upon articles for any purpose (including a commercial use) as long as you give appropriate credit to the original source and provide a link to the Creative Commons (CC) license. If you modify the material, you must indicate changes in a proper way.

Published by The Japan Society of Vacuum and Surface Science