AI Analysis Basic Course Second Installment: Application of Basis Profile Decomposition to the Powder X-ray Diffraction Method

Takumi Ohta*

Abstract

A method for estimating profiles of individual components (basis profile decomposition) from powder X-ray diffraction profiles of mixtures is described. In addition, a method for quantitative analysis combining the estimated profiles with the DD method™ is introduced. We show that it is possible to accurately estimate profiles of individual components by selecting the method according to the measured profiles or sample characteristics.

1. Introduction

Powder X-ray diffractometry is a nondestructive method for analyzing samples. In many cases, a sample is a mixture containing many components. For qualitative analysis, a peak search is performed on the measurement profile, and a list of peak positions is prepared. Then, the phase of each component is identified by matching the peak positions in a database with the list. For quantitative analysis, the Rietveld analysis using crystal structure information in a database and the DD method[(1), (2)] using known profile shapes can be performed. However, these analyses require crystal structure information or decomposition of patterns into individual components. Therefore, new analysis methods have been desired for quantitative analysis of compounds for which crystal structure information is not available or whose patterns cannot be decomposed into individual components by existing methods.

In order to extract characteristic profiles from multicomponent profiles, analysis methods using matrix factorization have been studied[(3), (4)]. In matrix factorization, a matrix (data matrix) X that arranges intensity information of multiple measured profiles is analyzed. If the number of measured profiles is N and the number of measurement points (for example, angles) is M, each row of the matrix X with N rows and M columns stores the intensity of the measured profile. Using matrix factorization, the data matrix can be decomposed into matrix W (weight matrix) with N rows and R columns and matrix B (basis matrix) with R rows and M columns. The matrix components are as follows.

$$X(n,m) = \sum_{r=1}^{R} W(n,r)B(r,m)$$
$$n = 1,2,...,N$$
$$m = 1,2,...,M \tag{1}$$

Each row of matrix B is considered to contain a characteristic profile. This is called the basis profile. The corresponding column of matrix W is considered to contain the weight of the basis profile included in the measured profile. A schematic diagram is shown in Fig. 1.

Here, the profile of the mixture is approximately the sum of profiles of pure substances weighted by the weight fractions. That is, the intensity $X(n,m)$ at the m-th measurement point of the n-th sample can be expressed using the intensity $B_{\text{pure}}(r,m)$ and the weight $W_{\text{pure}}(n,r)$ of the profile of r-th substance as

$$X(n,m) = \sum_{r=1}^{R} W_{\text{pure}}(n,r)B_{\text{pure}}(r,m) \tag{2}$$

where R is the number of pure substances contained in the sample. By comparing Equations (1) and (2), we can see that the profile of the pure substance and its weight fraction contained in the profile of the mixture can be estimated using matrix factorization. However, to interpret the solution of matrix factorization in this way, the characteristic profile must be regarded as the profile of a pure substance. To improve the interpretability and decomposition accuracy of the matrix factorization solution, a calculation method for matrix factorization must be devised.

In this paper, we introduce matrix factorization methods for powder X-ray diffraction profiles according to the characteristics of the measured profiles or samples. These functions are implemented

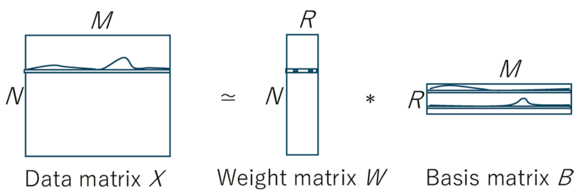

Fig. 1. Matrix factorization. The data matrix is represented by the product of the weight matrix and the basis matrix.

* X-ray Research Laboratory, Rigaku Corporation.

in SmartLab Studio II[5] and are called the basis profile decomposition.

2. Quantitative Analysis Using Basis Profile Decomposition

We introduce quantitative analysis using profiles obtained by the basis profile decomposition. Specific methods of basis profile decomposition are described in section3 and later. When a profile obtained by the basis profile decomposition can be regarded as a profile of a pure substance, quantitative analysis by the DD method can be performed using the profile and the chemical formula of the substance. If the chemical formula is unknown, an appropriate chemical formula can be used to examine the change of relative quantitative values.

The characteristic of this quantitative analysis is that it is not necessary to prepare a pure substance or a database. For example, when measuring a reaction system, it is difficult to extract a single intermediate product. If there is no database, phase identification and the Rietveld analysis cannot be performed. In such cases, the basis profile decomposition can be used to extract characteristic profiles contained in multiple measurement profiles. By using these profiles in the DD method, quantitative analysis becomes possible.

3. Examples of Application

In this section, the methods implemented in SmartLab Studio II are described. In addition, the results of applying each method to powder X-ray diffraction profiles and the results of quantitative analysis by the DD method are shown.

3.1. Alternating least squares method

First, basis profile decomposition by the alternating least squares method is described. In this method, the matrices W and B in Equation (1) are iteratively updated by the least squares method[3]. In the update, a non-negative constraint is added to the matrices. This constraint takes into account the fact that X-ray diffraction profiles have non-negative intensities and their weights are non-negative. This constraint is applied to all the following methods. On the other hand, this method does not include *a priori* information about the sample. Since the calculation time of this method is short, it is suitable for initial analysis to understand the characteristics of the measured data.

The method is applied to an indomethacin sample that

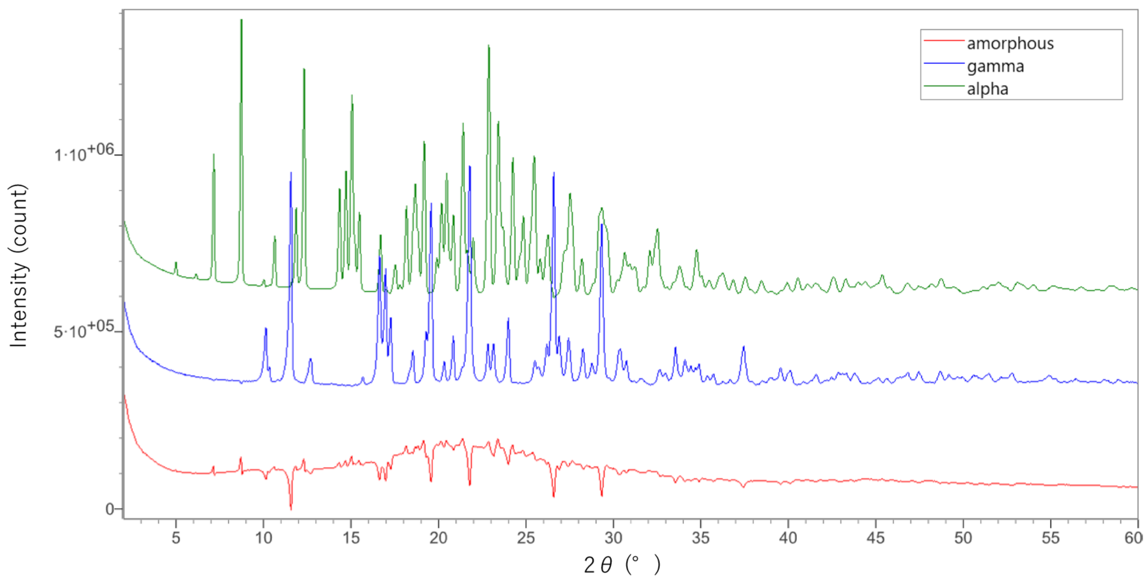

Fig. 2. Basis profiles obtained by the alternating least squares method.

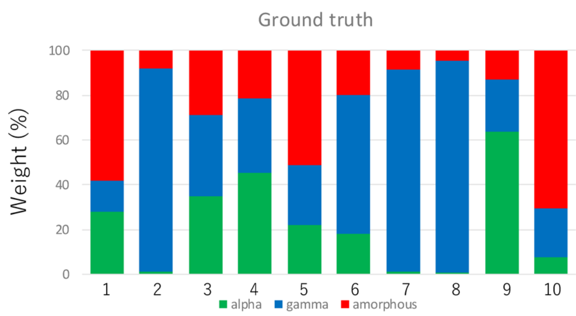

Fig. 3. Weight fractions of alpha, gamma, and amorphous indomethacin in 10 samples.

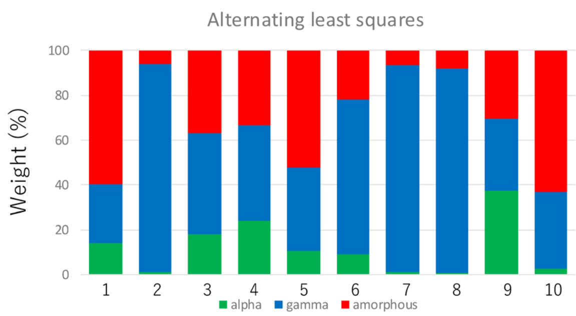

Fig. 4. Quantitative values by DD method using basis profiles obtained by the alternating least squares method.

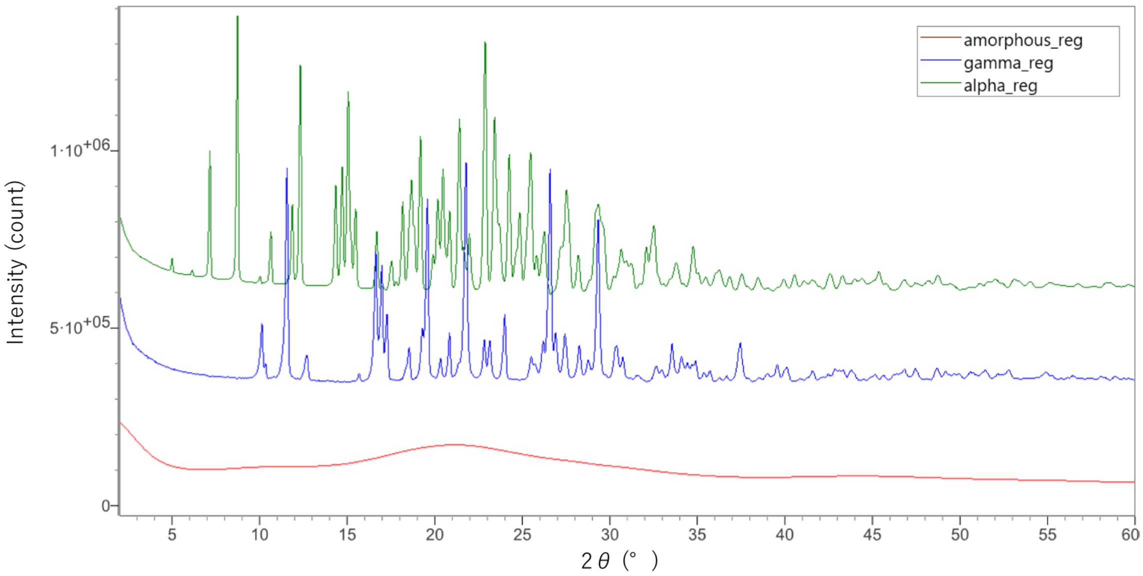

Fig. 5. Basis profile obtained by the amorphous regularization method.

contains three types of indomethacin: alpha, gamma, and amorphous. The 101 profiles in total were prepared by varying the weight fraction of each type. The three basis profiles obtained by the method are shown in Fig. 2. By visually confirming the characteristic shape of each basis profile, the name of the compound was applied to the label of the basis profile (the same applies below). The intensity of the amorphous basis profile decreases at the location where another basis profile has a peak. In addition, peaks that do not exist in the profiles of pure substances sometimes appear in the basis profiles.

The weight fraction is calculated by using the obtained basis profile for the known profile in the DD method. In this study, 10 indomethacin mixtures were prepared. The weight fractions of pure substances in each mixture are shown in Fig. 3. By regarding the weight fraction as the true value, we validate below the quantitative analysis by the basis profile decomposition and the DD method. Figure 4 shows the results of the DD method. The root mean squared error (RMSE) between the quantitative value by the DD method and the true value is 10.3%. Thus, quantitative analysis using the basis profile decomposition and the DD method can calculate the weight fraction. Although there is an average quantitative error of 10%, this method is useful because it does not require the preparation of known profiles.

3.2. Amorphous regularization method

X-ray diffraction profiles of amorphous materials have broad peaks. When the alternating least squares method is applied to measured profiles containing amorphous materials as shown in Section 3.1, the shape of the basis profile corresponding to the amorphous materials is not smooth (Fig. 2). It is necessary to devise a method to obtain a profile that is closer to the true amorphous profile.

The amorphous regularization method assumes that the sample contains one amorphous material. When updating the matrices, the basis profile corresponding to the amorphous material is searched. If the amorphous profile has a broad peak, the spatial variation of the profile is considered to be small. Then, we introduce a quantitative index called the relative total variation (RTV)

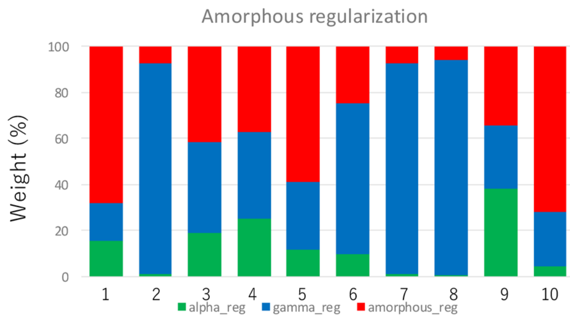

Fig. 6. Quantitative values obtained by the DD method using the basis profiles obtained by the amorphous regularization method.

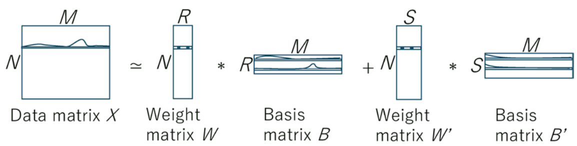

Fig. 7. Matrix factorization for the semi-supervised minimization method. The data matrix is represented by two weight matrices and two basis matrices.

$$\mathrm{RTV} = \frac{\|\mathrm{grad} f\|}{\|f\|} \qquad (3)$$

Here, f is the basis profile corresponding to a row of the basis matrix B. The RTV is calculated for all the basis profiles, and the basis profile corresponding to the smallest RTV is regarded as the amorphous one. By applying a strong smoothing process to this basis

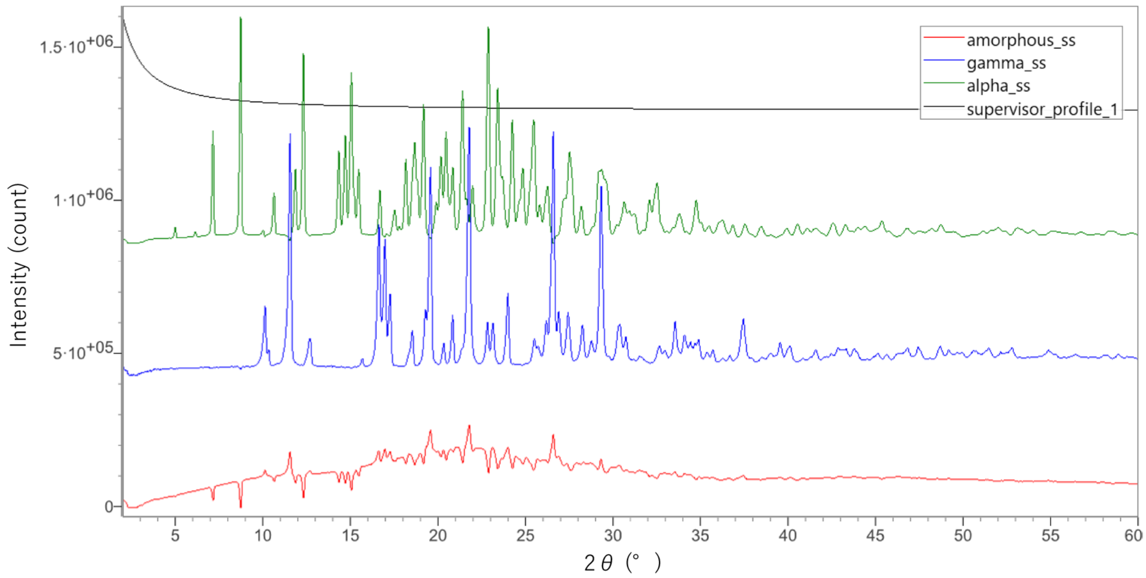

Fig. 8. Basis profiles obtained by the semi-supervised minimization method. The black profile is the weighted sum of all monomials used.

profile during the update of the matrices, a smooth profile is obtained. If the sample contains more than one amorphous substance, the basis profiles corresponding to that number of the smallest RTVs can be regarded as the amorphous ones.

This method is applied to the sample of indomethacin in Section 3.1. The obtained basis profiles are shown in Fig. 5. It can be seen that the basis profile (amorphous_reg) corresponding to the amorphous substance is smoother than that (amorphous) in the previous section.

The results of the DD method using the obtained basis profiles are shown in Fig. 6. The RMSE between the quantitative values and the true values by the DD method is 9.6%. In the sense of RMSE, the quantitative values are better than those by the alternating least squares method.

3.3. Semi-supervised minimization method

In the previous sections, the basis profiles were obtained by matrix factorizations without using known profiles. The semi-supervised minimization method incorporates known information of samples by using known profiles in matrix updates. For example, known profiles can be profiles of pure substances, backgrounds, and functional forms.

In the semi-supervised minimization method, the measured profile is represented by the sum of unknown and known profiles. The intensity $X(n,m)$ at the m-th measurement point of the n-th sample can be represented by the intensity $B(r,m)$ and the weight $W(n,r)$ of the r-th unknown profile and the intensity $B'(s,m)$ and the weight $W'(n,s)$ of the s-th known profile as

$$X(n,m) = \sum_{r=1}^{R} W(n,r)B(r,m) + \sum_{s=1}^{S} W'(n,s)B'(s,m) \quad (4)$$

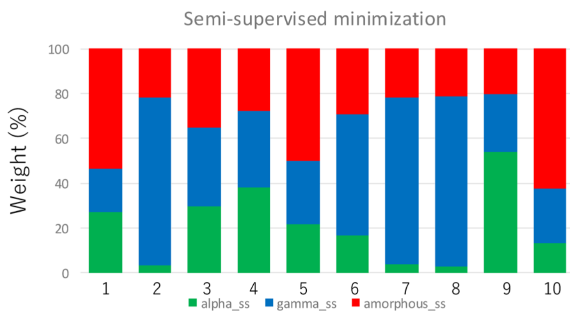

Fig. 9. Quantitative values obtained by the DD method using the basis profiles obtained by the semi-supervised minimization method.

where S is the number of known profiles. In other words, in the semi-supervised minimization method, the data matrix is decomposed into two types of weight matrices and basis matrices. Figure 7 shows a schematic diagram. When updating the matrices, the basis matrix B and the weight matrices W' and W are updated while the basis matrix B' is fixed. The optimization is based on the method of majorizer minimization[(6), (7)].

An analysis using an inverse power polynomial as a known profile is shown. The inverse power polynomial is expressed by the sum of the following monomials with x as the measurement point.

$$x^l \quad \left(l = -\frac{1}{2}, -1, -\frac{3}{2}, -2, \ldots \right) \quad (5)$$

By storing these profiles in each row of basis matrix B', the background shape can be represented by an inverse power polynomial formed by adding monomials with nonnegative weights. Powder X-ray diffraction profiles often show a background whose intensity increases toward lower angles. In such a case, it is effective to represent the background by an inverse power polynomial.

We apply this method to the sample of indomethacin in the previous section. The obtained basis profiles are shown in Fig. 8. Here, the profile (supervisor_profile_1) obtained by adding all the used monomials with weights is shown. In the following, this is called the estimated background.

The results of the DD method using the obtained basis profiles are shown in Fig. 9. In the DD method, the estimated background (supervisor_profile_l) was used as the blank data for background removal. The RMSE between the quantitative value obtained by the DD method and the true value is 8.4%. In the sense of RMSE, the quantitative value is improved from the result obtained by the alternating least squares method.

3.4. Constrained alternating least squares method

In powder X-ray diffraction, a profile common to all measured profiles may appear. For example, there are cases when an equal amount of reference substance is added to all samples, some substances do not change in the reaction system, and there is a common background from the equipment.

In the constrained alternating least squares method, the matrices are updated on the assumption that there is a profile common to all measured profiles. Specifically, such a profile is supposed to be stored in the *R*-th row of the basis matrix. When updating the matrices, the *R*-th column of the weight matrix is constrained to have the same value.

The method is applied to 10 samples obtained by summing the profiles of three pure substances (referred to as $A1_2O_3$, SiO_2 and ZnO, respectively). However, the contents of $A1_2O_3$ are assumed to be common to all samples, and the weights of the other substances are

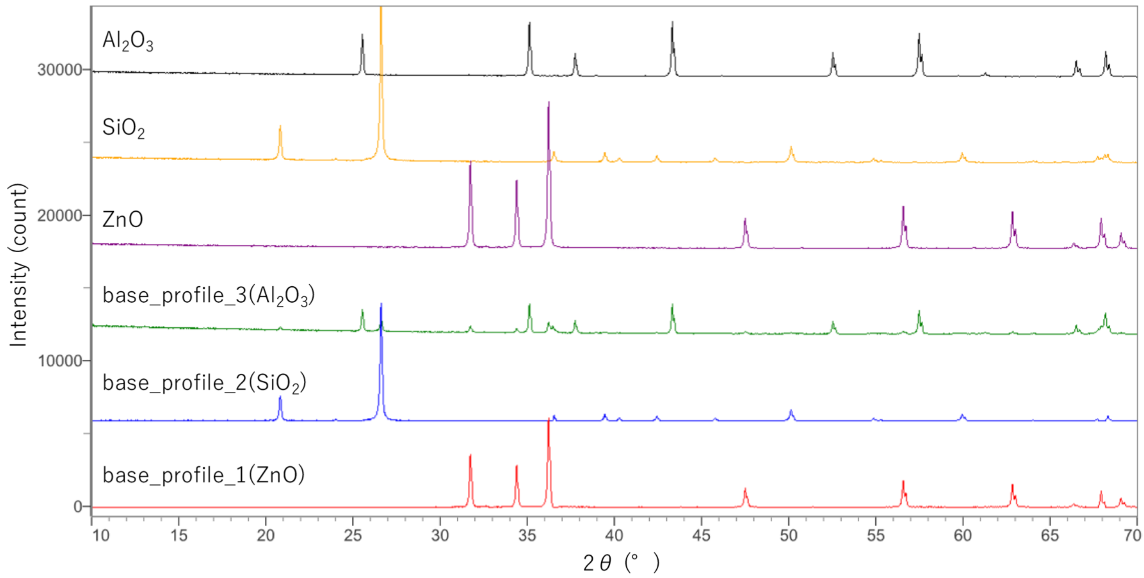

Fig. 10. Basis profiles obtained by the constrained alternating least squares method and profiles of pure substances.

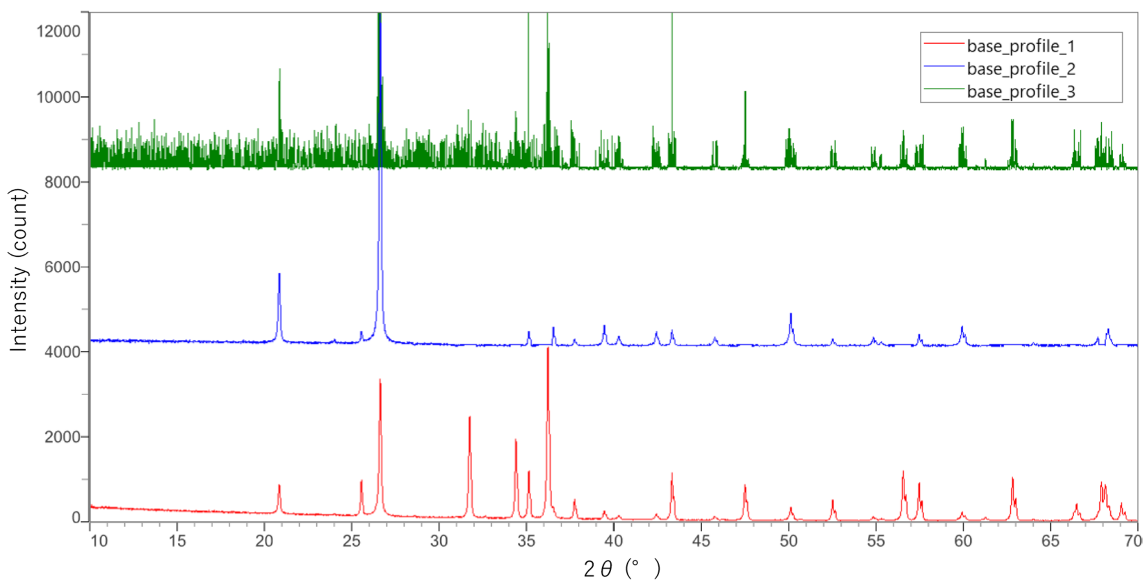

Fig. 11. Basis profiles obtained by the alternating least squares method.

assumed to be different for each material.

Figure 10 shows the basis profiles obtained by this method and the profiles of pure materials. It can be confirmed that the basis profiles corresponding to pure materials are similar to those of pure materials. The basis profiles obtained by the alternating least squares method are shown in Fig. 11 for comparison. It can be seen that the peaks of pure substances exist in several basis profiles. In other words, the accuracy of the decomposition was improved by imposing the constraint. Since it is difficult to regard the basis profiles obtained by the alternating least squares method as the profiles of pure substances, the quantitative values obtained by the DD method are not compared.

4. Conclusion

A method for estimating the profiles of individual components using basis profile decomposition and a method for quantitative analysis of powder X-ray diffraction profiles of mixtures have been described. With the amorphous regularization method, the semi-supervised minimization method, and the constrained alternating least squares method, we have obtained profiles with higher interpretability and accuracy than the alternating least squares method by incorporating the measured profiles in the powder X-ray diffraction method or the known information of the sample.

The methods introduced in this paper can be applied to other measurements. In the literature[8], the amorphous regularization method was applied to the powder X-ray diffraction method of pharmaceutical samples measured while changing temperature and humidity. Phase transition behavior was analyzed without using the known profiles. In the literature[9], the semi-supervised minimization method was applied to the measured X-ray emission spectroscopy spectra of battery samples. By using the known spectra in decomposition, an unknown spectrum derived from a component whose content changes during measurement was successfully extracted. For details, please refer to the respective literature.

References

(1) H. Toraya: *J. Appl. Cryst.*, **49** (2016), 1508–1516.
(2) H. Toraya: *Adv. X-Ray. Chem. Anal.*, *Japan*, **55** (2024), 35–55.
(3) P. Paatero and U. Tapper: *Environmetrics*, **5** (1994), Issue 2, 111–126.
(4) P. Smaragdis, B. Raj and M. Shashanka: *Proc. LVA/ICA*, (2007), 414–421.
(5) *Rigaku Journal*, **31** (2015), No.2, 25–26.
(6) D. R. Hunter and K. Lange: *J. Comput. Graph. Stat.*, **9** (2000), No.1, 60–77.
(7) D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi and K. Kondo: *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, **E97-A** (2014), Issue 5, 1113–1118.
(8) T. Ohta, K. Kihara and A. Sasaki: *Adv. X-ray Anal.*, **68** (2024), 56–63.
(9) T. Ohta, H. Takahara and H. Kobayashi: *Adv. X-ray Anal.*, **68** (2024), 77–84.

