HOKUGA 北海学園学術情報リポジトリ
学校法人北海学園
北海学園大学
北海商科大学

タイトル	Two-Dimensional Correlation of Fractional Derivative Spectra in Near-Infrared Region for Analyzing Correlation with Constituent Concentrations
著者	UOZUMI, Jun
引用	北海学園大学工学部研究報告(46): 133-140
発行日	2019-02-14

Hokkai-Gakuen Organization of Knowledge Ubiquitous through Gaining Archives.

Two-Dimensional Correlation of Fractional Derivative Spectra in Near-Infrared Region for Analyzing Correlation with Constituent Concentrations

Jun UOZUMI*

Abstract

For the purpose of analyzing complex structures of fractional and fractional absolute derivative spectra in near-infrared spectroscopy, two-dimensional correlation spectroscopy is applied by extending it to represent correlation with the constituent concentration of interest. To this end, average over external variable for the ordinary 2D correlation spectroscopy is replaced by ensemble average over many samples and inserting a variable for the concentration of the target constituent. The third-order correlation defined by a simple insertion of the concentration variable is found to convey no significant information of the constituent. The fourth-order correlation defined by an insertion of the square of the concentration variable is found to work for this purpose. It is shown that care must be taken in interpreting auto- and cross-peaks in the fourth-order correlation spectra because of the square insertion of the variable.

1. Introduction

In near-infrared (NIR) spectroscopy, derivatives of spectra sometimes play a key role in improving prediction performance.[1)] Since the higher-order differentiation enhances high frequency noise components as well as fine structures in the spectra, however, the order of traditional derivatives is practically limited to one and two. To introduce additional flexibility in the order of differentiation, it can be extended to any positive number ν by means of scaling filtering in the Fourier domain, which leads to a fractional derivative (FD).[2,3)] Shift and inversion of peaks associated with first, second and fractional derivatives sometimes make it difficult to identify the exact wavelengths of absorption peaks. Such unfavorable deformation can be suppressed by employing a fractional absolute derivative (FAD).[4–6)] It was shown that FD and FAD give rise to better prediction performance when an adequate derivative order ν is chosen.[7)] As ν increases, however, FD and FAD spectra get complicated gradually, growing finer new peaks, the mutual relations of which become also complicated. In this paper, generalized two-dimensional (2D) correlation spectra are investigated for examining correlation prop-

* Department of Electronics and Information Engineering, Faculty of Engineering, Hokkai-Gakuen University

erties within and among raw, FD and FAD spectra, by introducing a new approach to extract 2D correlation peaks that are correlated with individual constituents.

2. Materials and methods

Samples and data

NIR spectra of 1100–2500 nm with 2 nm interval and chemical values for moisture, amylose and protein concentrations were obtained from 31 samples of rice flour in different varieties. From the spectra, FDs and FADs of various orders ν were calculated after MSC pretreatment using MATLAB, from which 2D correlation spectra of several different types were calculated. The raw spectra after the MSC pretreatment are denoted by $S(\lambda)$ in the following.

Fractional and fractional absolute derivatives

FD of order ν of a function $f(\lambda)$ is denoted and defined by[5,6]

$$D_{\lambda}^{\nu} f(\lambda) = \int_{-\infty}^{\infty} F(\mu)\,(i\,2\pi\mu)^{\nu} \exp(i\,2\pi\mu\lambda)\,d\mu, \tag{1}$$

where $F(\mu)$ is the Fourier transform of $f(\lambda)$. This is an extension of the ordinary derivative to that with an arbitrary positive order ν. As a modification of FD, FAD is denoted and defined by[5,6]

$$D_{|\lambda|}^{\nu} f(\lambda) = \int_{-\infty}^{\infty} F(\mu)\,|2\pi\mu|^{\nu} \exp(i\,2\pi\mu\lambda)\,d\mu, \tag{2}$$

which provides completely non-shifted derivative peaks. As is in ordinary derivatives, FD and FAD defined in eqs. (1) and (2) suffer from high-frequency noises as ν increases. To suppress them, a Gaussian low-pass filter is actually inserted in the integrands in eqs. (1) and (2).

Two-dimensional correlation analysis

Generalized 2D correlation spectroscopy is a powerful tool for analyzing correlation properties between same or different spectra at different wavelengths.[8] The 2D correlation spectrum of spectra $A(\lambda)$ and $B(\lambda)$ are defined by

$$R(\lambda_1, \lambda_2) = \left\langle \Delta A(\lambda_1, t)\, \Delta B(\lambda_2, t) \right\rangle_t, \tag{3}$$

where $\Delta A = A - \langle A \rangle_t$ and $\Delta B = B - \langle B \rangle_t$, and $\langle\ \rangle_t$ stands for an average with respect to an external variable t. Since we do not use any controllable external variable in this paper, however, we replace $\langle\ \rangle_t$ by an ensemble average $\langle\ \rangle$ over many samples, and therefore we have

$$R(\lambda_1, \lambda_2) = \langle \Delta A(\lambda_1)\, \Delta B(\lambda_2) \rangle . \tag{4}$$

In this case, the 2D spectrum of eq. (4) expresses overall correlations reflecting variations of all the constituents in the samples.

In many cases in NIR spectroscopy, 2D correlation peaks are required that reflect correlations exclusively with a certain target constituent. Though it is possible to employ the constituent concentration as the external perturbation variable to this end,[9] a relatively large number of samples are needed. As an alternative approach, we consider correlations of a higher order. The first idea would be the third order correlation involving the concentration p of a constituent of interest :

$$R_3(\lambda_1, \lambda_2) = \langle \Delta A(\lambda_1)\, \Delta B(\lambda_2)\, \Delta p \rangle, \tag{5}$$

where $\Delta p = p - \langle p \rangle$. This function seems to have high values where the three variables vary all in phase. In reality, however, this is not the case, as seen from the fact that any odd order moment of a zero-mean Gaussian variate vanishes. Let us consider then the fourth order correlation of the form of

$$R_4(\lambda_1, \lambda_2) = \langle \Delta A(\lambda_1)\, \Delta B(\lambda_2) (\Delta p)^2 \rangle . \tag{6}$$

We also discuss normalized versions of 2D correlation spectra in eqs. (4)–(6). Equation (4) is normalized in a common manner to be

$$\rho(\lambda_1, \lambda_2) = \frac{R(\lambda_1, \lambda_2)}{\sigma_A(\lambda_1)\sigma_B(\lambda_2)} = \frac{\langle \Delta A(\lambda_1)\, \Delta B(\lambda_2) \rangle}{\sigma_A(\lambda_1)\sigma_B(\lambda_2)}, \tag{7}$$

where $\sigma_A(\lambda)$ and $\sigma_B(\lambda)$ stand for the standard deviations of $A(\lambda)$ and $B(\lambda)$, respectively. To normalize $R_4(\lambda_1, \lambda_2)$, we consider that the fourth-order correlation of any zero-mean Gaussian variates reduces to the sum of the pair products of the second-order correlations,[10] namely,

$$R_4(x_1, x_2, x_3, x_4) = \langle x_1 x_2 x_3 x_4 \rangle = \langle x_1 x_2 \rangle \langle x_3 x_4 \rangle + \langle x_1 x_3 \rangle \langle x_2 x_4 \rangle + \langle x_1 x_4 \rangle \langle x_2 x_3 \rangle . \tag{8}$$

This relation indicates that the fourth-order correlation should be normalized as

$$\rho_4(x_1, x_2, x_3, x_4) = \frac{\langle x_1 x_2 x_3 x_4 \rangle}{3\sigma_1\sigma_2\sigma_3\sigma_4} = \frac{1}{3}\left(\frac{\langle x_1 x_2 \rangle}{\sigma_1\sigma_2} \frac{\langle x_3 x_4 \rangle}{\sigma_3\sigma_4} + \frac{\langle x_1 x_3 \rangle}{\sigma_1\sigma_3} \frac{\langle x_2 x_4 \rangle}{\sigma_2\sigma_4} + \frac{\langle x_1 x_4 \rangle}{\sigma_1\sigma_4} \frac{\langle x_2 x_3 \rangle}{\sigma_2\sigma_3} \right), \tag{9}$$

since $\rho_4 = 1$ when $x_1 = x_2 = x_3 = x_4$. Therefore, $R_4(\lambda_1, \lambda_2)$ can be normalized as

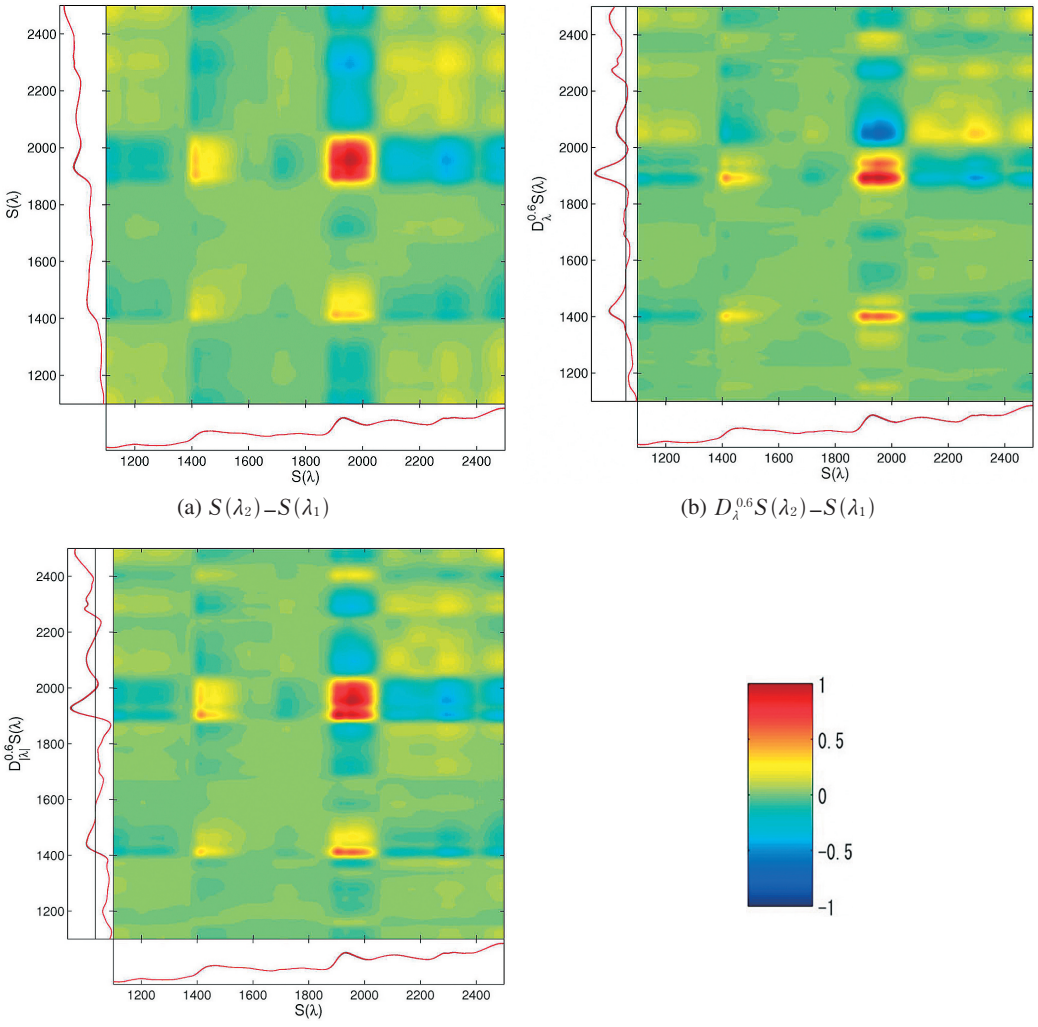
$$\rho_4(\lambda_1, \lambda_2) = \frac{R_3(\lambda_1, \lambda_2)}{3\sigma_A(\lambda_1)\sigma_B(\lambda_2)\sigma_p^2} = \frac{\langle \Delta A(\lambda_1)\, \Delta B(\lambda_2) (\Delta p)^2 \rangle}{3\sigma_A(\lambda_1)\sigma_B(\lambda_2)\sigma_p^2}. \tag{10}$$

Though we do not have a suitable reasoning like eq. (8) for the third-order correlation, let us normalize it similarly as eq. (10) at this moment;

$$\rho_3(\lambda_1,\lambda_2)=\frac{R_3(\lambda_1,\lambda_2)}{3\sigma_A(\lambda_1)\sigma_B(\lambda_2)\sigma_p}=\frac{\langle \Delta A(\lambda_1)\Delta B(\lambda_2)\Delta p\rangle}{3\sigma_A(\lambda_1)\sigma_B(\lambda_2)\sigma_p}. \tag{11}$$

3. Results and Discussions

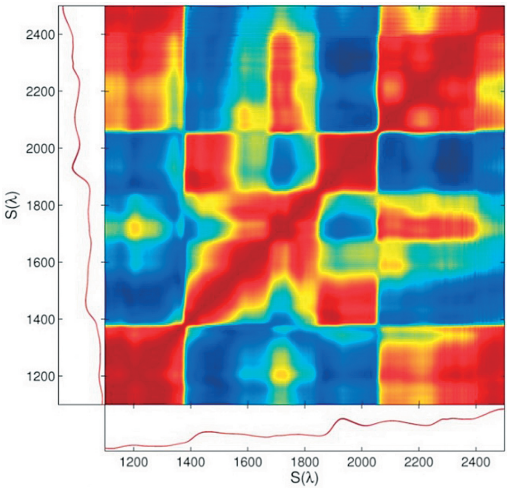
Figures 1(a)–(c) show 2D correlation spectra $R(\lambda_1,\lambda_2)$ of (a) raw spectra $S(\lambda)$, (b) FD spectra of 0.6th order and (c) FAD spectra of 0.6th order, respectively, all versus the raw spectra $S(\lambda)$. These spectra are denoted here by $R(\lambda_1,\lambda_2):S(\lambda_2)–S(\lambda_1)$, $R(\lambda_1,\lambda_2):D_{\lambda}^{0.6}S(\lambda_2)–S(\lambda_1)$ and $R(\lambda_1,\lambda_2):D_{|\lambda|}^{0.6}S(\lambda_2)–S(\lambda_1)$, respectively, or $R:S–S$, $R:D_{\lambda}^{0.6}S–S$ and $R:D_{|\lambda|}^{0.6}S–S$, respectively, for short.



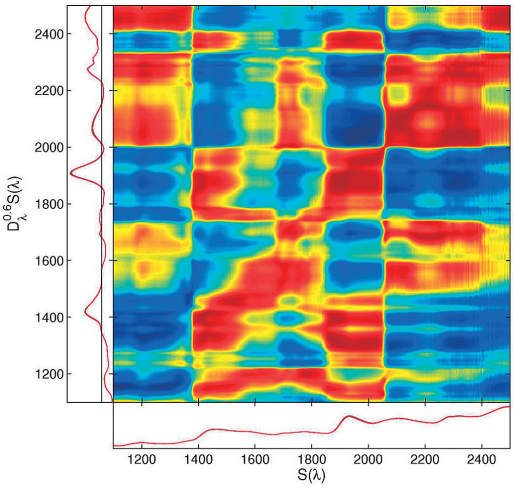
(a) $S(\lambda_2)–S(\lambda_1)$ (b) $D_{\lambda}^{0.6}S(\lambda_2)–S(\lambda_1)$

(c) $D_{|\lambda|}^{0.6}S(\lambda_2)–S(\lambda_1)$ (d) Color map for the correlation value.

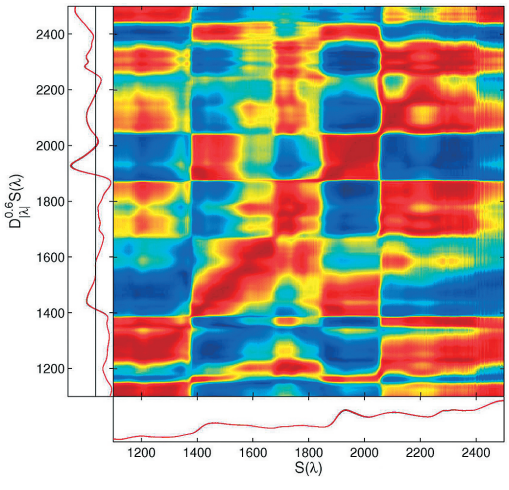
Fig. 1 2D correlation spectra $R(\lambda_1,\lambda_2)$ of the second order.



(a) $S(\lambda_2)-S(\lambda_1)$



(b) $D_{\lambda}^{0.6}S(\lambda_2)-S(\lambda_1)$



(c) $D_{|\lambda|}^{0.6}S(\lambda_2)-S(\lambda_1)$

Fig. 2 Normalized 2D correlation spectra $\rho(\lambda_1, \lambda_2)$ of the second order.

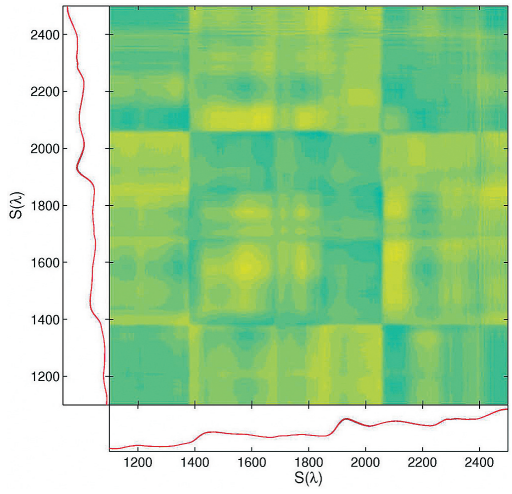


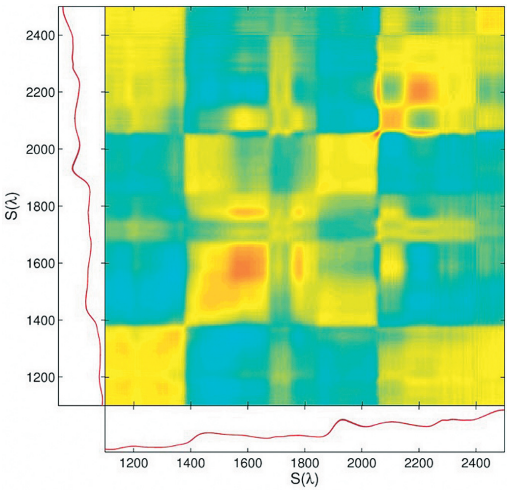
Fig. 3 Normalized 2D correlation spectra $\rho_3(\lambda_1, \lambda_2)$: $S(\lambda_2)-S(\lambda_1)$ of the third order.

The color map for the 2D correlation value is shown in **Fig. 1**(d). In **Fig. 1**(a), a strong autopeak appears at 1900−2030 nm due to a strong absorption peak in this region. Similar quasi-autopeaks also appear in **Figs. 1**(b) and (c). The term "quasi-" is used because the peaks are separated slightly in λ_2 direction into smaller ones and deformed from the autopeak as they would be without differentiation by λ_2. The shape of these quasi-autopeaks expresses how the correlation is deformed and separated by fractional differentiation. Note that the correlation peaks in **Fig. 1**(b) are slightly shifted to smaller λ_2 (downward) as compared with **Figs. 1**(a) and (c) due to the peak-shift effect of FD. In this respect,

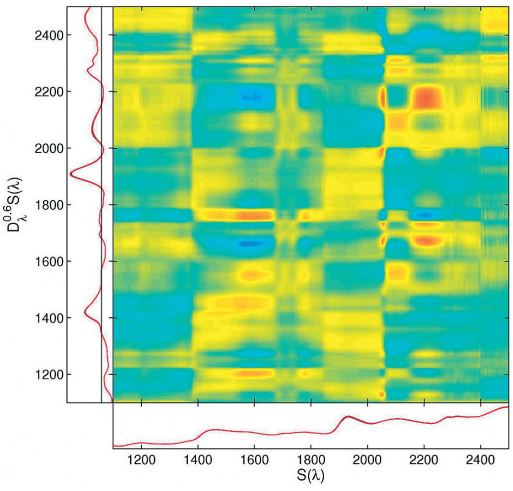
the spectrum $R : D_{|\lambda|}^{0.6} S - S$ is rather similar to $R : S - S$ as compared with $R : D_{\lambda}^{0.6} S - S$.

Dependence of the 2D correlation spectrum $R(\lambda_1, \lambda_2)$ on the magnitude of the absorbance can be removed by normalizing it in the form of eq. (7), results corresponding to **Figs. 1**(a)–(c) being shown in **Figs. 2**(a)–(c), respectively. The correlation peaks in **Figs. 2**(b) and (c) are again separated in λ_2 direction due to the effect of differentiation. In these figures, however, very many correlation peaks appear due to the contributions from all the constituents, making interpretation of the spectrum difficult.

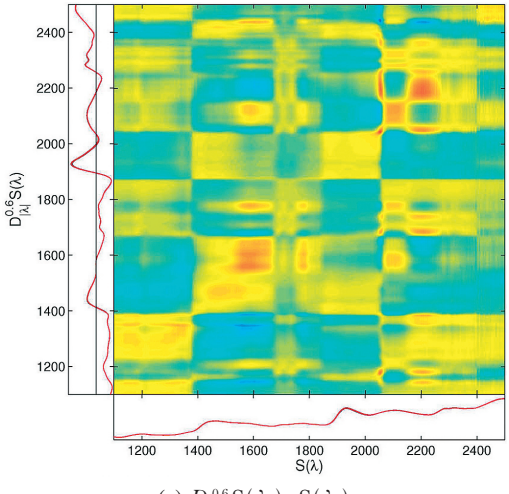
To extract correlation peaks that are correlated with a certain constituent, a normalized third-order correlation ρ_3: $S(\lambda_2) - S(\lambda_1)$ was calculated from eq. (11) with protein as the target constituent and is



(a) $S(\lambda_2) - S(\lambda_1)$



(b) $D_{\lambda}^{0.6} S(\lambda_2) - S(\lambda_1)$



(c) $D_{|\lambda|}^{0.6} S(\lambda_2) - S(\lambda_1)$

Fig. 4 Normalized 2D correlation spectra $\rho_4(\lambda_1, \lambda_2)$.

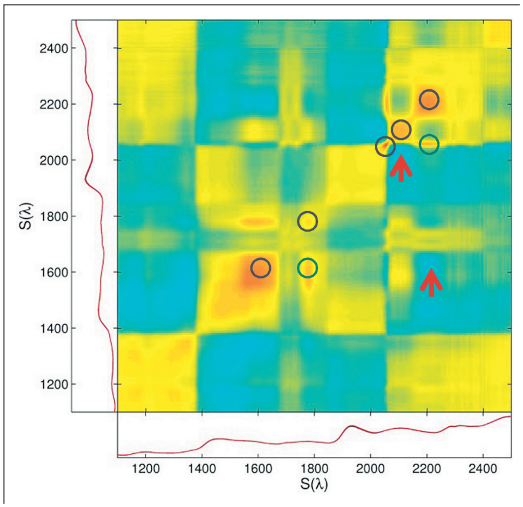


Fig. 5 $\rho_4(\lambda_1, \lambda_2)$: $S(\lambda_2) - S(\lambda_1)$ with some marks at auto- and cross-peaks.

shown in **Fig. 3**. As predicted theoretically, the third-order correlation spectrum is almost vanishing, and it is confirmed that this type of correlation conveys almost no information of the target constituent.

According to the theoretical result of eq. (10), the fourth-order correlation $\rho_4(\lambda_1, \lambda_2)$ was calculated with protein as the target constituent and is shown in **Figs. 4**(a)–(c) for the same combinations of spectra as **Figs. 1** and **2.** This figure clearly shows that the 2D correlation peaks focused on protein are extracted and it is expected that the cross-peaks provide information of the relation between two spectra, such as those before and after fractional differentiation.

To examine correlation peaks in **Fig 4**(a), this spectrum is show again in **Fig. 5** with some symbols. In this spectrum, some quasi-autopeaks, e.g., at 1570, 1780, 2050, 2100 and 2200 nm are observed as indicated by blue circles, and cross-peaks are seen at (1570, 1780) and (2050, 2200) nm as indicated by green circles. However, there are no cross peak at (2050, 2100) and (1570, 2200) nm, where two arrows point. Care should be taken in interpreting ρ_4 since the sign of variations in Δp is not distinguished in eq. (10). The wavelengths of 2050 and 2200 nm are known to be assigned to protein, and there does exist the cross-peak between these wavelengths. However, the quasi-autopeaks at 1570, 1780 and 2100 nm prove to be false correlation with protein since these wavelengths do not have cross-peaks with 2050 or 2200 nm. Namely, if $\Delta A(\lambda_1)$ increases and $\Delta B(\lambda_2)$ decreases when Δp increases, autopeaks will appear both at λ_1 and λ_2 because $\langle [\Delta A(\lambda_1)]^2 (\Delta p)^2 \rangle$ and $\langle [\Delta B(\lambda_1)]^2 (\Delta p)^2 \rangle$ are positive, but $\langle \Delta A(\lambda_1) \Delta B(\lambda_2) (\Delta p)^2 \rangle$ is negative. On the other hand, cross-peaks at 1670 and 1720 nm correlated with 2200 nm show that they become correlated with protein for $\nu = 0.6$, as is seen in **Fig. 4**(c), while not correlated in the raw spectrum. Therefore, it is concluded that the fourth-order correlation $\rho_4(\lambda_1, \lambda_2)$ is suitable for analyzing structural change in NIR spectra as the derivative order changes in FD or FAD.

4. Conclusions

Derivative spectra of NIR spectroscopy have very many peaks produced by decomposition of overlapped or complex peaks. In particular, spectral structures of fractional and fractional absolute derivatives change in a complex manner depending upon the order of differentiation. To help examining those structures, two-dimensional correlation spectroscopy is applied by extending it to be correlated with the constituent concentration of interest. To this end, average over external variable for the ordinary 2D correlation spectroscopy is replaced by ensemble average over many samples and inserting a variable for the concentration of the target constituent. The third-order correlation defined by a simple insertion of the concentration variable is found to convey no significant information of the constituent.

The fourth-order correlation defined by a insertion of the square of the concentration variable is found to work well for this purpose. A suitable way of normalization is given and shown to work correctly. It is shown that care must be taken in interpreting auto- and cross-peaks in the fourth-order correlation spectra because of the square insertion of the variable. It is shown that the proposed 2D correlation spectra are effective in interpreting complex structures of fractional and fractional absolute derivative spectra in NIR region.

The content of this paper has been presented in conferences and has been partly published in their abstracts.[11, 12)]

REFERENCES

1. H. W. Siesler, Y. Ozaki, S. Kawata and H. M. Heise : *Near-Infrared Spectroscopy*, Wiley-VCH, Weinhelm, Germany (2002).
2. J. M. Schmitt : Fractional derivative analysis of diffuse reflectance spectra, *Appl. Spectrosc.*, 52(6) pp. 840–846 (1998).
3. C.-C. Tseng, S.-C. Pei and S.-C. Hsia : Computational fractional derivatives using Fourier transform and digital FIR differentiator, *Signal Processing*, 80, pp. 151–159 (2000).
4. D. K. Buslov : Modification of derivatives for resolution enhancement of bands in overlapped spectra, *Appl. Spectrosc.*, 58(11) pp. 1302–1307 (2004).
5. J. Uozumi and H. Ogasawara : MLR analysis of rice by means of extended derivatives of NIR spectra, *Proc. Japan–Thailand Joint Symp. Nondestructive Evaluation Technol.*, p. 103 (2004).
6. J. Uozumi : Phase-controlled fractional derivatives for near infrared spectral processing, *Asian Journal of Physics*, 27 (9–12) (2018), to be published.
7. J. Uozumi and M. Murata : Optimized noise reduction in extended derivative spectroscopy for NIR calibration of rice constituents —A preliminary study, *Proc. Japan-Korea Joint Symp. Near Infrared Spectrosc.*, p. 195 (2006).
8. I. Noda and Y. Ozaki, *Two-Dimensional Correlation Spectroscopy*. Wiley, Chichester, UK (2004).
9. Y. Liu, S. R. Delwiche and R. A. Graybosch : Two-dimensional correlation analysis of near infrared spectral intensity variations of ground wheat, *J. Near Infrared Spectrosc.*, 17, pp. 41–50 (2009).
10. A. Papoulis and S. U. Pillai : *Probability, Random Variables and Stochastic Processes*, McGraw Hill, New York, USA (2002), p. 258.
11. J. Uozumi and Y. Kawakami : Two-dimensional correlation analysis of NIR fractional derivative spectra : Correlation with constituent concentrations, *Abstract Book of NIR-2009* (The Fourteenth International Conference on Near Infrared Spectroscopy (NIR-2009), Bangkok, Thailand, November 7–16, 2009), p. 72.
12. J. Uozumi : 2D correlation spectra based on higher order moments for correlation analysis with constituent concentrations, *Final Program Book of Abstracts of FACSS 2010* (Federation of Analytical Chemistry and Spectroscopy Societies, Raleigh, USA, October 17–21, 2010), p. 187.