Two-dimensional correlation analysis of NIR fractional derivative spectra: correlation with constituent concentrations

Jun Uozumi and Yoshika Kawakami[a]
Faculty of Engineering, Hokkai-Gakuen University, Sapporo 064–0926 Japan.
E-mail: uozumi@eli.hokkai-s-u.ac.jp
[a]Present address: East Japan Railway Company

Introduction

Derivatives of spectra sometimes play a key role in improving prediction performance of NIR spectroscopy. Although the order of traditional derivatives is practically limited to one and two, 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).[1] 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).[1] It was shown that FD and FAD give rise to better prediction performance when an adequate derivative order ν is chosen.[2] As ν increases, however, FD and FAD spectra gradually get complicated, growing finer new peaks, the mutual relations of which become also complicated. In this paper, generalised two-dimensional (2D) correlation spectra are investigated for examining correlation properties within and among raw, FD and FAD spectra, by introducing a new approach to extract 2D correlation peaks that are correlated with individual constituents.

Materials and methods

Samples and data

NIR spectra of 1100–2500 nm with a 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[1]

$$D_{\lambda}^{\nu} f(\lambda) = \int_{-\infty}^{\infty} F(\mu)(i2\pi\mu)^{\nu} \exp(i2\pi\mu\lambda) d\mu, \tag{1}$$

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[1]

$$D_{|\lambda|}^{\nu} f(\lambda) = \int_{-\infty}^{\infty} F(\mu)|2\pi\mu|^{\nu} \exp(i2\pi\mu\lambda) d\mu, \tag{2}$$

which provides completely non-shifted derivative peaks. As is in ordinary derivatives, FD and FAD defined in Equations (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 Equations (1) and (2).

Two-dimensional correlation analysis

Generalised 2D correlation spectroscopy is a powerful tool for analysing correlation properties between same or different spectra at different wavelengths.[3] The 2D correlation spectrum of spectra $A(\lambda)$ and $B(\lambda)$ are defined by

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

where $\Delta A = A - \langle A \rangle$ and $\Delta B = B - \langle B \rangle$, and $\langle\ \rangle$ 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 omit it and regard $\langle\ \rangle$ as an average over samples. In this case, the 2D spectrum of Equation (3) 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,[4] 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:

$$R_3(\lambda_1, \lambda_2) = \langle \Delta A(\lambda_1) \Delta B(\lambda_2) \Delta p \rangle, \tag{4}$$

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 zero-mean Gaussian variate vanishes. Let us consider then the fourth order correlation of the form of

$$R_4(\lambda_1, \lambda_2) = \langle \Delta A(\lambda_1) \Delta B(\lambda_2) (\Delta p)^2 \rangle. \tag{5}$$

We also discuss normalised versions of 2D correlation spectra in Equations (3)–(5) defined as

$$\rho(\lambda_1,\lambda_2)=\frac{R(\lambda_1,\lambda_2)}{\sigma_A(\lambda_1)\sigma_B(\lambda_2)}=\frac{\langle\Delta A(\lambda_1)\Delta B(\lambda_2)\rangle}{\sigma_A(\lambda_1)\sigma_B(\lambda_2)}, \tag{6}$$

$$\rho_3(\lambda_1,\lambda_2)=\frac{R_3(\lambda_1,\lambda_2)}{3\sigma_A(\lambda_1)\sigma_B(\lambda_2)\sigma_p}=\frac{\langle\Delta A(\lambda_1)\Delta B(\lambda_2)\Delta p\rangle}{3\sigma_A(\lambda_1)\sigma_B(\lambda_2)\sigma_p}, \tag{7}$$

$$\rho_4(\lambda_1,\lambda_2)=\frac{R_4(\lambda_1,\lambda_2)}{3\sigma_A(\lambda_1)\sigma_B(\lambda_2)\sigma_p^2}=\frac{\langle\Delta A(\lambda_1)\Delta B(\lambda_2)(\Delta p)^2\rangle}{3\sigma_A(\lambda_1)\sigma_B(\lambda_2)\sigma_p^2}. \tag{8}$$

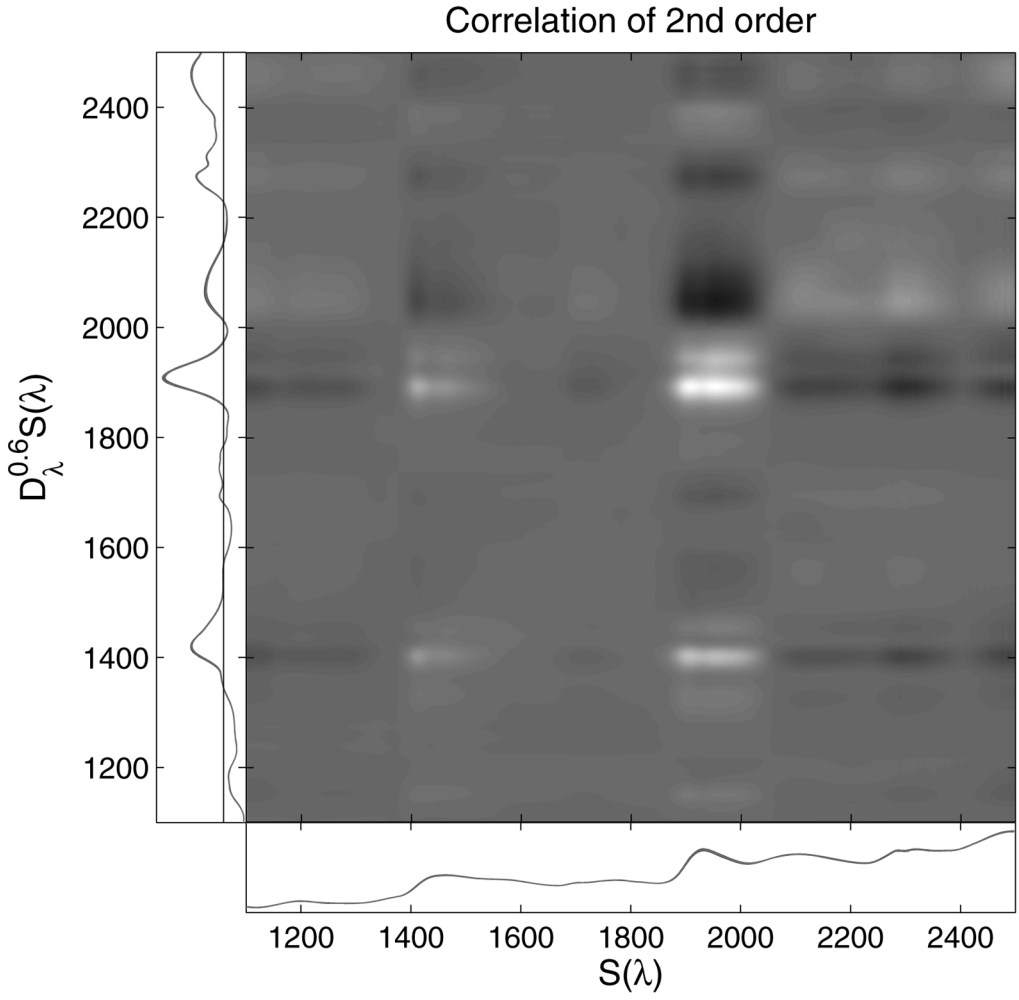

Figure 1. Correlation $R(\lambda_1, \lambda_2)$ of raw and 0.6th order FD spectra.

Results and discussion

Figure 1 shows a generalised 2D correlation spectrum $R(\lambda_1, \lambda_2)$ of raw spectra $S(\lambda)$ (x-axis) and FD spectra of 0.6th order (y-axis), which is denoted here by $R(\lambda_1,\lambda_2):D_{\lambda}^{0.6}S(\lambda_2)-S(\lambda_1)$, or $R:D_{\lambda}^{0.6}S-S$ for short.

A strong quasi-autopeak appears at 1900–2030 nm due to a strong absorption peak in this region. The term "quasi-" is used because the peak is separated slightly in λ_2 direction and deformed from the autopeak as it would be without differentiation by λ_2. The shape of this quasi-autopeak expresses how the correlation is deformed and separated by fractional differentiation. Figure 2 shows a FAD version of Figure 1, $R:D_{|\lambda|}^{0.6}S-S$.

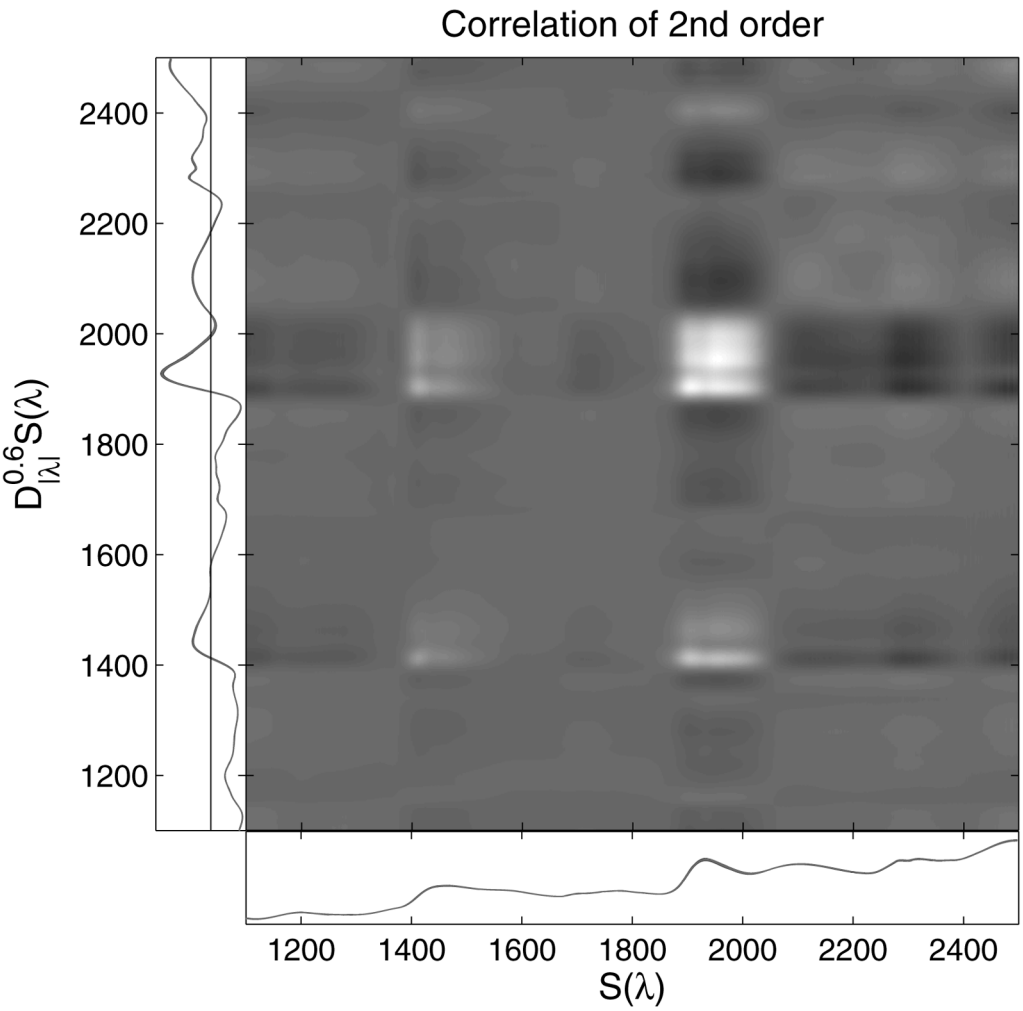

Figure 2. Correlation R(λ_1, λ_2) of raw and 0.6th order FAD spectra.

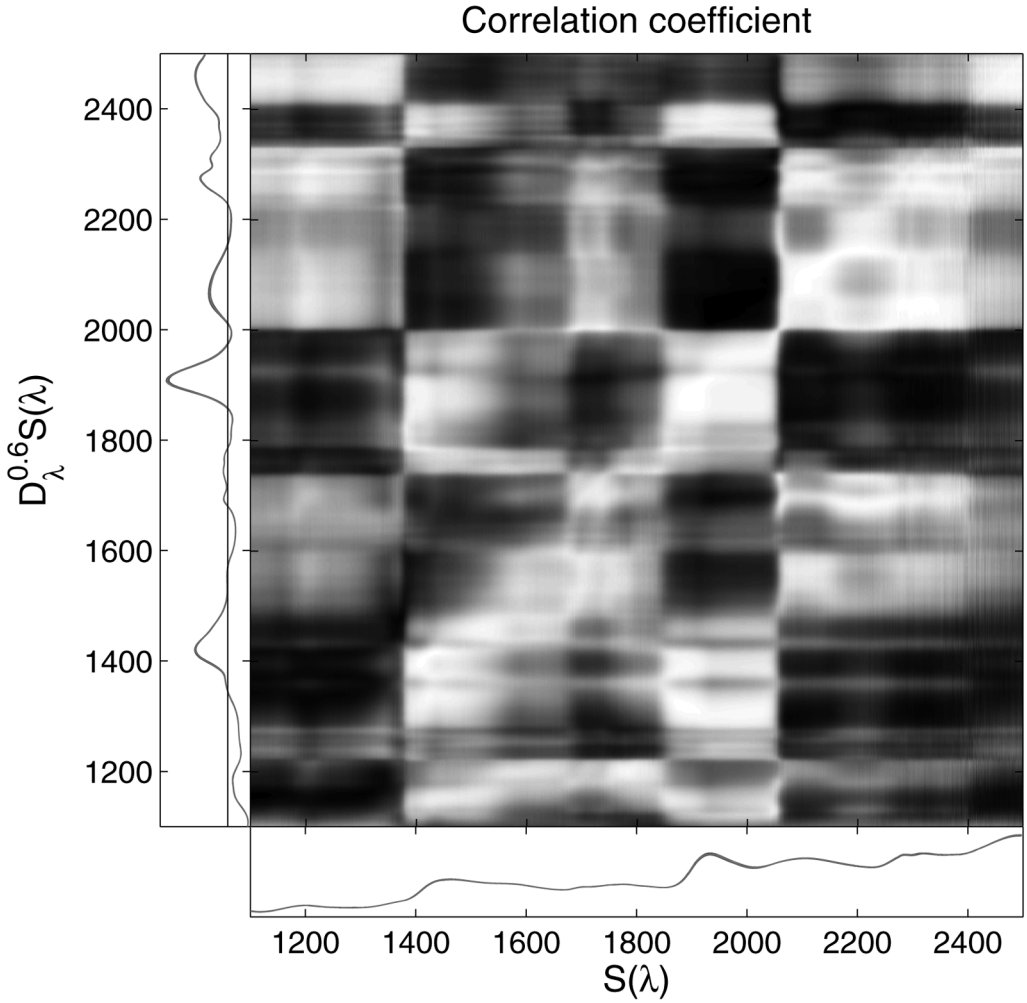

Figure 3. Normalised correlation $\rho(\lambda_1, \lambda_2)$ of raw and 0.6th order FD spectra.

Note that the correlation peaks in Figure 1 are slightly shifted to smaller λ_2 (downward) as compared with Figure 2 due to the peak-shift effect of FD. It follows therefore that the spectrum $R:D_{|\lambda|}^{0.6}S-S$ is rather similar to $R:S-S$ 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 normalising it in the form of Equation (6), a result for $\rho:D_\lambda^{0.6}S-S$ being shown in Figure 3.

In this figure, however, very many correlation peaks are contributed from different constituents, making interpretation of the spectrum difficult. To extract correlation peaks that are correlated with a certain constituent, a normalized fourth-order correlation $\rho_4:D_\lambda^{0.6}S-S$ was calculated with protein as the target constituent and is shown in Figure 4.

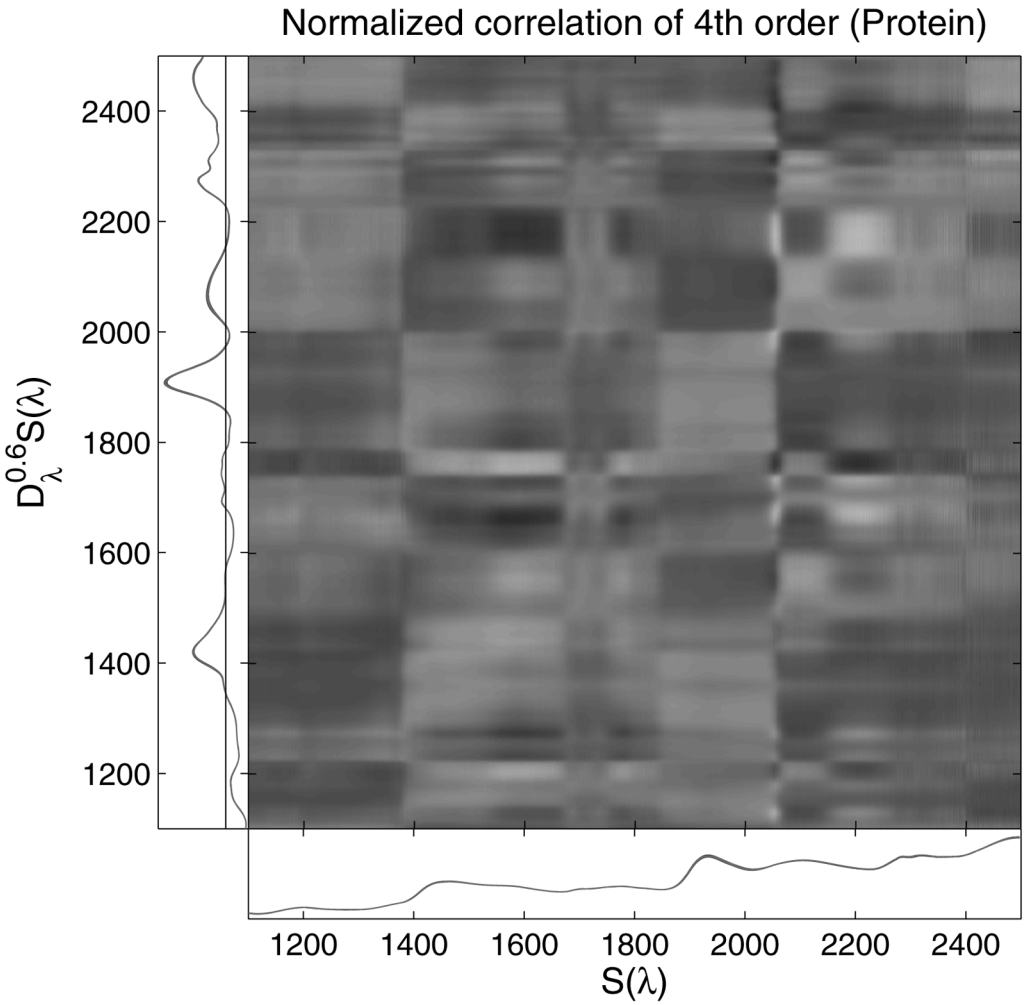

Figure 4. Normalised correlation $\rho_4(\lambda_1, \lambda_2)$ of raw and 0.6th order FD spectra and protein concentration.

In this figure, some quasi-autopeaks, e.g., at 1570, 1780, 2050, 2100 and 2200 nm are observed and cross peaks are seen at (1570, 1780) and (2050, 2200) nm. However, there are no cross peak at (2050, 2100) and (1570, 2200) nm. Care should be taken in interpreting ρ_4 since the sign of variations in Δp is not distinguished in Equation (8). For 2050 and 2200 nm are known to be assigned to protein, the quasi-autopeaks at 1570, 1780 and 2100 nm prove to be false correlation with protein. On the other hand, cross peaks at 1670 and 1720 nm correlated with 2200 nm show that they become correlated with protein for $\nu = 0.6$, while not correlated in the raw spectrum.

2D correlation spectra given by Equation (7), such as $\rho_3 : S - S$, were also calculated and found to exhibit no significant peaks as predicted. Therefore, it is concluded that the fourth-order corre-

lation $\rho_4(\lambda_1, \lambda_2)$ is suitable for analysing structural change in NIR spectra as the derivative order changes in FD or FAD.

References

1. J. Uozumi and H. Ogasawara, *Proc. Japan-Thailand Joint Symp. Nondestructive Evaluation Technol*., p. 103 (2004).
2. J. Uozumi and M. Murata, *Proc. Japan-Korea Joint Symp. Near Infrared Spectrosc*., p. 195 (2006).
3. I. Noda and Y. Ozaki, *Two-Dimensional Correlation Spectroscopy.* Wiley, Chichester, UK (2004).
4. Y. Liu, S.R. Delwiche and R.A. Graybosch, *J. Near Infrared Spectrosc.* **17,** 41 (2009).