How much of a near infrared spectrum is useful? Sparse regularization—let the data decide!

R. S. Anderssen[1] F. R. de Hoog[2] I. J. Wesley[3]
A. B. Zwart[4]

(Received 11 February 2013; revised 27 June 2014)

Abstract

In information recovery from indirect measurements of the phenomenon of interest (e.g., near infrared spectra of milk powders or pharmaceuticals, or Raman spectra of explosives or anaesthetics) the available data can be partitioned into two separate components: (i) the information which encapsulates the answer to the question under examination (the proportion of casein, the major protein component, in milk powder, the presence or absence of explosives, the monitoring of anaesthetic and respiratory levels during surgery); and (ii) a considerable amount of superfluous information, the presence of which compromises the reliability of the answer to the question of interest. In

http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/6744
 Published August 12, 2014, as part of the Proceedings of the 16th Biennial Computational Techniques and Applications Conference. ISSN 1446-8735. (Print two pages per sheet of paper.)

such spectroscopic situations, for the identification of the information that encapsulates the answer, a variety of techniques are used such as partial least squares, neural networks and support vector machines. With respect to the available calibration data, the support vector machines procedure performs an *implicit* form of *sparse regularization*. In this article, the aim is to show how, using the Beer–Lambert law and derivative spectroscopy, the sparse regularization is performed in an *explicit* manner. This information can be subsequently utilized to construct, using statistical regression, an appropriate predictor. Here, the goal is to give a *proof-of-concept* for the application of derivative spectroscopy as an explicit sparse regularization protocol. For this, the calibration data consists of near infrared spectra of milk powder spiked with known amounts of casein, while the property of interest is the proportion of casein in the milk powder.

Contents

1 Introduction

Because of the speed and accuracy with which near infrared (NIR) spectra of biological and non-biological samples can now be recorded, the information that they contain is playing an increasingly relevant role in daily decision making. Is the mango ready to pick? Has the coffee powder been adulterated? Is the wine in the bottle the same as stated on the label? How much should the farmer be paid for the wheat? What is the casein content in this milk powder sample?

In order to answer such questions, the initial step is to identify wavelength regions in the recorded spectra of the samples that carry information relevant to the specific question under consideration. The relevant wavelength regions change as the questions change. The reason for this change is that a spectrum records the molecular vibrations of all the different components making up a material, whereas the information required for decision making usually only relates to the presence or absence of some specific component in the material: the sugar content of the mango; the presence of starch in the coffee powder; the components that are stated on the label; the gluten in the wheat; the casein in the milk powder. In addition, because the molecular structure of the individual components of a material (e.g., casein in milk powder) are different, their molecular vibrations occur in different distinct wavelength bands. For example, at 20°C, the NIR wavelength intervals where water molecules vibrate are centred around 760, 970, 1190, 1450, 1940 nm. For proteins, there are always vibrations in the wavelength interval around 2180 nm, which correspond to amide linkage in the protein backbone.

The situation is complicated by overlapping wavelength bands of different components of the material. For example, the wavelength bands of the amylose and amylopectin components of starch have strong overlapping [8]. In addition, the protein vibrations centered around 2180 nm tend to be confounded by the absorbance due to starch at 2100 nm [7]. Consequently, where possible, for the component of interest, one is interested in the wavelength intervals

which are not confounded by molecular vibrations associated with some other component(s). If such intervals are present, then they are the ideal wavelength regions to utilize for the construction of a predictor, using an appropriate regression procedure.

The aim of this article is to show how derivative spectroscopy is utilized to determine whether or not there are wavelength intervals where no overlapping occurs. The basic assumption that underlies the proposed methodology is that, for the spectra that make up the calibration data, no overlapping occurs for the component of interest if there exist wavelengths for which the ordering of the fourth derivative values of the spectra are exactly the same as the proportions of that component in the sample. As explained algebraically in Section 3.1, this is implemented computationally by using some appropriate measure of how well the ordering of the derivative values correlates with the ordering of the proportional presences. Any correlation measure would be appropriate and would represent a useful expedient for asssessing the level of confounding as a function of wavelength. The importance of this approach is that it naturally allows for the possibility that there are no wavelength intervals where there is no confounding. When this occurs, it is important to know where the confounding is minimal, since this represents an identifier for the wavelength regions which have a strong connection to the property to be calibrated and predicted.

The aim is to give a proof-of-concept for the proposed method. For this, the calibration data are the NIR spectra for milk powder spiked with known amounts of casein, while the property of interest is the proportions of casein in the milk powder.

Traditionally, the identification of the wavelength intervals was performed using some appropriate calibration procedure such as partial least squares (PLS), neural networks or support vector machines. The calibration data consists of a set of representative spectra and the associated measured values of the property of interest (e.g., proportions of some key component) for which a predictor is to be constructed. Such methods perform and exploit an

implicit form of *sparse regularization*. They approximately identify, via some iterative procedure applied to the calibration data, a subset of wavelength intervals for which there is a strong correlation with the values of the property and simultaneously utilize this information in the construction of the required predictor, exploiting the linearity of the relationship between the spectra and the values of the property. The essential background about NIR spectra and sparse regularization is discussed in Section 2.

The relevance of the proposed derivative spectroscopy methodology is that it represents an *explicit* sparse regularization protocol in that the existence or otherwise of non-confounded wavelength intervals is determined directly from the values of the fourth derivatives of the spectra. The derivative spectroscopy analysis of the spiked milk power data is examined in Section 3.

From a practical perspective, an additional motivation is that the design costs for instruments performing specific tasks, such as monitoring the sugar content in mangoes, is reduced if a small number of non-confounded wavelength intervals are available and known.

2 Background and notation

2.1 Structure of the spectra of the spiked samples

Let $S_{\text{casein}}(\lambda)$ and $S_{\text{non}}(\lambda)$ denote the (theoretical) NIR spectra of pure casein and the non-casein component in milk powder, respectively (ignoring the scattering effect due to sample packing). The $J-1$ spiked samples, numbered $j = 1, 2, \ldots, J-1$, were prepared by adding different proportions of casein p_j, $0 = p_1 < p_2 < \cdots < p_{J-1}$, to unspiked milk powder. Prior to spiking, the milk powder contains proportional amounts α and $(1-\alpha)$ of casein and non-casein, respectively. Therefore, for the jth sample we determine the

proportion of casein and non-casein in the milk powder:

$$\begin{aligned} \text{SS}_j &= p_j\ [\text{added casein}] + (1-p_j)[\alpha + (1-\alpha)]\ [\text{milk powder}] \\ &= [\alpha + (1-\alpha)p_j]\ [\text{casein}] + (1-p_j)(1-\alpha)\ [\text{non-casein}]. \end{aligned}$$

The proportion of casein and non-casein in the jth spiked sample are $a_j = [\alpha + (1-\alpha)p_j]$ and $(1-a_j) = (1-p_j)(1-\alpha)$, respectively. As a direct consequence of the Beer–Lambert law [7], the observed NIR spectral responses $S_{\text{MP},j}(\lambda)$, $j = 1, 2, \ldots, J-1$, of the spiked samples are the proportional sums of the spectral responses of the casein and non-casein components:

$$S_{\text{MP},j}(\lambda) = a_j S_{\text{casein}}(\lambda) + (1-a_j) S_{\text{non}}(\lambda)\,, \quad j = 1, 2, \ldots, J-1\,. \tag{1}$$

The non-spiked milk powder sample corresponds to $p_1 = 0$ and, consequently, to $a_1 = \alpha$. The special case where the sample consists only of casein is now included as the Jth sample with $a_J = 1$.

The ability to decompose the observed spectra of the samples, as shown in equation (1), is the reason why the spectroscopic analysis of a (biological) material can be utilized to determine the proportions of the various components in that material. The underlying molecular rationale is that the measured spectral response of a pure substance, such as casein, at a wavelength λ, is proportional to the number of side chains of that substance vibrating at that wavelength and that the number of side chains is proportional to the weight of the pure substance.

The Beer–Lambert law was initially formulated for fluids [7, 6, 4]. For wavelength dependent absorbance $\sigma(\lambda)$ and path length of absorbance through the fluid ℓ, this law gives the total absorbance of a material with K components:

$$A(\lambda) = \sigma(\lambda)\ell \left(\sum_{k=1}^{K} N_k \right), \tag{2}$$

where N_k denotes the density (number per unit volume) of the kth molecular component.

2.2 NIR spectroscopy analysis

An NIR spectrum of a biological material records, as a function of wavelength λ, the intensity of the vibrations of the various side chains of the molecules in the sample. For a pure substance such as casein (virtually the only protein in unadulterated milk), its NIR spectrum $S_{\text{casein}}(\lambda)$ (Figure 1(a)) is an identifying signature. To remove the background linear particle scatter effect in an NIR spectrum, such as that in $S_{\text{casein}}(\lambda)$, it is common practice to use the second derivative, with respect to wavelength, $S^{(2)}_{\text{casein}}(\lambda) = d^2S_{\text{casein}}/d\lambda^2$ (Figure 1(b)) as the casein signature. However, because it is the negative peaks in $S^{(2)}_{\text{casein}}(\lambda)$ that identify the positive peaks in the measured $S_{\text{casein}}(\lambda)$, it is more natural to work with the fourth derivative of the spectrum $S^{(4)}_{\text{casein}}(\lambda)$ (Figure 1(c)), since its positive peaks are in phase with the positive peaks in $S_{\text{casein}}(\lambda)$.

The *sparse fingerprint* of $S_{\text{casein}}(\lambda)$ is defined to be the set of wavelength intervals $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ at which the intensity of the fourth derivative $S^{(4)}_{\text{casein}}(\lambda)$ exceeds some representative positive threshold δ. The value of δ is chosen to localize the intervals to contain the theoretical wavelengths of the individual side chain vibrations (such as an O-H vibration) with the resulting width of each interval being a measure of the spread of the wavelength vibrational energy caused by the side chain configuration within the material being studied. Importantly, $\mathbb{S}^{(4)}_{\text{casein};0}(\lambda)$ identifies the set of positive peaks in $S^{(4)}_{\text{casein}}(\lambda)$. As illustrated in Figure 1, as well as having its positive peaks in phase with the positive peaks in the spectrum, the fourth derivative performs a localization through its implicit resolution enhancement behaviour. Consequently, a representation such as $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ becomes the sparse identifier for the pure substance. Some representative examples of the intervals identified by $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ for the casein spectrum are plotted in Figure 2 for two different values of δ, with the blue bands corresponding to the smaller value of δ. They illustrate how, as δ increases, the sparsity becomes more pronounced, with the narrower green bands corresponding to the larger value of δ. The blue and green bands are replotted along the wavelength axis to highlight how the sparsity becomes sharper.

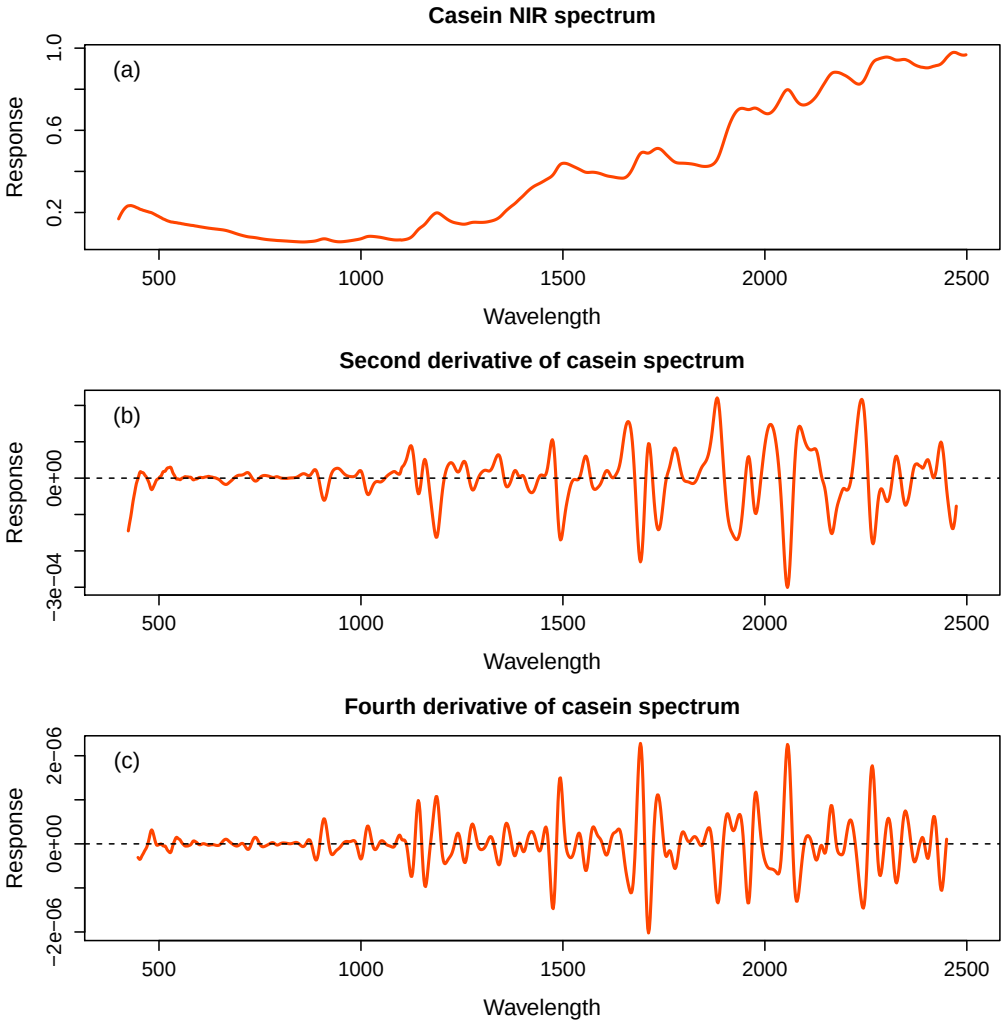

Figure 1: The NIR casein spectrum: (a) the spectrum; (b) the second derivative; (c) the fourth derivative.

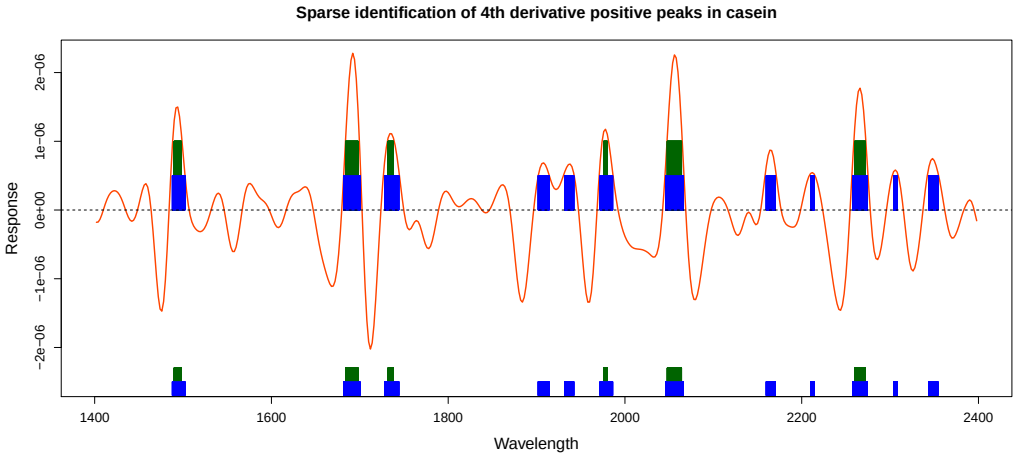

Figure 2: Plots of $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ of the casein spectrum for two different values of δ, illustrating how, as δ increases, the sparsity becomes more pronounced. The green bands correspond to the larger value of δ and are clearly narrower and fewer than the blue bands. Below the spectra, the blue and green bands are replotted along the wavelength axis to highlight the increase in sparsity.

For a mixture of biological molecules consisting of K separate molecular components (as in milk powder or wheat grains), each with spectra $S_k(\lambda)$, $k = 1, 2, \ldots, K$, some of the intervals will overlap forming the sets $\mathbb{S}^{(4)}_{k;\delta}(\lambda)$, $k = 1, 2, \ldots, K$. Such overlapping corresponds to the confounding discussed in Section 1.

2.2.1 The experimental protocol

When the goal is to determine the proportion of some key molecule in a given mixture (such as casein in milk powder), one is given, for a number of representative samples, their (NIR) spectra and the corresponding proportions of the key molecule. Often, the proportions are measured. However, in

many practical situations the actual measurement protocol is time-consuming, expensive and sometimes dangerous. This problematic situation arises when determining the proportional presence of protein in a wheat sample as the experimental/estimation protocol is based on knowing its nitrogen content, since nitrogen is only found in the proteins. In addition, such situations greatly limit the number of spectra and corresponding proportional presences that can be recorded to answer the question under investigation. Because an appropriate simple experimental/estimation protocol is not available, the alternative is to perform calibration and prediction [5, 6, 1]. As mentioned previously, a measured spectrum contains a considerable amount of superfluous information which does not directly help answer the question under consideration. It is the exploitation of this fact that is the rationale behind methodologies such as PLS, neural networks and support vector machine protocols. In one way or another, these methods identify, through a calibration procedure, the set of wavelength intervals that encapsulate the information necessary to define a predictor to answer the question.

The milk powder example described previously is representative of a wide range of practical situations, including testing for adulteration. There, only some representative wavelength intervals associated with the presence of one specific molecule (casein) are required. When a pure sample of the single molecule is available, as in casein and many other adulteration scenarios, the methodology is reversed. The (proportional) presence of the specific molecule in the material of interest is orchestrated by spiking that material with different known amounts of the pure molecule. Through the utilization of derivative spectroscopy, the implementation of sparse regularization is reduced to experimentally determining the intervals $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ where changes are occurring as a result of the spiking.

The goal is the identification of the subset of intervals $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ which identify side chain vibrations in the casein component of milk powder and that do not interact with the non-casein components. This highlights how the application under consideration introduces additional constraints on the nature of the sparse regularization required. The sparse regularization must

take account of cross-interactions that occur between vibrations in some side chain components in the casein and non-casein components of the milk powder. For a given threshold δ, it is necessary to identify the intervals $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ for the side chain vibrations in the casein which do not cross-interact with the non-casein components.

2.3 Derivative spectroscopy analysis of (milk powder) NIR spectra

The focus of the current deliberations is the identification of features in a milk power spectrum that can be utilized to predict proportional casein content. This identification could be done with PLS [5, 6]. However, as already mentioned, the sparse regularization that PLS performs is quite implicit. The nature of the regularization performed by PLS can be explained in terms of simultaneous minimization and total least squares [1].

In this article, the goal is to show how to perform the identification as an explicit sparse regularization procedure by using derivative spectroscopy to exploit the structure in the NIR spectra of milk powder spiked with casein. The essence of the situation is illustrated in Figure 3, where the spectra of normal milk powder and pure casein are plotted, as well as the difference between the two spectra. In theory, the difference should highlight and identify the non-casein components in the milk powder. In practice, this is not achieved because, due to sample particle scattering spectra contain independent linear trends which change from one sample to the next [7, 6, 4], and because some of the non-casein components in the milk powder have side chains which vibrate in a similar manner to the side chains of casein.

Because the scattering effect is essentially a linear function of the wavelength λ, it is removed by taking the second derivative of the data. In addition, the implicit utility of the second derivative is that it performs a resolution enhancement [2], as illustrated in Figure 4. This motivates and validates the explicit utility of derivative spectroscopy.

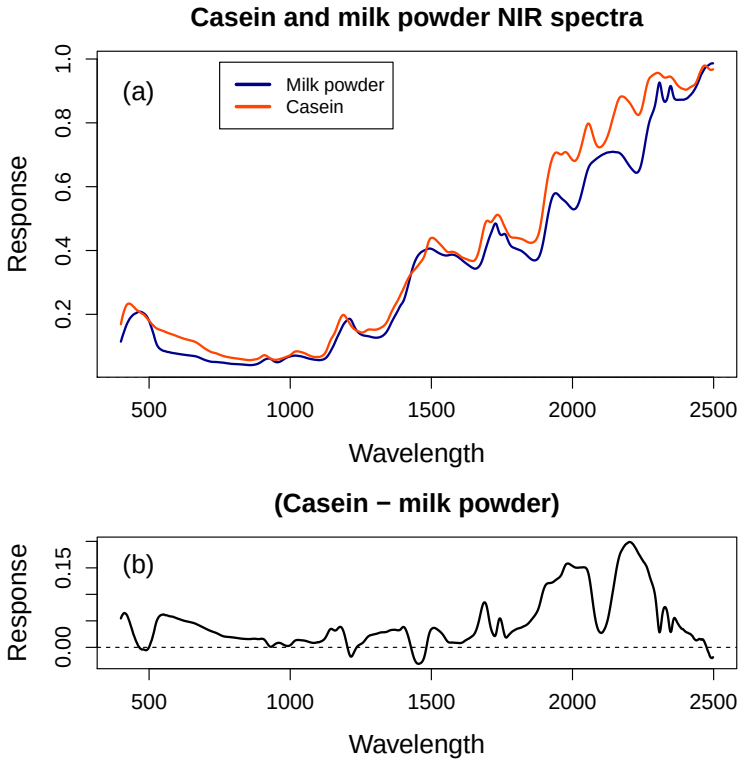

Figure 3: Comparison of a representative milk powder spectrum with that for casein: (top) typical milk power and casein spectra; (bottom) the difference between these two spectra.

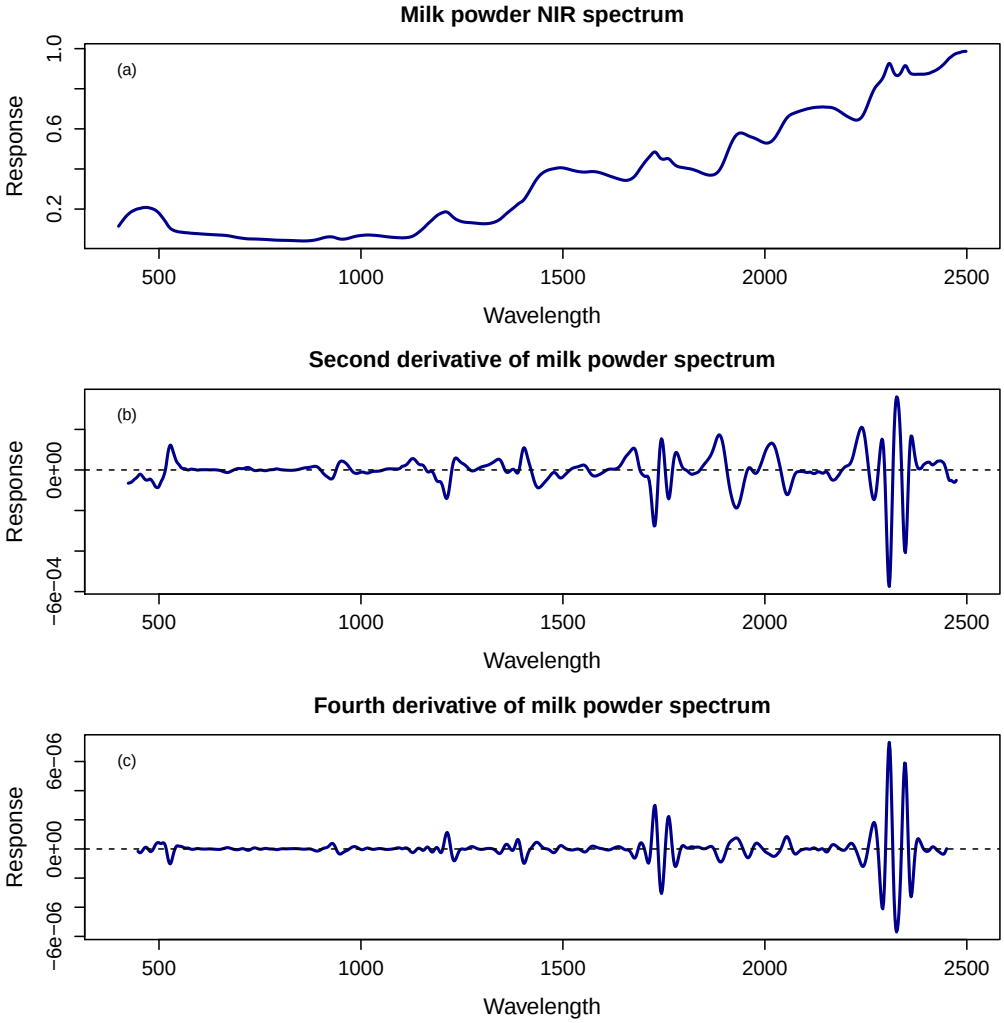

Figure 4: Resolution enhancement for the milk powder spectrum from the second and fourth derivatives.

The successful numerical differentiation of highly accurate data on a very fine grid (such as the NIR data being considered here) is possible through the use of moving average differentiators with an appropriately large footprint [2]. As explained by Anderssen and Hegland [2], it is the size of the footprint, as a function of the level of discretization determining the fineness of the grid, that performs the regularization which accommodates a high order of differentiation before instability becomes apparent. The utility of numerical differentiation to perform resolution enhancement has a long history and, historically, was given the name *derivative spectroscopy* [3].

3 Sparse regularization for spiked NIR data

Here, the focus is the spiking of samples of the same milk powder with different proportions p_j of casein, $j = 1, 2, \ldots, J$. In terms of the notation introduced in Section 2.1, the proportions of casein in the samples are, in ascending order:

$$a_1 = \alpha \text{ [unspiked milk powder]} < a_2 < \cdots < a_J = 1 \text{ [only casein]}.$$

For $j = 1, 2, \ldots, J$ the corresponding recorded spectra are represented as the row vectors

$$\mathbb{MIP}_{a_j}^{T} = \left[\mathbb{MIP}_{a_j}(\lambda_1), \mathbb{MIP}_{a_j}(\lambda_2), \ldots, \mathbb{MIP}_{a_j}(\lambda_K)\right],$$

where $\mathbb{MIP}_{a_j}(\lambda_k)$ denotes the milk powder spectra at wavelength λ_k when the proportion of casein is a_j. The above sparsity identification can equally well be applied to the spectrum of the unspiked milk powder (without added casein). The counterpart of Figure 2 then becomes that shown in Figure 5. A comparison of the plots in these two figures shows clearly how, even for different related situations, the nature of the sparsity changes.

The rectangular matrix array generated by row vectors $\mathbb{MIP}_{a_j}^{T}$ with $j =$

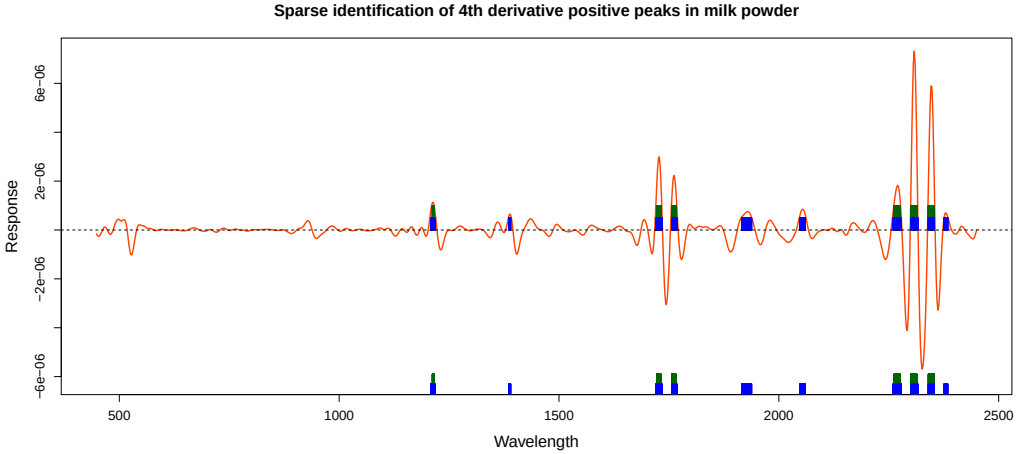

Figure 5: Plots of $\mathbb{S}^{(4)}_{\text{MP};\delta}(\lambda)$ for the milk powder (MP) for two different values of δ.

$1, 2, \ldots, J$, is

$$\mathbb{MIP} = \begin{bmatrix} \mathbb{MIP}_{a_1}(\lambda_1) & \mathbb{MIP}_{a_1}(\lambda_2) & \cdots & \mathbb{MIP}_{a_1}(\lambda_K) \\ \mathbb{MIP}_{a_2}(\lambda_1) & \mathbb{MIP}_{a_2}(\lambda_2) & \cdots & \mathbb{MIP}_{a_2}(\lambda_K) \\ \vdots & \vdots & \cdots & \vdots \\ \mathbb{MIP}_{a_J}(\lambda_1) & \mathbb{MIP}_{a_J}(\lambda_2) & \cdots & \mathbb{MIP}_{a_J}(\lambda_K) \end{bmatrix}. \tag{3}$$

The matrix of the fourth derivatives of these spectra is $\mathbb{MIP}^{(4)}$, with $\mathbb{MIP}^{(4)}_{a_j}(\lambda_k)$ denoting the values of the fourth derivatives of the milk powder spectra at wavelength λ_k when the proportional amount of casein is a_j. The opportunity that this set of spectra provides for identifying the appropriate intervals to be used as predictors of casein content is illustrated in Figure 6, where two localized wavelength regions of the spectra are highlighted. The top plot is indicative of strong confounding between the side chain vibrations of the casein and the non-casein components in the milk powder, whereas the bottom highlights minimal confounding. In the bottom plot, the heights (intensities) of peaks to the left of the 1200 nm wavelength correlate strongly

and positively with the levels of spiking, whereas for the peaks to the right the heights correlate inversely with the levels of spiking.

Figure 6 highlights how, for different levels of spiking, the fourth derivatives change as a function of the wavelength. In particular, there are two distinct situations, as described below.

1. For appropriate choices of δ there are wavelength intervals of $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ (such as in the wavelength region 1100–1300 nm) where the corresponding peaks of the $\mathbb{MP}^{(4)}_{a_j}$, $j = 1, 2, \ldots, J$, have very similar profiles, indicating that the associated casein side chain vibrations are independent of the side chain vibrations in the non-casein components of the milk powder.

2. For appropriate choices for δ, there are wavelength intervals of $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ (such as in the wavelength region 450–600 nm) where the corresponding peaks of the $\mathbb{MP}^{(4)}_{a_j}$, $j = 1, 2, \ldots, J$, have clearly different profiles, indicating that there are interactions of side chain vibrations between the casein and the non-casein component of the milk powder.

From a sparse regularization perspective, only the intervals associated with Situation 1 are appropriate as predictors of the casein content in milk powder. Since the intervals $\mathbb{S}^{(4)}_{\text{casein};\delta}(\lambda)$ for pure casein are known, the identification reduces to finding the subset of these intervals for which the peaks in $\mathbb{MP}^{(4)}$ have the same ordering as the proportions of casein $a_1 < a_2 < \cdots < a_J$. For the intervals associated with Situation 1, the heights of the peaks correlate closely with the proportional ordering. Because such invervals do not involve confounding of the molecular vibrations of the casein with the molecular vibrations of the non-casein components, this is an immediate consequence of the Beer–Lambert law. Even for Situation 2 the Beer–Lambert law still holds, but now the NIR response to molecular vibrations of non-casein components are added to the NIR responses of the casein component.

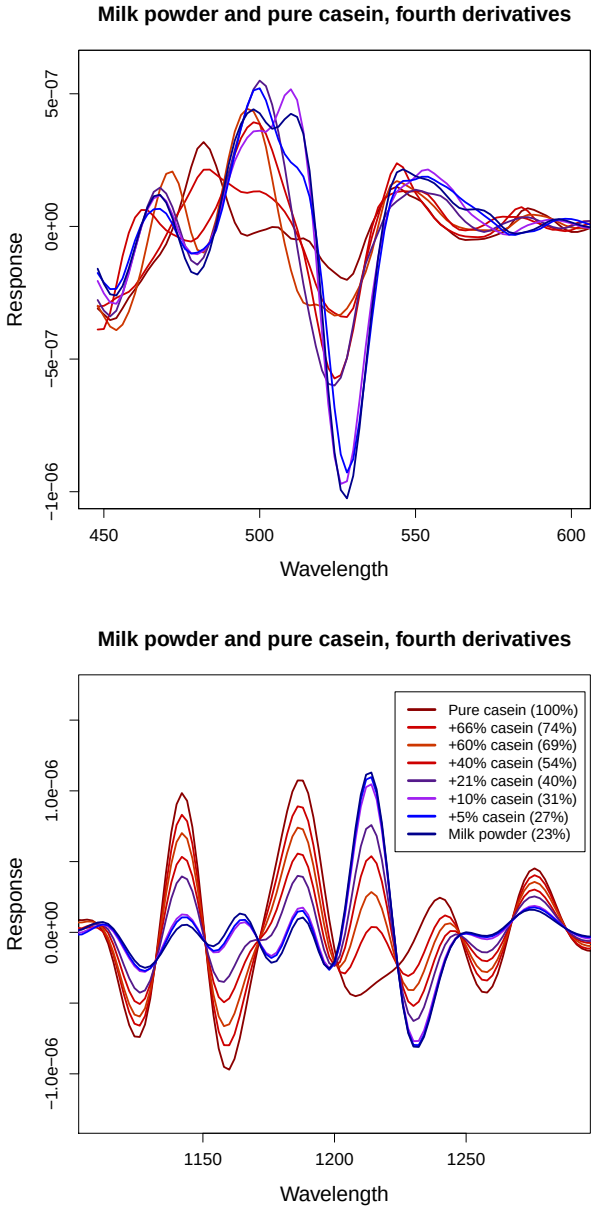

Figure 6: Plots of $\mathbb{MP}^{(4)}_{a_j}(\lambda_k)$, for $j = 1, 2, \ldots, J$, for two localized wavelength regions showing: (top) severe confounding; (bottom) minimal confounding.

3.1 A simple algorithm to test how fourth derivative values correlate with the levels of the spiking.

As already noted in Section 1, any correlation measure could be used to test how well the ordering of the fourth derivatives $\mathbb{MIP}^{(4)}_{a_j}$, $j = 1, 2, \ldots, J$, correlate with the ordering of the proportions of casein $a_1 < a_2 < \cdots < a_J$, as long as the data to which it is applied is suitably scaled.

An assessment of the extent to which the ordering of the fourth derivatives correlates with the level of spiking is perform using the spectral values $\mathbb{MIP}^{(4)}_{a_j}(\lambda_k)$, $j = 1, 2, \ldots, J$, for each λ_k. Its implementation involves working with the columns of the $J \times K$ rectangular matrix $\mathbb{MIP}^{(4)}$ defined in equation (3). The relative changes in the fourth derivative values are highlighted by removing the magnitude effect, which corresponds to mean centering each of the columns of $\mathbb{MIP}^{(4)}$. The mean centering (meanc) matrix of $\mathbb{MIP}^{(4)}$ is

$$\mathbb{MIP}^{(4)}_{\text{meanc}} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K] ,$$

with column vectors $\mathbf{s}_j$.

Because the same volume of milk powder is used for each spiking, the activity of non-casein components in the columns of $\mathbb{MIP}^{(4)}_{\text{meanc}}$ are removed along with the magnitude of the activity associated with casein. Let $\mathbf{a}^*$ denote the column vector obtained by mean centering the column vector $(a_1, a_2, \ldots, a_J)$. The simple test is to assess how well some multiple of an $\mathbf{s}_k$ correlates with $\mathbf{a}^*$; namely, evaluate

$$E_k = (\hat{\alpha}_k \mathbf{s}_k - \mathbf{a}^*)^T (\hat{\alpha}_k \mathbf{s}_k - \mathbf{a}^*) , \quad \hat{\alpha}_k = \frac{\mathbf{s}_k^T \mathbf{a}^*}{\mathbf{s}_k^T \mathbf{s}_k} . \tag{4}$$

The values of E_k, as a function of k, and hence λ_k, are plotted as a 'rug' along the bottom of Figure 7, with white and red denoting a good and a poor correlation, respectively; that is, minimum and maximum confounding. The pink shading denotes the extent of the confounding which occurs between the minimum and the maximum. The role of the rug is to give an easily

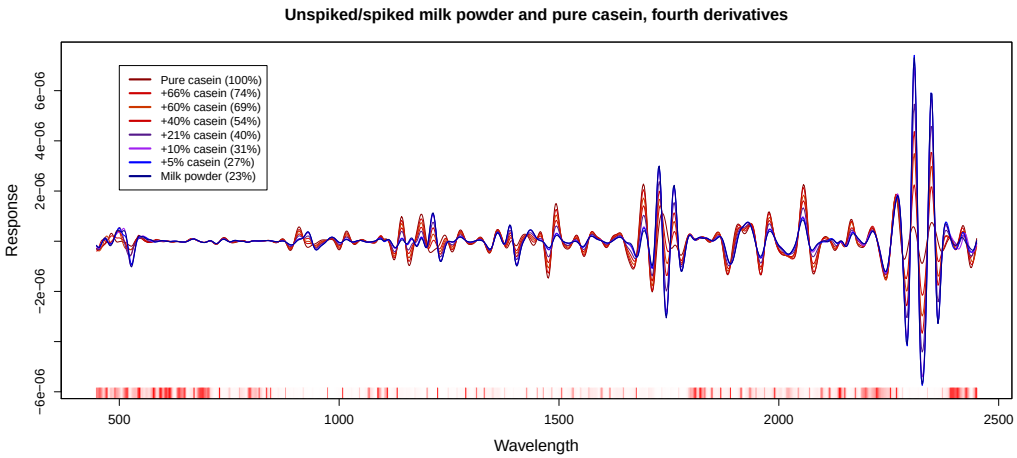

Figure 7: A plot of E_k as a function of k as a rug, compared with the fourth derivative spectrum of milk powder with different spiking.

accessible visual characterization of the level of confounding as a function of the wavelength λ_k .

4 Conclusion

Here, the goal is to give a *proof-of-concept* for the application of derivative spectroscopy as an explicit *sparse regularization* protocol. For this, the calibration data consists of the NIR spectra for milk powder spiked with known amounts of casein, while the property of interest is the proportional presence of the casein in the milk powder.

Acknowledgements The milk powder and casein spectra were supplied by Steve Holroyd of Fonterra in New Zealand. The reviewers' helpful advice resulted in a clearer explanation of the spiking and the explicit sparse

regularization.

References

[1] R. S. Anderssen, F. R. de Hoog, and I. J. Wesley. Information recovery from near infrared data. In W. McLean and A. J. Roberts, editors, *Proceedings of the 15th Biennial Computational Techniques and Applications Conference, CTAC-2010*, volume 52 of *ANZIAM J.*, pages C333–C348, July 2011. http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/3909. C797, C798

[2] R. S. Anderssen and M. Hegland. Derivative Spectroscopy – An enhanced role for numerical differentiation. *J. Integral Equat. Appl.*, 22(3):355–367, 2010. doi:10.1216/JIE-2010-22-3-355. C798, C801

[3] R. S. Anderssen, M. Hegland, and I. J. Wesley. Resolution enhancement for infrared spectroscopic data. In *MODSIM2011, 19th International Congress of Modelling and Simulation*, pages 371–377. Modelling and Simulation Society of Australian and New Zealand, 2011. http://www.mssanz.org.au/modsim2011/A4/anderssen3.pdf. C801

[4] R. S. Anderssen, B. G. Osborne, and I. J. Wesley. The application of localisation to near infrared calibration and prediction through partial least squares regression. *JNIRS*, 11(1):39–48, 2003. doi:10.1255/jnirs.352. C793, C798

[5] T. Naes, T. Isaksson, T. Fearn, and T. Davies. *A User-Friendly Guide to Multivariate Calibration and Classification*. NIR Publications, Chichester, UK, 2002. http://eprints.ucl.ac.uk/151973/. C797, C798

[6] B. G. Osborne. Near-infrared spectroscopy in food analysis. In *Encyclopedia of Analytic Chemistry*, pages 1–14. J. Wiley & Sons, Chichester, UK, 2006. doi:10.1002/9780470027318.a1018. C793, C797, C798

[7] B. G. Osborne, T. Fearn, and P. H. Hindle. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Longman Scientific & Technical, Harlow, UK, 1993. McGraw-Hill Series in Higher Mathematics. C790, C793, C798

[8] I. J. Wesley, B. G. Osborne, R. S. Anderssen, S. R. Delwiche, and R. A. Graybosch. Chemometric localization approach to NIR measurement of apparent amylose content of ground wheat. *Cereal Chem.*, 80(4):462–467, 2003. doi:10.1094/CCHEM.2003.80.4.462. C790

Author addresses

1. **R. S. Anderssen**, CSIRO Mathematics, Informatics and Statistics GPO Box 664, Canberra, ACT 2601, Australia.
mailto:Bob.Anderssen@csiro.au

2. **F. R. de Hoog**, CSIRO Mathematics, Informatics and Statistics GPO Box 664, Canberra, ACT 2601, Australia
mailto:Frank.deHoog@csiro.au

3. **I. J. Wesley**, Grain Growers Limited, PO Box 7, North Ryde, NSW 1670, Australia.

4. **A. B. Zwart**, CSIRO Mathematics, Informatics and Statistics GPO Box 664, Canberra, ACT 2601, Australia
mailto:Alec.Zwart@csiro.au