Journal of Computational Physics 295 (2015) 307–321

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

On computing first and second order derivative spectra

Indrajit G. Roy *

Spaceage Geoconsulting, PO Box 6154, Conder, ACT 2906, Australia

ARTICLE INFO

Article history:
Received 26 October 2014
Received in revised form 11 February 2015
Accepted 13 April 2015
Available online 21 April 2015

Keywords:
Robust derivative
Balancing principle
Regularization parameter
Spectral data

ABSTRACT

Enhancing resolution in spectral response and an ability to differentiate spectral mixing in delineating the *endmembers* from the spectral response are central to the spectral data analysis. First and higher order derivatives analysis of absorbance and reflectance spectral data is commonly used techniques in differentiating the spectral mixing. But high sensitivity of derivative to the noise in data is a major problem in the robust estimation of derivative of spectral data. An algorithm of robust estimation of first and second order derivative spectra from evenly spaced noisy normal spectral data is proposed. The algorithm is formalized in the framework of an inverse problem, where based on the fundamental theorem of calculus a matrix equation is formed using a Volterra type integral equation of first kind. A regularization technique, where the balancing principle is used in selecting *a posteriori* optimal regularization parameter is designed to solve the inverse problem for robust estimation of first order derivative spectra. The higher order derivative spectra are obtained while using the algorithm in sequel. The algorithm is tested successfully with synthetically generated spectral data contaminated with additive white Gaussian noise, and also with real absorbance and reflectance spectral data for fresh and sea water respectively.

1. Introduction

Enhanced resolution without bringing noise artifact in spectroscopic technique is an essential step in spectroscopic analysis of a complex substance. This is because spectral contributions by the components within a substance under investigation often overlap in the measured spectral band making it difficult to identify and separate their contributions individually. In fact, such issue has been well recognized over many years and was attempted to resolve as early as middle of twentieth century [1–4]. There is a record of various approaches over more than six decades to enhance resolution, chiefly either through Fourier based self deconvolution [5,6] or by incorporating derivative of the spectra; in that derivative method has taken a significant place in today's spectroscopy technique.

Initial attempt in incorporating derivative on the measured spectra were restricted with the hardware by designing either appropriate optical set up, such as scanning spectrophotometer of various make or through appropriate electronic circuitry. A detailed discussion on hardware based method is available in [7]. However, derivative spectroscopy through hardware implementation has limitations, such as 1) restricted limit on the order of derivative in the measured spectra, 2) increasing difficulty in noise control with the increase in order of derivative and 3) a shift in amplitude and wavelength of the derivatives due to variation in the scan speed for optical scan based technique.

* Tel.: +61 2 6294 2510.
E-mail address: spaceage.geocon@gmail.com.

http://dx.doi.org/10.1016/j.jcp.2015.04.015
0021-9991/

Computational approach, on the other hand, is highly flexible. The noise reduction in the measurement becomes greatly simplified. With today's technology of high speed computing and flexibility, one can have, almost a real time, derivative spectra from the measurement of normal spectra. Despite such advantage a robust, reliable computationally efficient method of derivative estimation, especially of large and noisy dataset remains elusive. Even today the difference method of computing derivative is most popular and is widely used technique, as it is found to be a simplest means of computing derivative from evenly spaced tabulated data using Excel™ spreadsheet. But such simplicity in computing comes with a price of compromising fidelity by increasing the effect of noise in data. More densely (i.e. reduction in data interval) the spectral data is measured the effect of noise in data becomes more pronounced in computed derivative using a difference scheme. Unreliability even grows further with the increase in noise magnitude and the order of derivatives. With a common belief, the reduction in reliability in estimated derivative via difference method is somewhat counter intuitive, as from the first principle of differential calculus computational accuracy should increase with decreasing data interval as long as the denominator remains non-zero or does not attain the underflow limit of a computer.

The measured data, however, contain random noise in general, and such noise has dominant effect on the high frequency part of the measurements. Since, a difference scheme which is essentially based on polynomial interpolation acquires the attribute of low pass filtering. Larger the data spacing higher is the filtering effect. An optimal data spacing which is most sought after is, however, hard to achieve. It was demonstrated in [8] that for the noise level δ in data a minimal error of magnitude $2\sqrt{\delta}$ can be attained for a second order central difference scheme, if the denominator δx ($=l\Delta$, with l an integer) approximates closely to δ. It then turns out that one needs to discard some data samples from a closely sampled dataset to gain the desired accuracy within a predetermined noise level. Such strategy is, however, not encouraging as A) one is often left with a little option of pre-selecting the data spacing (or the sampling rate) and B) the choice of larger δx leads to a lose of information from a highly sampled dataset.

Other popular approach, mostly embraced by geophysics community, is Fourier method in computing derivative of evenly distributed data. The motivation in adopting the technique is chiefly due to computational efficiency while handling a large data through Fast Fourier transform (FFT) algorithm, although additional steps involved, such as trend removal, appropriate windowing and zero padding makes the method somewhat tedious. Despite all, the most worrying aspect is that the method is sensitive to additive white Gaussian noise (AWGN). Higher the order of derivative, more susceptible it becomes with the noise in data. Therefore, derivative computation without noise removal using an appropriate filter cannot be an option. The requirement of filtering to tackle the noise issue (especially with AWGN) makes Savitzky–Golay (SG) moving polynomial method [9,10] an alternative approach for derivative spectroscopy. The performance of SG method, however, depends on the size of the window and the degree of polynomial chosen. Larger window would cause a high level of smoothness; conversely significantly small window size may not be able to attain the desired level of noise reduction. Another popular approach is polynomial fitting method [11,12] where the measured dataset is fitted with an appropriate polynomial globally and then derivative is computed on the polynomial. The limitation of such method is to find an optimal order of polynomial. Often, selecting a higher order polynomial brings in computational instability due to Runge effect [12].

On the other hand, the recent trend is towards regularization technique which would provide the much desired stable solution in derivative estimation while remaining not too expensive computationally. There are various strategies in estimating derivative using regularization technique. For example, in [13] a regularized spline technique was used in estimating derivative of discrete unevenly spaced dataset. Other approaches are: 1) use of integral equation, such as Fredholm integral equation of first kind [35], Volterra integral equation [14,15] together with Tikhonov regularization technique, 2) use of integral equation and total variation regularization in estimating derivative from noisy data set [16], 3) finite difference technique with optimally chosen step size [36], use of dynamical system method [37].

We propose a novel technique which is sufficiently robust and computationally efficient in estimating derivative of noisy spectral data. We formalize the derivative estimation problem as an inverse problem. Using the fundamental theorem of calculus a matrix operator equation is designed from a Volterra type integral equation of first kind. The inverse problem is solved in the framework of a constrained optimization via regularization where an optimal value of regularization parameter is determined using '*balancing principle*'. We demonstrate robustness and efficiency of the proposed algorithm using numerical experiments on synthetically generated data which are contaminated with AWGN. We then use it in analyzing absorption and reflectance spectra within the visible range to study phytoplankton and other dissolved organic matters in pure water [32] and sea water [33] respectively.

The paper is organized as followed: In Section 2 we give a brief discussion on the theoretical background for our proposed algorithm. In Section 3 we discuss about the results obtained from the numerical experiments conducted on synthetically generated spectral data contaminated with AWGN. We also demonstrate with illustrations how conventional Fourier transform based method and central difference scheme in derivative computation fail when data contain AWGN even in a mild level. We then applied our method in estimating first and second order derivatives of absorbance and reflectance spectral data within the visible range to study the phytoplankton and other dissolve organic matters in both fresh water lake and sea water to demonstrate the applicability of the algorithm. The paper then concludes with a brief conclusion.

2. Formulation

Suppose that $f(x)$ is a continuously differentiable, smooth function within an interval $[a, b]$ for which $u(x)$, a continuous function, is the first derivative. Actually, we deal with evenly sampled discrete 1D data, such as $\{f_1, f_2, \ldots, f_N\}$ defined over