Nuclear Instruments and Methods in Physics Research A 691 (2012) 1–4

Contents lists available at SciVerse ScienceDirect

Nuclear Instruments and Methods in Physics Research A

journal homepage: www.elsevier.com/locate/nima

Continuum determination in spectroscopic data by means of topological concepts and Fourier filtering

Nikolaos Kourkoumelis *

Department of Medical Physics, Medical School, University of Ioannina, Ioannina 45110, Greece

ARTICLE INFO

Article history:
Received 21 March 2012
Received in revised form
26 June 2012
Accepted 26 June 2012
Available online 10 July 2012

Keywords:
Continuum
Spectroscopy
Background estimation
Convex hull

ABSTRACT

A novel implementation of a convex hull minimization algorithm for the determination of the continuum in x-ray and γ-ray spectroscopic data is presented in this paper. The method is semi-automatic, performed in three successive steps and requires user intervention regarding the values of two parameters. The first controls the low pass filtering and the second the reduction of the estimated points which construct the background. On the other hand, it does not rely on an explicit mathematical model of the background or the signal (FWHM, lineshapes) and keeps the original spectrum data unaltered.

1. Introduction

A typical, albeit critical, procedure in spectroscopic data analysis is the background elimination of the experimental signal. Continuum removal is usually the initial preprocessing step in quantitative spectrum analysis and significantly affects all subsequent study. Theoretically, spectra are composed of peaks superimposed upon a slow varying background originating from counts fluctuations in the receiving channel. Due to a number of effects however, from counting statistics to other specific physical phenomena, this condition is not met and background shape is far from linear.

Numerous effective methods have been proposed for x-ray and γ-ray spectrometry. These include: peak stripping [1] where a comparison is made between the intensity of neighboring channels, the Statistics Sensitive Nonlinear Iterative Peak-Clipping Algorithm (SNIP) [2–4], being an improvement of the previous method and very effective in the complete removal of background in the vicinity of the peaks, second and third order spline interpolation [5] via estimation of the local minima on either side of the peaks, Fourier transform filtering [6], zero-area digital filtering [7,8] where baseline has the value of zero in the absence of peaks and its shape is varied between upper and lower boundaries similar to a smoothed second derivative spectrum, polynomial fitting [9] or its optimized extension of orthogonal polynomial decomposition [10] where polynomial fitting is performed and the weights of the least-squares fit are iteratively adjusted to include only channels belonging to the continuum, iterative methods [11], where in each iterative step the peak height is decreased while one point in each side of the smoothed spectrum is added until the slope variance reaches a certain value, Bayesian probability algorithms [12], spectra statistical properties (counts fluctuations) [13] and continuum estimation based on mathematical morphology [14] which is actually a filtering technique based on dilation and erosion of geometric structures (i.e. the signal). The vast implementation methods for background determination are a proof that there is not an optimum one. Parameters like signal-to-noise ratio (S/N), peak lineshape asymmetry and full width at half maximum (FWHM), inherent characteristics like Compton edges for gamma rays, experimental setup etc. may significantly affect the background estimation.

In this study, we present a generic geometric approach the for background removal of energy dispersive x-ray fluorescence and γ-ray spectroscopic data based on Fourier filtering and the application of topological concepts and algorithms.

2. Method

The method proposed in this study consists of three steps. The first one, is a common but very basic assumption in all filtering methods stating that the whole signal, S, is a composition of a low frequency component ($B(E)$, background) which propagates very slowly and independently the true information ($P(E)$, peaks). The latter represent the high frequency component, so that $S(E)=P(E)+B(E)$. Fourier transformation is, by definition, a

* Tel.: +30 2651007594.
E-mail address: nkourkou@cc.uoi.gr

0168-9002/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.nima.2012.06.061

technique to determine the frequency content of a time-dependent signal. Using the Discrete Fourier Transform (DFT) we may sample the signal to a finite number of equidistant data points so that time is denoted by energy channels (E). The DFT of $S(E)$ can be written as

$$F(\omega)=\sum_{E=0}^{N-1} S(E)\mathrm{e}^{-i\omega E}=P(\omega)+B(\omega) \tag{1}$$

where ω is the frequency component per channel and N the number of channels.

This low-pass filter attenuates high frequencies and retains low frequencies practically unchanged. The application of the inverse Fourier transform gives a first approximation of the background component as the convolution of the original signal:

$$\mathrm{S}'(\mathrm{E})=\frac{1}{N}\sum_{\omega=0}^{N-1} \mathrm{F}(\omega)\mathrm{e}^{i\omega E} \tag{2}$$

where $S'(E)$ is a smoothed form of the original spectrum due to the application of a cutoff frequency (ω_c) value defined by the user. The ω_c value increases as the spectrum becomes more complicated since the $S'(E)$ has to capture the high degree of peaks variation. Conversely, its value is low for spectra with well-defined peaks. For noisy data the optimal value for ω_c is dictated by the S/N ratio since by definition it has to suppress high frequencies and smooth the lower ones.

The second step is to decompose the signal into convex sets [15]. The theory underlying this process has already been described by Brunetti et al. [16] who named this approach "projection on convex sets". The method we propose is in essence analogous but has different implementation and requires no fitting iterations; it is purely geometrical and driven solely by the topological descriptors of each spectrum. This step starts by constructing a one-dimensional table consisting of "0" and "1" which refer to the positive and negative slope of the filtered signal. Hence, via a pattern search of "01" and "10", we divide the whole spectrum into regions of peaks and valleys to which we apply a convex hull minimization routine. Convex hull is a classical geometrical problem of finding the minimal convex polygon which contains all points of a set. In other words, given a finite set of points $P=\{p_1, p_2,\dots, p_n\}$ as a list of ordered pairs of cartesian coordinates $\{(x_1, y_1),(x_2, y_2),\dots,(x_n, y_n)\}$, (channel vs. counts), convex hull of P is the smallest convex set C such that $P \in C$. It can be applied to any set of positive points propagating monotonically such as spectroscopic data and is typically given by the expression:

$$C=\left\{\sum_{n=1}^{N} \lambda_n p_n : \lambda_n \geq 0 \text{ and } \sum_{n=1}^{N} \lambda_n = 1\right\} \tag{3}$$

For the calculated convex regions we construct a linear median of the local data dispersion which splits the polygon into two parts also by taking into account the varied local slope. Since we have already divided the spectrum into convex sets of peaks and valleys, the median effectively allocates the data to background and peaks. The problem with this algorithmic implementation though, is to ensure continuation of the background between the adjacent convex regions. The third step introduces a user adjusted variable multiplied by the mean derivative of all data points (the median of all the slopes) of the signal and defines the number of the linking points which must be included in the estimated background. This is actually a reduction of points participating in the final baseline array based on the initial assumption that the background has quite low mean derivative compared to the high frequency part of the signal. The final determination of the continuum is summarized as follows:

$$S'^{(E)}=\{P \in C_s \text{ and } P<j\} \tag{4}$$

where C_s is the part of the resulting convex set corresponding to background and j the slope average of the signal. An outcome resulting in a smoother spline may be possible with "overlap and add" algorithms [17] and represents a future development.

3. Results

To assess the efficiency of the method we have evaluated three different synthesized spectra with diverse background shapes. First, we have used a set of four Lorentzian peaks of unit height but of gradually diminishing half-widths to illustrate distortion effects. To make data more realistic, Gaussian noise (s_n=0.04) has been added. Fig. 1(a)–(c) depicts the three steps required for the determination of the background.

The two adjustable parameters play a significant role for the correct background approximation. If the ω_c has a very high value the number of convex regions increases and part of the peaks are considered as background. Similarly, when j adopts very high values, Eq. (4) indicates that steep signals "contaminate" the points considered as background. The comparison of Fig. 1(b) and (c) clearly shows the importance of j, especially in noisy spectra.

Fig. 2 shows an energy-dispersive X-ray fluorescence (EDXRF) spectrum with a ^{238}Pu excitation source and a Si-PIN detector adopted from Ref. [6] and the background calculated from it. The spectrum is far more complicated and therefore the two adjustable parameters have higher values compared to the previous case. Strong peaks are easily distinguished and information from weak peaks is efficiently retained.

A synthesized γ-ray spectrum as described in Ref. [11] is depicted in Fig. 3. Three Gaussian lineshapes represent the spectroscopic data with superimposed random noise and Compton continuum. As expected, the peak width measured in channels of peaks in the previous case of x-ray spectrum is bigger than in γ-ray one and the latter exhibits some significant asymmetric background features due to distinct Compton edges. Thus, the two user defined variables must have even higher values than before in order to follow the challenging Compton scattering.

It is evident from all figures that the method proposed here successfully predicts the background signal from diverse spectroscopic data related to x-ray and γ-ray. It does not require any previous knowledge about specific experimental details or inherent parameters like asymmetric lineshapes and FWHMs. In addition, knowledge of the underlying mathematical model is not necessary. A disadvantage however, might be the reproducibility of the estimation among users since it requires the fine-tuning of two adjustable parameters making the whole process semi-automatic.

4. Conclusions

A novel implementation of an accurate and fast technique aiming to determine the continuum independent of manifestations of physical phenomena is presented in this study. The method requires distinct steps based on Fourier filtering and convex hull minimization. It is computationally inexpensive and it does not require fitting convergence or successive iterations. Results showed comparable results with other background elimination methods found in the literature. The accuracy is confirmed by synthetic and experimental spectra. A significant advantage is that it does not rely on an explicit mathematical model of the continuum since, quite often, background modeling