Nuclear Instruments and Methods in Physics Research A 600 (2009) 478–487

Contents lists available at ScienceDirect

Nuclear Instruments and Methods in Physics Research A

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

An algorithm for determination of peak regions and baseline elimination in spectroscopic data

Miroslav Morháč *

Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 845 11 Bratislava, Slovakia

ARTICLE INFO

Article history:
Received 17 September 2008
Received in revised form
20 November 2008
Accepted 30 November 2008
Available online 7 December 2008

Keywords:
Peak regions
Smoothed first derivatives
Convolution
Sensitive Nonlinear Iterative Peak (SNIP)
clipping algorithm
Clipping window
Background estimation

ABSTRACT

In the paper we propose a new algorithm for the determination of peaks containing regions and their separation from peak-free regions. Further based on this algorithm we propose a new background elimination algorithm which allows more accurate estimate of the background beneath the peaks than the algorithms known so far. The algorithm is based on a clipping operation with the window adjustable automatically to the widths of identified peak regions. The illustrative examples presented in the paper prove in favor of the proposed algorithms.

1. Introduction

One of the basic problems in the analysis of spectroscopic data is the separation of useful information contained in peaks from the useless information (background, noise). Baseline removal, as the first preprocessing step of spectra, critically influences subsequent analysis steps. The accuracy and reliability of the background analysis depend critically on the treatment in order to resolve strong peak overlaps, to account for continuum background contributions, and to distinguish artifacts to the responses of some detector types. A background approximation must pay particular attention to the reliable estimation of the background continuum under low-statistics peaks, both isolated and in close proximity to strong peaks. In order to process data from numerous analyses efficiently and reproducibly, the background approximation must be as much as possible free of user-adjustable parameters [1].

Background estimation is an omnipresent problem for quantitative spectral analysis. Numerous techniques for background estimation were designed with the goal for the efficient performance for specific problems [2]. The approaches to the background treatment cover Fourier transform techniques [3], digital filters, iterative line stripping [4], peak clipping [5] and the treatment of the first and the second derivative spectra [6]. In Ref. [7] the authors propose an algorithm based on Bayesian probability theory and in Ref. [8] the investigation of selected baseline removal techniques has been carried out.

In the works [9–12] the peak-free regions had to be identified in order to construct the background beneath peaks. Once background regions have been located an approximation to the shape of the spectral background must be found, both in background and in peak containing regions of the spectrum. Then to estimate the shape of the baseline one can employ the methods based on polynomial fits [13]. To identify peak-free regions of γ-ray spectra in Ref. [14] a zero-area digital filter is proposed.

A very efficient method of background estimation, based on a Statistics-sensitive Nonlinear Iterative Peak-clipping algorithm (SNIP), has been developed in Ref. [1]. It is considered as the most efficient method. Unlike to the previously mentioned methods, the SNIP method implicitly determines peak regions and peak-free regions. In Ref. [15] we extended the SNIP method for multi-dimensional spectra. In multidimensional spectra, the algorithm must be able to recognize not only continuous background but also to include all the combinations of coincidences of the background in some dimensions and the peaks in the other ones. In Ref. [16] we proposed several improvements of the SNIP algorithm. Further we have derived a set of modifications of the algorithm that allow estimation of specific shapes of background and ridges as well.

The SNIP algorithm is a multi-pass clipping loop which replaces each channel value $y(i)$ with the smaller of the two values $y(i)$ and the mean $[y(i-j)+y(i+j)]/2$, where $j \in \langle 1, m \rangle$ and m is

*Tel.: +421 2 5941 0536; fax: +421 2 5477 6085.
E-mail address: Miroslav.Morhac@savba.sk

0168-9002/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.nima.2008.11.132

a given free parameter [15]. To propose a suitable m parameter one can use

$$2m + 1 = w \tag{1}$$

where w is the real width of an object (peak, doublet, multiplet) that should be preserved. Choosing m too big can cause the remaining background residua under peaks. On the contrary choosing it too small can cause that the baseline underruns the real peaks and thus after subtraction decreases their net areas.

Therefore, the aim of the paper is to propose:

- an efficient algorithm for the determination of peak regions and
- to define the algorithm of baseline determination with adjustable clipping window along the spectrum according to peak regions.

2. Estimation of widths of peaks regions

There are many different algorithms and approaches to detect peaks in data. The main problem lies in distinguishing true peaks from statistical fluctuations, Compton edges and other undesired spectrum features. The estimation of widths of peaks regions is tightly connected with the localization and identification of peaks in the spectrum. It is beyond the scope of this paper to give a complete survey of all existing peak searching algorithms. However, a great majority of them is based on convolution operations. The use of the convolution method to locate the peaks is a well-established approach and is utilized in many of the available algorithms. It removes statistical fluctuations and is insensitive to linear background under Gaussians [17].

Without loss of generality let us assume that a peak is perfectly Gaussian-shaped given by

$$G(x) = A \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) \tag{2}$$

where x is the channel number and A, a and σ are peak intensity, peak position and standard deviation determining the peak width, respectively. Slavic in Refs. [18,19] proposed a so called Gaussian product function that is derived from four successive points

$$P_1(x) = [G(x)G(x-1)]/[G(x-2)G(x+1)] = \exp(2/\sigma^2). \tag{3}$$

One can observe that this function depends only on σ. However, it has to be calculated for all possible channel values x. When looking at the function, obviously for peak-free regions $\sigma \to \infty$ and consequently $p_1(x)$=1. Otherwise, in peaks regions $p_1(x)$=$const > 1$. This condition can be utilized for the identification of the peaks and together with Eq. (3) also for the determination of peak width and thus the peak regions. In our study, to make the function symmetrical around the peak position, we slightly modified Eq. (3) to

$$P_1(x) = G(x)^2/[G(x-1)G(x+1)] = \exp(1/\sigma^2). \tag{4}$$

To increase the sensitivity of the Gaussian product function, according to Ref. [18], we define

$$P_m(x) = G(x)^2/[G(x-m)G(x+m)] = \exp(m^2/\sigma^2) \tag{5}$$

where m represents the order of the Gaussian product function.

Let us now study the proposed algorithm by giving some illustrative examples. In Fig. 1a one can see an ideal Gaussian (without noise and background). In Fig. 1b we present Gaussian product functions defined in Eq. (5) for m=1, 2, and 5, respectively. Apparently, the functions have a constant value $P_m(x_0)$ inside of the peak and the value equal to 1 outside of it, where x_0 is the position of the peak. Then using Eq. (5) we can express

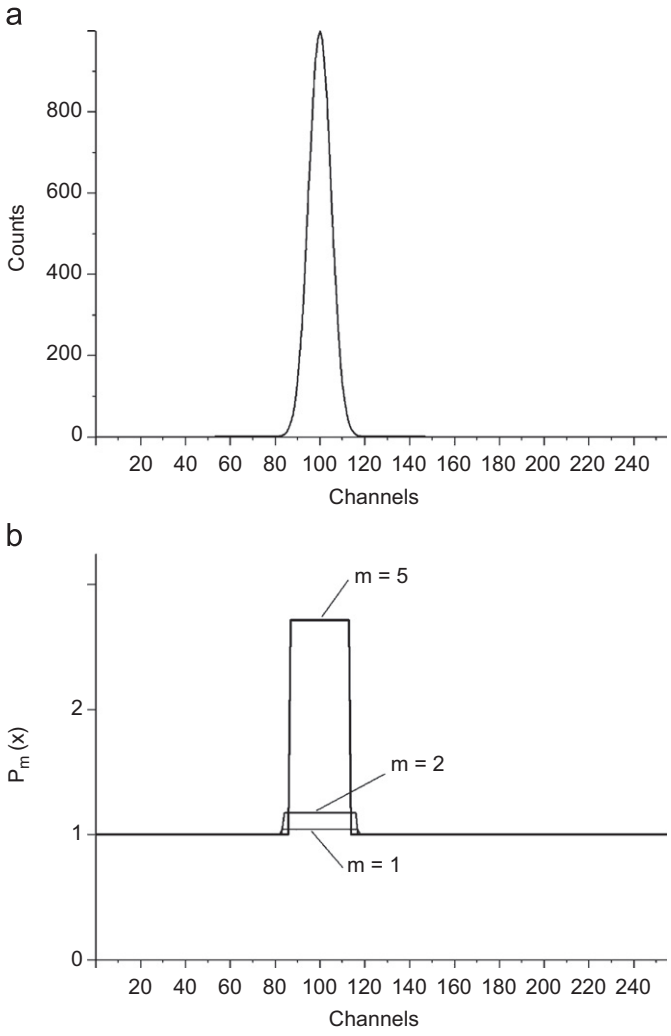

Fig. 1. An example of ideal Gaussian (σ=5) (a) and Gaussian product functions for m=1,2, and 5, respectively, (b).

$$\sigma = \frac{m}{\sqrt{\log P_m(x_0)}}. \tag{6}$$

Using the read out value $P_m(x_0)$ and Eq. (5) one can calculate with rather good precision σ of the peak.

For the sake of illustration in Fig. 2 we show the same example with additive noise with the amplitude of 10% of the maximum value of the data. The level of noise in the Gaussian product functions is outside of the peak magnified to such an extent that it is bigger than inside of the peak and can hardly be used for its identification.

Further, with the aim to damp the oscillations in the calculated Gaussian product function let us modify Eq. (5) to

$$P_m(x) = (G(x)+d)^2/[(G(x-m)+d)(G(x+m)+d)] \tag{7}$$

where d is a damping constant. In Fig. 3 we introduce an example of the damped Gaussian product functions of the order m=5 for the data with the noise equal to 1 and 10% of the maximum value, respectively (Fig. 2a). The damping constant d in this example was chosen to be the maximum value of the spectrum. One can observe a substantial decrease of the oscillations that allows identification of the peak. Nevertheless, these data cannot be used

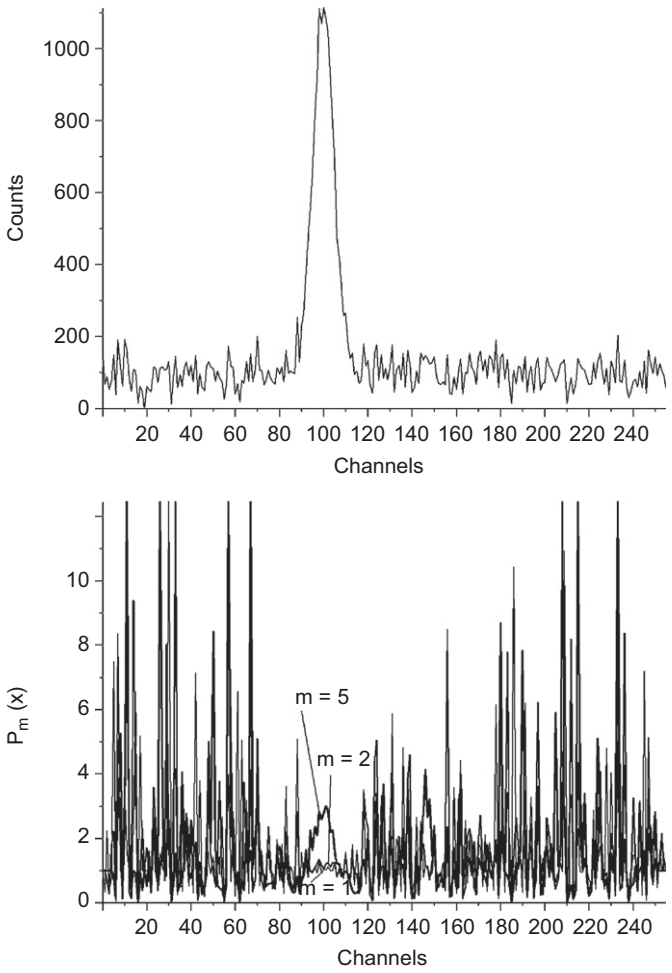

Fig. 2. An example of ideal Gaussian (σ=5), background and noise with the amplitude of 10% of maximum value in the data (a) and Gaussian product functions for m=1,2, and 5, respectively, (b).

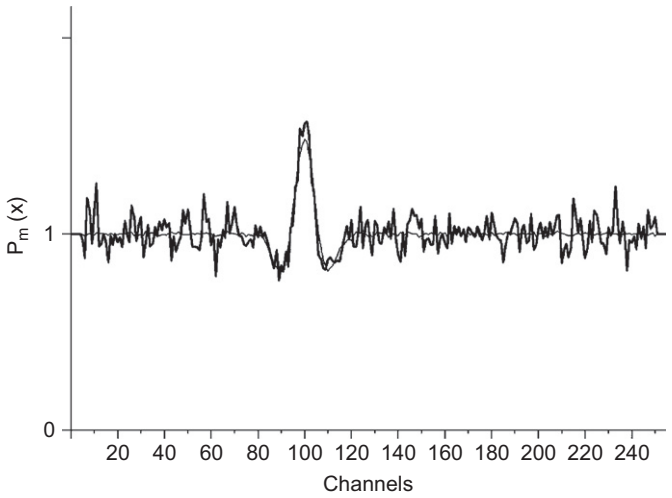

Fig. 3. Damped Gaussian product functions for m=5 for the data with noise with the amplitude of 1% (thin line) and 10% (thick line) of maximum value, respectively.

for the determination of σ of the peak as the right side of Eq. (5) is not more valid.

In Fig. 4a we present a synthetic spectrum consisting of 10 peaks with σ ranging from 1–10, with considerable level of

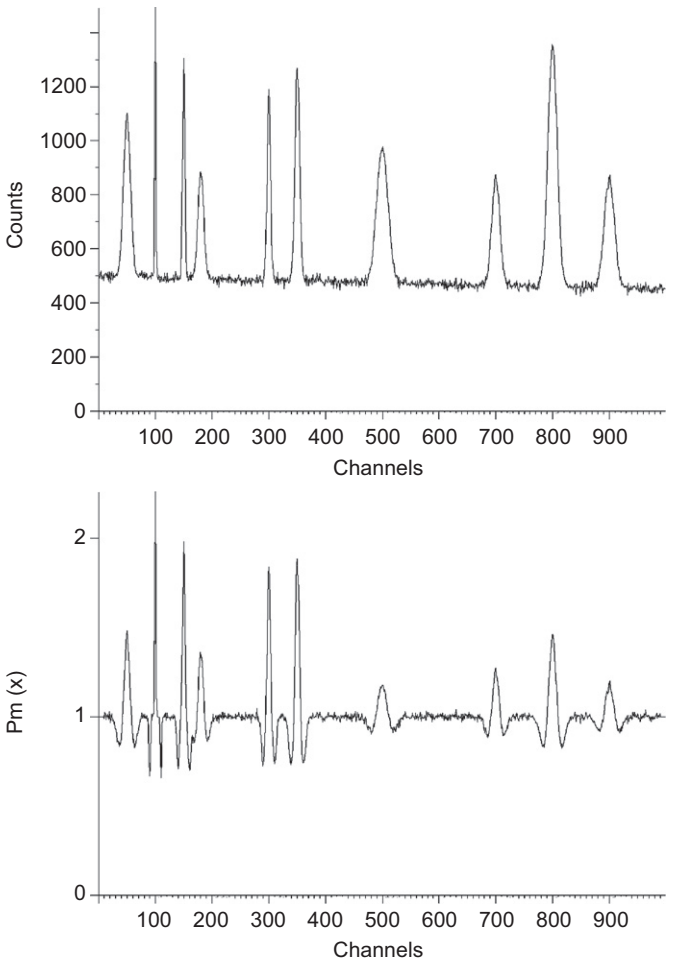

Fig. 4. A synthetic spectrum consisting of 10 peaks with σ from the range $\in \langle 1, 10\rangle$ (a) and its damped Gaussian product functions for m=10 (b).

Gaussian noise (with amplitude of 5% of maximum value) and linear background (see Table 1).

In Fig. 4b one can see its corresponding Gaussian product function of the order m=10. Apparently this data are smooth enough and outline clearly the positions of peaks in the original spectrum. Again due to reasons given above they cannot be utilized for the determination of σ and thus peaks intervals.

Sophisticated methods to locate the peaks and the intervals where they lie are based on the use of the convolution methods. This is a well established approach and is utilized in many codes. In peak searching algorithms usually the following two characteristics of a local maximum of continuous function are used [20]:

(1) Near the maximum, the curve is convex and so its second derivative becomes negative, having a minimum value around the maximum of the peak.
(2) While passing a peak position the first derivative changes sign.

It is assumed that peaks can be described by Gaussian functions (2) and that the background may be approximated by a linear function within short intervals. Thus in such an interval, the spectrum as a function of channel number is

$$y(x) = G(x) + B + Cx \tag{8}$$

where B, C are constants describing the background. The second derivative $y''(x)$ becomes independent of the background and it vanishes for any interval in which there is no peak.

Table 1
Positions and σ of peaks in the synthetic spectrum in Fig. 4a.

Peak #	Generated peaks (position/σ)
1	50/6
2	100/1
3	150/2
4	180/5
5	300/3
6	350/4
7	500/10
8	700/7
9	800/8
10	900/9

In Refs. [21,22] the authors propose a peak-search method based on a two-pass convolution of the spectrum with the first derivative of the Gaussian. Besides of the identification of peaks the method can be utilized for the determination of peaks intervals and peaks areas. In Ref. [23] the authors suggest an approach to the peak search algorithms in which the optimal convolution functions are used to extract the peak area, position and width from the convolved spectrum. The approach keeps the statistical errors down to the optimal level for all the relevant quantities. The method in Ref. [21] is based on convolution of the spectrum with the first derivative of an unnormalized Gaussian function (filter)

$$f'(t) = \left[\exp\left(-\frac{(t-x)^2}{2\delta^2}\right)\right]' = -\frac{t-x}{\delta^2}\exp\left(-\frac{(t-x)^2}{2\delta^2}\right) \tag{9}$$

where δ is the width of the filter. The convolution with the Gaussian (2) yields

$$c_1(x) = \int_{-\infty}^{\infty} G(t)f'(t)\,dt = \frac{\sqrt{2\pi}(x-a)\sigma\delta}{(\sigma^2+\delta^2)^{3/2}}\exp\left(-\frac{(x-a)^2}{2(\sigma^2+\delta^2)}\right). \tag{10}$$

By repeating the operation of convolution of the filter (9) with Eq. (10) we get

$$c_2(x) = \frac{2\pi\delta^2\sigma((x-a)^2-\sigma^2-2\delta^2)}{(\sigma^2+2\delta^2)^{5/2}}\exp\left(-\frac{(x-a)^2}{2(\sigma^2+2\delta^2)}\right). \tag{11}$$

These two convolutions filter out the background and smooth the data by removing statistical disturbances. Function (11) has a negative minimum value at the position of the peak. Using the crossing-over points of Eq. (11) with zero value one can determine

$$\sigma = \sqrt{(x_0-a)^2-2\delta^2} \tag{12}$$

where a is the position of the peak and x_0 is the appropriate zero crossing-over point.

Let us now illustrate the efficiency of the above suggested algorithm. In Fig. 5a we show the synthetic spectrum from Fig. 4a twice convolved with the first derivative of the non-normalized Gaussian filter (9) with δ=5. The negative minima in the data distinctly determine the positions of peaks a. Employing Eq. (12) and the crossing-over points with zero value x_0 one can determine also the left and right edge σ_l and σ_r of appropriate peaks and taking $\langle a-3\sigma_l, a+3\sigma_r\rangle$ also their intervals. In Fig. 5b we show again the same spectrum from Fig. 4a with found peaks denoted by markers and peaks regions outlined with thick lines. When looking at Fig. 5b one can immediately observe that due to the proximity of peaks 3 and 4 at the positions $\approx$150 and $\approx$180, there are errors in the estimation of the left and the right edge, respectively.

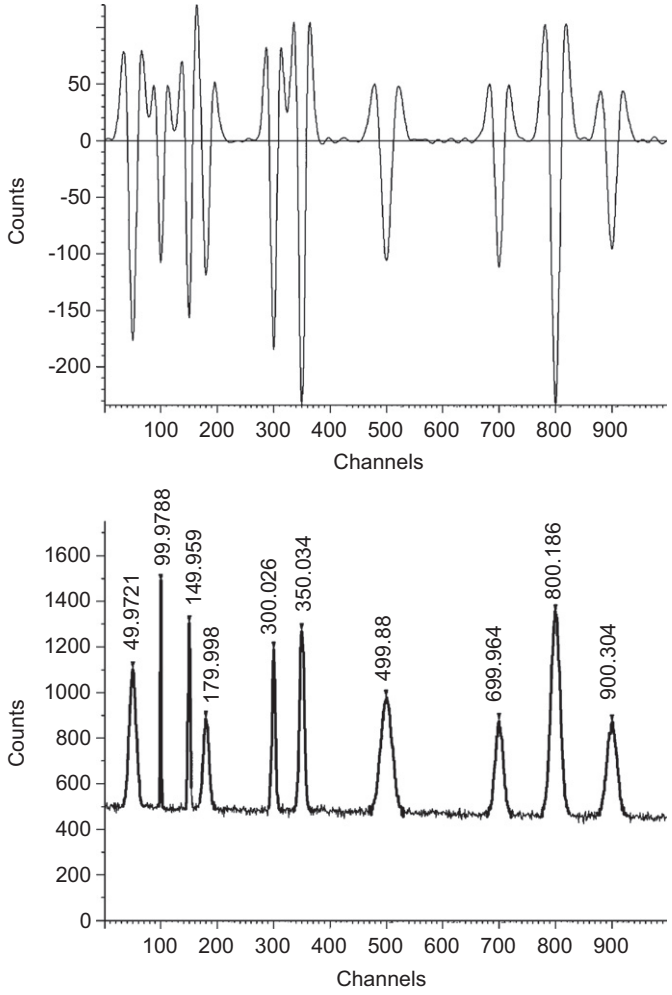

Fig. 5. The synthetic spectrum from Fig. 4a twice convolved with the first derivative of Gaussian function (a) and determined peaks positions (denoted by markers) and their regions (drawn as thick lines) (b).

Further in Fig. 6a we present a noisy spectrum containing 10 overlapping peaks of rather different widths (see Table 2) with outlined peaks regions determined by employing the above described algorithm and in Fig. 6b its two-pass convolution with the first derivative of a Gaussian function (δ=10).

The spectrum was generated using a ROOT benchmark [24]. Again in several places the estimates of peaks regions are incorrect. It is most evident for the two rightmost overlapping peaks. So we can conclude that this algorithm is not suitable for overlapping doublets or multiplets. The problem lies in the inaccurate estimate of the crossing-over points with zero value x_0 for closely positioned peaks. The estimate of σ according to Eq. (12) is rather sensitive to this error.

Therefore it is unavoidable to improve this algorithm. During the study of the examples given above we observed that the estimate of σ for appropriate peaks depends sensitively on the chosen value of δ which was a free parameter. When studying the function $c_2(x)$ by solving $dc_2(x)/d\delta$=0 it can be proved [21] that it reaches its maximum value for $\delta=\sigma$.

Further let us define the inverted positive double pass first derivative of the Gaussian

$$g(x) = \begin{cases} -c_2(x) & \text{if } c_2(x)<0 \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

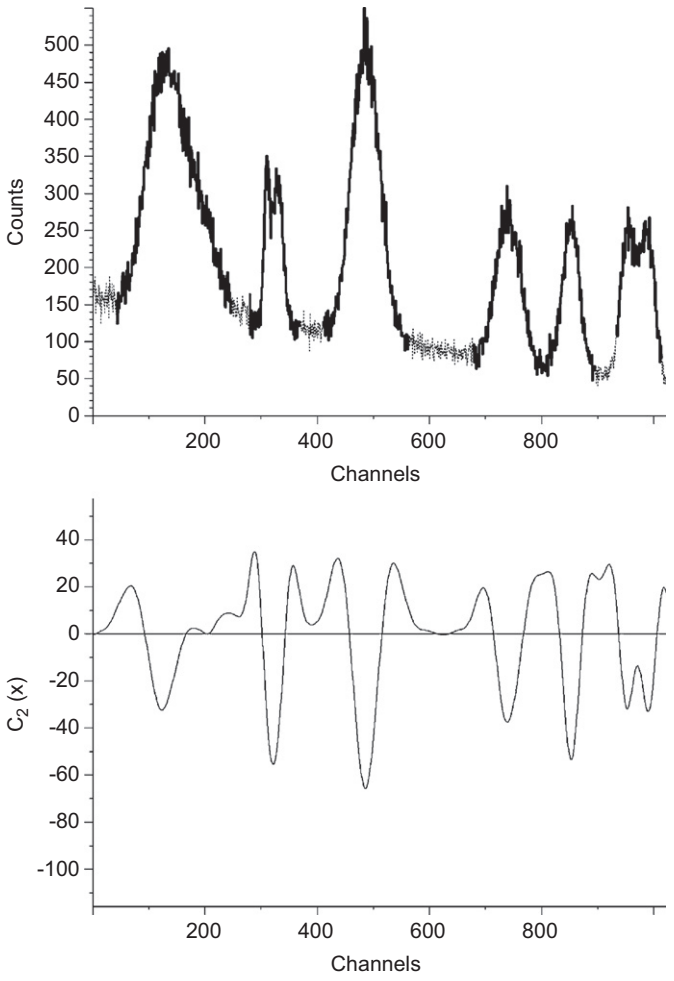

Fig. 6. Noisy spectrum containing 10 peaks of rather different widths with outlined peaks regions (a) and its two-pass convolution with the first derivative of Gaussian function (δ=10) (b).

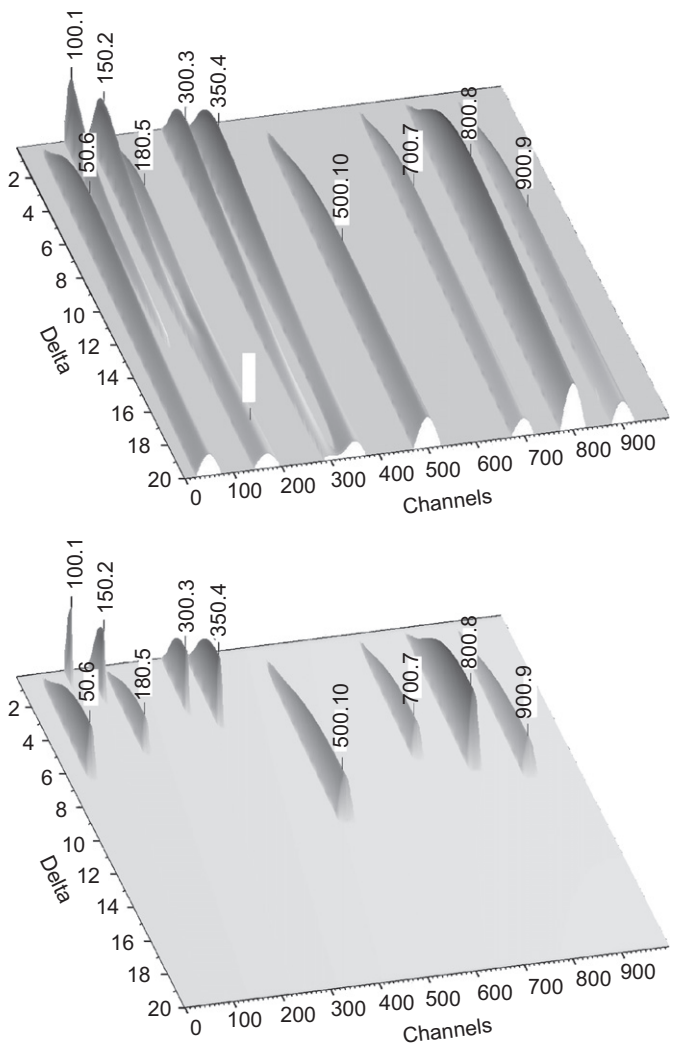

Fig. 7. Two-dimensional matrix $M(x,\delta)$ for the spectrum from Fig. 4a (a) and the same matrix after cutting-off ridges according to Eq. (14) (b).

Table 2
Positions and σ of peaks in the synthetic spectrum in Fig. 6a.

Peak #	Generated peaks (position/σ)
1	118/26
2	162/41
3	310/4
4	330/8
5	482/22
6	491/24
7	740/21
8	852/15
9	954/12
10	989/13

The basic idea for the new proposed algorithm is to scan the whole range of possible σ, i.e., to generate a series of functions $g(x)$ to form a matrix $M(x,\delta)$ from them and subsequently to find local maxima. The positions of the maxima give both positions of peaks and their σ. Let us generate a matrix $M(x,\delta)$ for the spectrum given in Fig. 4a for $\delta \in \langle 0.2, 20\rangle$with the step 0.2, i.e., we generate 100 rows of the matrix. In practice, in the software one has to set a maximum σ, which is divided into 100 distinct values and 100 functions $g(x)$ are generated and formed into the matrix. The example of the matrix for the spectrum from Fig. 4a is shown in Fig. 7a. By finding local maxima in these smoothed data one can determine the positions of peaks as well as their σ. The first coordinate determines the channel of the peak position, the second one determines its σ. The maxima with both values are denoted by markers. Both are identical with peaks parameters given in Table 1 except for the fake peak at the position 170 with δ=18. Some additional measures must be taken to ignore the fake peaks, which arise for higher values of σ and for closely positioned peaks (in this case for the peaks at the positions 150 and 180). It was found that for real peaks σ_l and σ_r estimated according to Eq. (12) should satisfy the criterion

$$\sigma_l \geqslant \delta \quad \text{or} \quad \sigma_r \geqslant \delta. \tag{14}$$

This criterion is not satisfied for the peak at the position 170 with δ=18. The matrix $M(x,\delta)$ with cut-off ridges according to condition (14) is shown in Fig. 7b. One can see that the fake peak has been removed from the matrix.

The spectrum from Fig. 4a with regions estimated using a new proposed algorithm is presented in Fig. 8. Apparently the peak regions of all peaks (including peaks 3 and 4) are estimated more accurately than in Fig. 5b.

The example of the matrix with cut-off ridges for the spectrum from Fig. 6a is shown in Fig. 9.

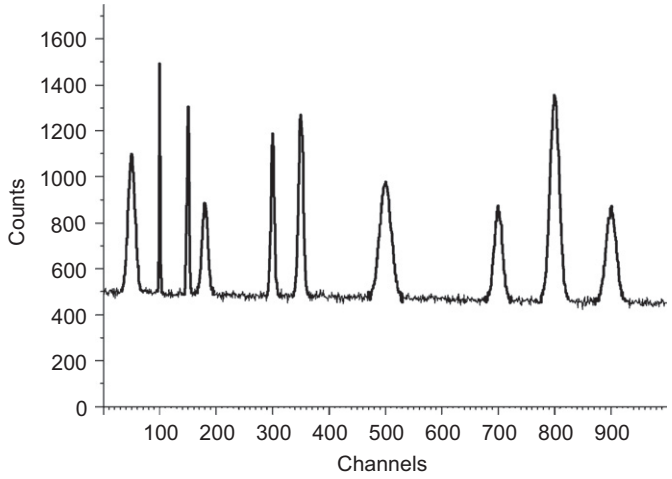

Fig. 8. Spectrum from Fig. 4a with regions estimated using a new proposed algorithm.

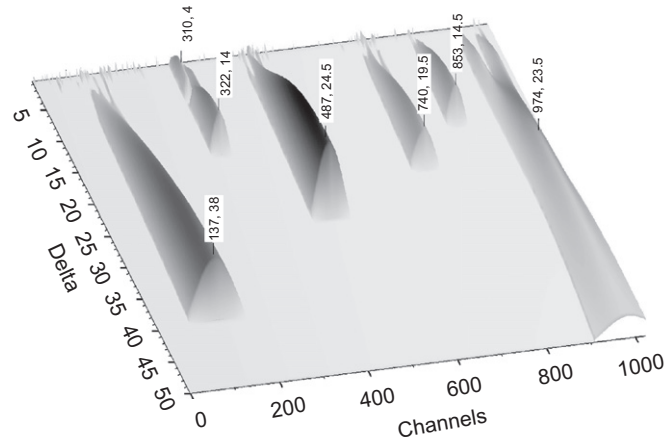

Fig. 9. Two-dimensional matrix $M(x,\delta)$ for the spectrum from Fig. 6a after cutting-off ridges according to Eq. (14).

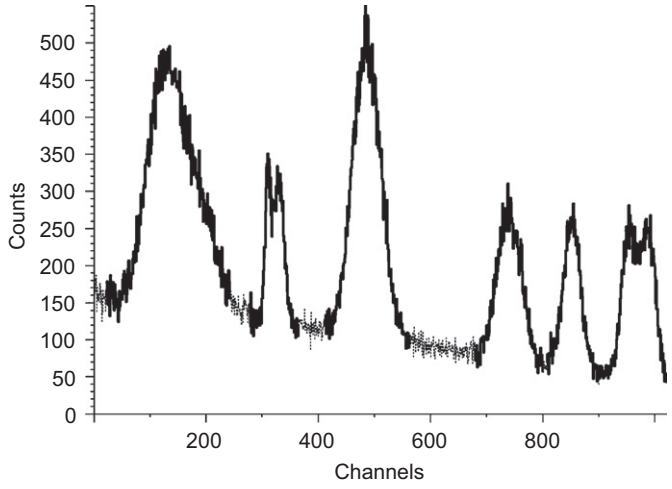

Fig. 10. Spectrum from Fig. 6a with regions estimated using a new proposed algorithm.

Finally in Fig. 10 we present the spectrum from Fig. 6a with regions estimated according to the proposed method. Again we can state that the estimate of peak regions is more accurate than that presented in Fig. 6a.

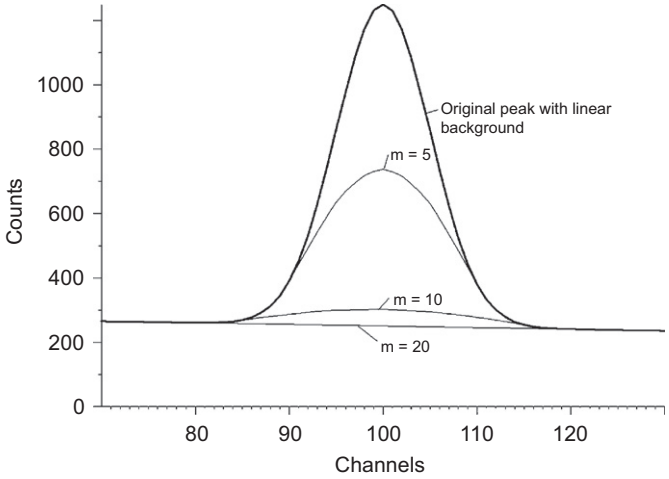

Fig. 11. Illustration of the influence of parameter m on the estimated background.

3. Background estimation algorithm with clipping window adaptive to peak regions widths

The SNIP peak clipping algorithm of background estimation has been developed in Ref. [1]. From the original spectrum $y(i)$ we calculate step by step the vectors $y_1(i)$, $y_2(i)$ up to $y_m(i)$ where m is a given user-adjustable parameter and i is $1 \leqslant i \leqslant n$. It should be chosen so that

$$2m + 1 = w \tag{15}$$

where w is the width of a preserved object. The new value in the channel i in the p-th iteration is obtained by

$$y_p(i) = \min\{y_{p-1}(i), \tfrac{1}{2}[y_{p-1}(i+p) + y_{p-1}(i-p)]\}. \tag{16}$$

After the vector $y_m(i)$ is calculated, we obtain the resulting baseline spectrum. The original SNIP algorithm in the form of C language commands [15] is expressed in the Algorithm A.

Algorithm A.

```
for(p=1;p<=m;p++){
  for(i=p;i<n−p;i++){
    a1=y[i];
    a2=(y[i−p]+y[i+p])/2;
    z[i]=min(a1,a2);
  }
  for(i=p;i<n−p;i++)
  y[i]=z[i];
}
```

We need one working vector $\boldsymbol{z}$. Both vectors have length n while m is a parameter given by the user. The influence of the parameter m on the estimated background is illustrated in Fig. 11.

One can see that the parameter m strongly influences the shape of the estimated background. A correct value of the parameter m is a necessary condition for the correct estimation of the background. Moreover, so far we assumed a constant width of the peaks.

However, for some kinds of spectra, the width of peaks can be changing with energy (e.g. high-energy spectra). In other applications the number of peaks can be enormous and consequently the peaks can be overlapping (e.g. γ-ray spectra). In both cases the estimation of one correct value m valid for the whole range of the spectrum is impossible and relation (15) is not

applicable. The widths of objects (individual peaks, clusters of peaks) in the spectrum cannot be described by one parameter m.

Let us now propose a new SNIP algorithm with a clipping window adaptive to peak regions widths. Let us assume we have determined peak regions and peak-free regions in the spectrum $\boldsymbol{y}$ using the algorithm derived in the previous section. Let us assume we have identified k regions. Then it is easy to construct a vector

$$\boldsymbol{r} = [0,0,\ldots 0,w_1,\ldots w_1,0,\ldots,0,w_2,w_2,\ldots, w_2,0,\ldots,0,\ldots,0,\ldots 0,w_k,w_k,\ldots w_k,0,0,\ldots,0]^{\mathrm{T}} \quad (17)$$

of the length n, where zero values stand in the positions in peak-free regions and the value w_j represents the width of the j-th region and stands in the positions of the j-th region. Let

$$m = \max\{w_j\}, \quad j \in \langle 1,k\rangle. \quad (18)$$

Then the SNIP algorithm with clipping window adaptive to peak regions widths can be defined as follows:

$$y_p(i) = \begin{cases} \min\{y_{p-1}(i), \frac{1}{2}[y_{p-1}(i+p)+y_{p-1}(i-p)]\} & \text{if } p \leqslant r(i) \\ y_p(i) & \text{otherwise} \end{cases} \quad (19)$$

where $p \in \langle 1,m\rangle$. The new proposed SNIP algorithm in the form of C language commands can be written as

Algorithm B.

```
for(p=1;p<=m;p++){
  for(i=p;i<n−p;i++){
    if(p<=r[i]){
      a1=y[i];
      a2=(y[i−p]+y[i+p])/2;
      z[i]=min(a1,a2);
  }
    else
      z[i]=y[i];
  }
  for(i=p;i<n−p;i++)
    y[i]=z[i];
}
```

4. Discussion and results

Initially the SNIP algorithm was proposed for an increasing clipping window, i.e. the window was changing starting from the value 1 up to the given value m (see Algorithm A). Due to the presence of the noise, the estimated background at the edges of peaks goes under the peaks. Since the operation of comparison in the above-proposed algorithm is nonlinear, the result achieved by employing the algorithm with decreasing window is rather different. The algorithm with decreasing clipping window removes the above-mentioned defect. The implementation of this very simple idea gives a substantially smoother estimated background [16]. The program implementation of the algorithm with decreasing window is given in Algorithm C.

Algorithm C.

```
for(p=m;p>=1;p−){
  for(i=p;i<n−p;i++){
    a1=y[i];
    a2=(y[i−p]+y[i+p])/2;
    z[i]=min(a1,a2);
  }
  for(i=p;i<n−p;i++)
    y[i]=z[i];
}
```

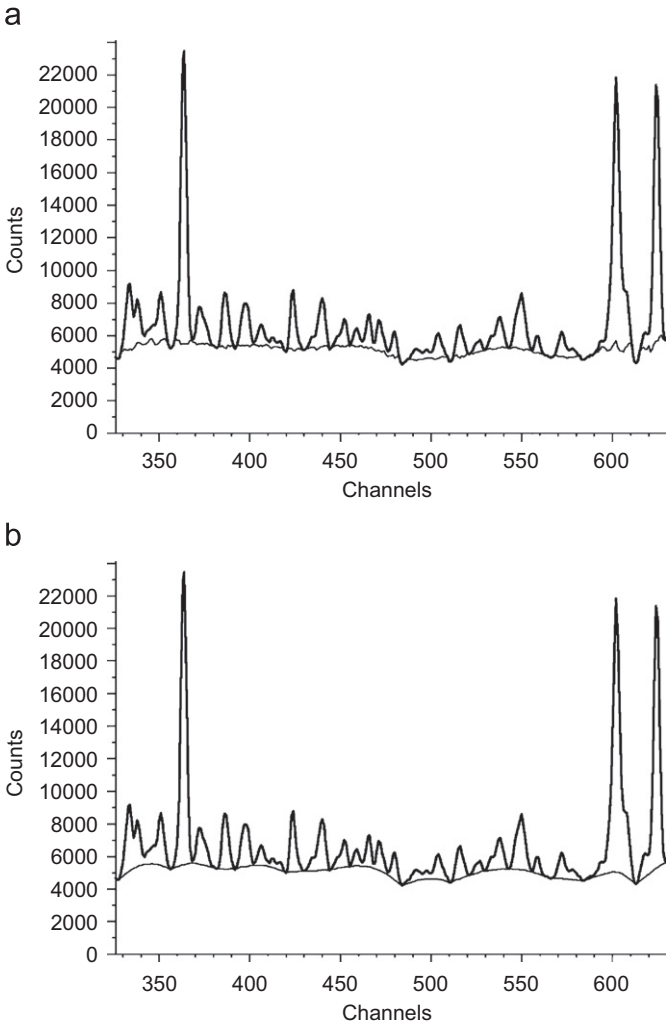

Fig. 12. An example of γ-ray spectrum with estimated background using increasing (a) and decreasing (b) clipping window, respectively.

Again, to illustrate both algorithms, in Fig. 12 we present an example of a γ-ray spectrum with estimated background using increasing and decreasing clipping window, respectively. Apparently the estimate with decreasing clipping window gives a more realistic estimate.

Further, for noisy data, the operation of comparison in Eq. (16) is a source of systematic error. The estimated background copies the minimum values of noise spikes. Therefore, we have developed a more sophisticated background estimation algorithm with simultaneous smoothing [16]. The new value in the channel i in the p-th iteration step is determined according to the following algorithm:

Algorithm D. Let us define

$$a = y_{p-1}(i), \quad b = \frac{1}{2}[y_{p-1}(i-p)+y_{p-1}(i+p)]$$
$$c = \frac{1}{2w+1}\sum_{j=-w}^{j=w} y_{p-1}(i+j) \quad (1)$$

where w is a smoothing window. Then

$$y_p(i) = \begin{cases} b & \text{if } b<a \\ c & \text{otherwise.} \end{cases}$$

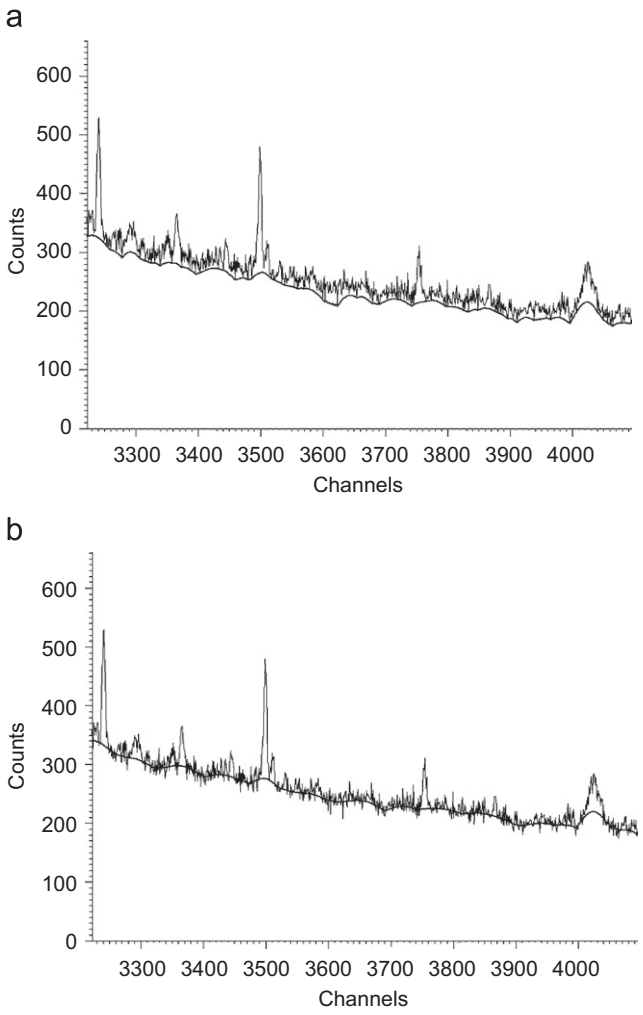

Fig. 13. An example of γ-ray spectrum with estimated background using SNIP algorithm with decreasing clipping window without (a) and with (b) simultaneous smoothing.

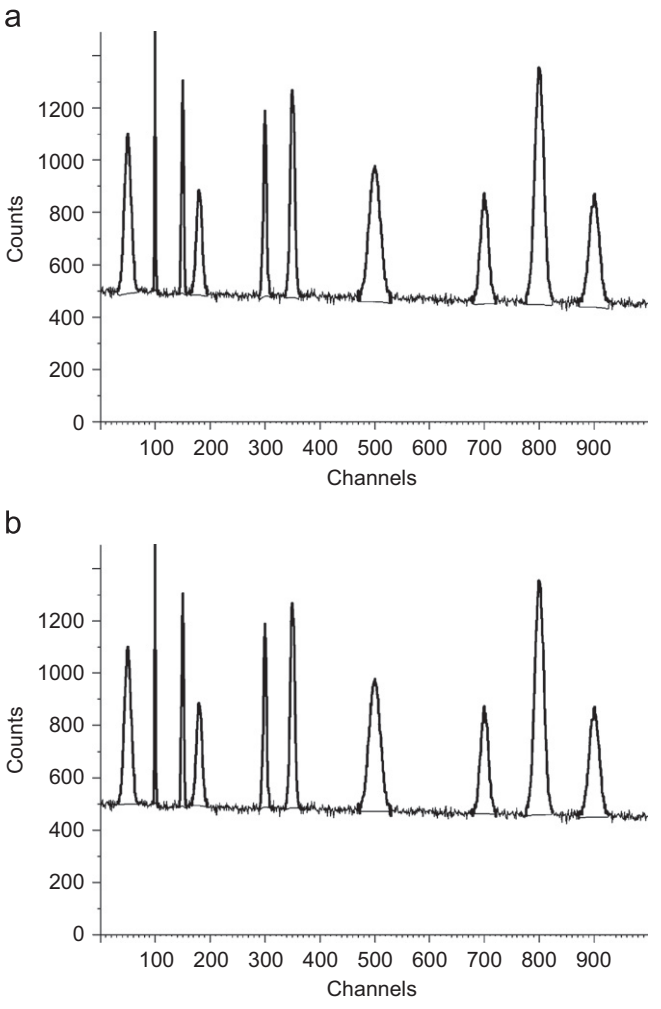

Fig. 14. Synthetic spectrum from Fig. 4a. with background estimated using SNIP algorithm with decreasing clipping window adjusted to peak regions widths using non-smoothing (a) and smoothing (b) approach.

In other words, we compare the average value of the smoothed spectrum in the symmetrically positioned points $(i-p)$, $(i+p)$ with the value (non-smoothed) in the point i. If the value in the point i is bigger, it is replaced by the average of smoothed values of the points $(i-p)$, $(i+p)$, i.e., the point i is situated inside of a peak region. Otherwise, the point i is situated outside of the peak region and we replace it by its smoothed value.

In Fig. 13 we can see an example of a γ-ray spectrum with estimated background using the SNIP algorithm with decreasing clipping window without and with simultaneous smoothing. For noisy data the example clearly proves in favor for the algorithm with simultaneous smoothing.

Now, let us turn back to the background estimation algorithm with clipping window adaptive to peak regions widths. One can freely combine this algorithm with both increasing/decreasing and non-smoothing/smoothing alternatives, i.e., algorithms C and D. In Fig. 14 we applied the above proposed algorithm B (with decreasing clipping window) combined with both non-smoothing and smoothing algorithms, respectively. The estimate is independent of the widths of the peaks. Outside of the peak regions the background coincides with the original data. However, inside of them, the estimate without smoothing (Fig. 14a) is influenced and biased by the present noise.

Next in Fig. 15 we present synthetic spectrum from Fig. 6a with background estimated using the above proposed algorithm with

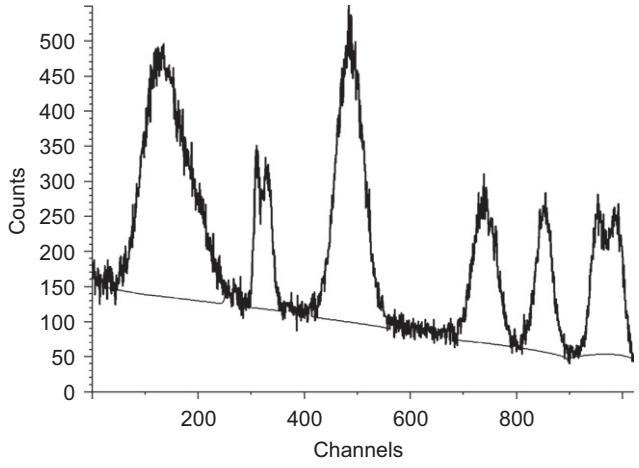

Fig. 15. Spectrum from Fig. 6a with background estimated using SNIP algorithm with simultaneous smoothing with decreasing clipping window adjusted to peak regions widths.

simultaneous smoothing and decreasing clipping window. It represents a very difficult task as it contains very closely positioned peaks of different widths (see Table 2). The estimate

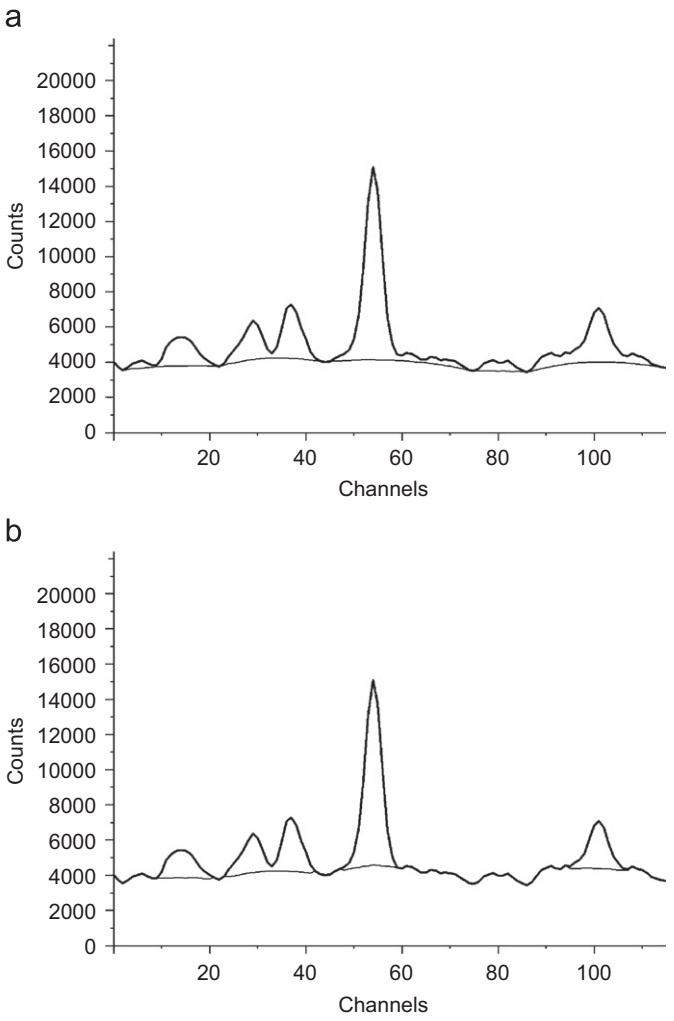

Fig. 16. Experimental γ-ray spectrum with estimated background using fixed width of clipping window given as user defined parameter (a) and width automatically adjustable to the widths of peak regions (b).

of the background is very close to the original linear background introduced into the synthetic data.

Let us proceed to the experimental spectra. In Fig. 16a we present an experimental γ-ray spectrum with an estimated background using a fixed width of the clipping window given as user defined parameter. In Fig. 16b we show the same spectrum with the background estimated with the decreasing clipping window automatically adjustable to the widths of the peak regions. One can see that the estimate with fixed window width in peak-free regions undergoes the data. After background subtraction there remain rests of data which can generate some false results.

Finally in Fig. 17 we show the noisy experimental γ-ray spectrum of Fig. 13 with the background estimated with the decreasing clipping window automatically adjustable to the widths of peak regions and simultaneous smoothing. Apparently, after comparison with Fig. 13 the presented result speaks in favor of the proposed algorithm.

5. Conclusion

At the beginning of the paper we study the algorithms for the estimation of peak regions suggested in Refs. [18,21]. We propose improvements of the methods and afterwards we outline new

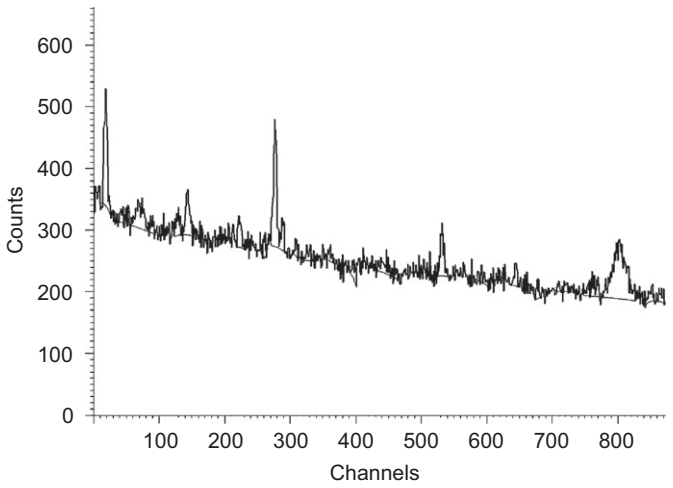

Fig. 17. Noisy γ-ray spectrum from Fig. 13 with the background estimated using the algorithm with clipping window adjusted to peak regions.

approach for the determination of the peak regions. Based on the double convolution operation with the first derivative of Gaussian presented in Ref. [21] we suggest a combined approach which allows for the determination of peak positions simultaneously with their widths. This method implicitly eliminates the background in the spectra.

However, peak searching algorithms based on convolution operation with either the second or the first derivative of Gaussian suffer from insufficient resolution. To improve the resolution very frequently the deconvolution methods are employed [25,26]. Before one can apply the deconvolution operation it is necessary to remove background from the spectrum. In Refs. [15,16] we proposed efficient algorithms for the determination of the background based on the SNIP approach. But these algorithms were proposed for the fixed clipping window width given as a parameter by user. In the paper we devised a new algorithm with the clipping window adaptable to the widths of peak regions obtained by the above defined algorithm. Together with other improvements presented in the paper (direction of the change of clipping window, simultaneous smoothing) the proposed background estimation algorithm provides a tool for more correct removal of the background than the algorithms known so far.

Moreover the algorithm for separation of peaks containing regions from peak-free regions can be utilized for other analysis operations. For instance it can be utilized for fitting purposes to confine the fitting regions. This can result in much better fit of the experimental data.

The above described algorithms are rather simple and allow simply their implementation on PC computers. The procedures are fully automatic and only a minimum intervention of the user is required. The method was implemented and integrated in the DaqProVis system [27,28].

Acknowledgment

The work is supported by the Grant Agency of Slovak Republic through Contract GAV 2/7117/27.

References

[1] C.G. Ryan, E. Clayton, W.L. Griffin, S.H. Sie, D.R. Cousens, Nucl. Instr. and Meth. B 34 (1988) 396.
[2] W. von der Linden, V. Dose, J. Padayachee, V. Prozesky, Phys. Rev. E 59 (1999) 6527.

[3] P.L. Ryder, in: K.F.J. Heinrich, D.E. Newbury, R.L. Myklebust, C.E. Fiori (Eds.), Proceedings of the Workshop on Energy Dispersive X-ray Spectrometry, Gaitherburg, 1979, NBS Special Publication 604, US Government Printing Office, Washington, DC, 1981, p. 177.
[4] P.J. Statham, X-ray Spectrum 5 (1976) 16.
[5] H.R. Ralston, G.E. Wilcox, in: J.R. Devoe (Ed.), Modern Trends in Activation Analysis, NBS Special Publication 312, US Government Printing Office, Washington, DC, vol II, 1969, p. 1238.
[6] F.H. Schamber, A modification of the linear least-squares fitting method which provides continuum suppression, in: T.G. Dzubay (Ed.), X-ray Fluorescence Analysis of Environmental Samples, Ann Arbor Science, Michigan, 1977, p. 241.
[7] R. Fischer, V. Dose, K.M. Hanson, W. von der Linden, Bayesian background estimation, in: J.T. Rychert et al.(Ed.), Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 19th International Workshop, American Institute of Physics, 2001, p. 193.
[8] G. Schulze, et al., Appl. Spectrosc. 59 (2005) 545.
[9] B. Groszwendt, Nucl. Instr. and Meth. 93 (1971) 461.
[10] Y. Karasaki, Nucl. Instr. and Meth. 133 (1976) 335.
[11] T.J. Kennet, W.V. Prestwich, R.J. Tervo, Nucl. Instr. and Meth. 190 (1981) 313.
[12] W. Westmeier, Nucl. Instr. and Meth. 180 (1981) 205.
[13] E. Clayton, PIXAN—The Lucas Heights PIXE Analysis Package, Australian Atomic Energy Commission, Report AAEC-M113, 1986.
[14] D.D. Burgess, R.J. Tervo, Nucl. Instr. and Meth. 214 (1983) 431.
[15] M. Morháč, J. Kliman, V. Matoušek, M. Veselsḱ, I. Turzo, Nucl. Instr. and Meth. A 401 (1997) 113.
[16] M. Morháč, V. Matoušek, Appl. Spectrosc. 62 (2008) 91.
[17] M.A. Mariscotti, Nucl. Instr. and Meth. 50 (1967) 309.
[18] I.A. Slavic, Nucl. Instr. and Meth. 112 (1973) 253.
[19] I.A. Slavic, Ph.D. Thesis, IBK 1011, 1971.
[20] Z.K. Silagadze, Nucl. Instr. and Meth. A 376 (1996) 451.
[21] A. Likar, T. Vidmar, J. Phys. D Appl. Phys. 36 (2003) 1903.
[22] A. Likar, T. Vidmar, M. Lipoglavšek, J. Phys. D Appl. Phys. 37 (2004) 932.
[23] A. Likar, T. Vidmar, Acta Phys. Slovaca 53 (2003) 165.
[24] R. Brun, et al., An Object-Oriented Data Analysis Framework, Users Guide 3.02c, CERN, 2002.
[25] M. Morháč, J. Kliman, V. Matoušek, M. Veselsḱ, I. Turzo, Nucl. Instr. and Meth. A 401 (1997) 385.
[26] M. Morháč, Nucl. Instr. and Meth. A 559 (2006) 119.
[27] M. Morháč, J. Kliman, V. Matoušek, I. Turzo I., Nucl. Instr. and Meth. A 389 (1997) 89.
[28] M. Morháč, V. Matoušek, I. Turzo, J. Kliman, Nucl. Instr. and Meth. A 559 (2006) 76.