Evaluation on least square method applied to gamma spectrum de-noising

Fei Li, Xinle Lang, Yi Chen, Liangquan Ge, Li Feng, Siwei Li

a Applied Nuclear Technology in Geosciences Key Laboratory of Sichuan Province, China
b College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, China
c Chengdu Shude foreign language school

* Corresponding author and e-mail: Liangquan Ge, glq@cdut.edu.cn

Abstract. Gamma energy spectrum is a way to confirm the type and content of radioactive nuclides in substances. In order to obtain effective data of gamma spectrum, it is necessary to analyze the data of gamma spectrum obtained by the detector. Nevertheless, due to the influence of statistical fluctuation and gamma generated by the environment itself during the measurement of gamma spectrum, the obtained gamma spectrum is not accurate enough as a result of needing spectral analysis of gamma spectrum, and noise reduction can make the spectral data smoother. Therefore, this paper studies the noise reduction algorithm for gamma spectrum analysis, analyzes the design and application of different algorithms in the process of noise reduction, and summarizes the advantages and disadvantages of this algorithm and others.

1. Introduction
Gamma spectrometry is regarded as an intuitive instrumental analytical technique for its fast, reliable and non-destructive qualitative and quantitative analysis of radionuclides. The analysis systems generally consist of scintillators, photomultiplier tubes, electronic circuits and back-end displays. The principle is that gamma rays released during the process of gamma decay of nuclides interact with the scintillator in the instrument in terms of photoelectric effect, Compton effect, electron pair effect and so on, which are finally output, collected, recorded and analyzed in the form of voltage pulses. Continuous development of detector technology has improved accuracy and precision, resulting in rapid progress in gamma spectrum analysis technology. Quantitative analysis of gamma spectrum data lead to errors due to the inherent statistical fluctuations, electronic system noise, measurement time and other factors. In order to obtain accurate radioactivity of various nuclides from gamma spectrum and keep the data feature of spectrum data as far as possible, it is necessary to reduce the influence of various noise[1][2]. The denoising algorithm is to eliminate the interference information in the original spectrum and form a reasonable smooth curve by means of statistical mathematics. At present, there are many smoothing methods for gamma spectrum, including sliding average method [3], barycenter method [4], wavelet transform method [5-6], Gauss method and least square method [7].

The least squares method is one of the most widely used curve fitting methods, which use linear regression to transform a non-linear problem into a linear problem and establish a fitting function, taking the optimal solution of the function as the root of the equation system. Barton et al selected Savitzky-
Golay smoothing method for data processing, which adopted the least square method for fitting and noise reduction at the window to improve the accuracy of the results to be measured [8]. Experimental results show that this algorithm has superior noise reduction capability and the SNR is improved by 100%. Zaiyu Duan et al. obtained gamma spectrum data by using NaI(Tl) scintillation detector, programmed on MATLAB to achieve smoothing processing, and improved the least square method by using multi-point smoothing. The method basically suppresses the statistical fluctuations, and the weak peaks are also evident, which shows that the least square method and the barycenter method have a good smoothing effect on energy spectrum processing in MATLAB [4]. Guobing Yu introduced the principle and algorithm of weighted least square method (WLSQ) in the study of NaI (TI) scintillation detector, and programmed it on MATLAB. Experimental results showed that the processing error of weighted least square method was lower than 10% compared with HPGe spectrometer [9]. Liang Chen proposed a nuclide recognition algorithm for portable gamma spectrometer. In the process of smoothing energy spectrum data, the time-domain least square method was used for noise reduction, which could prepare for peak searching and spectrum resolution in the following[10]. Ming Hu proposed the design principle of the least square vector machine to reduce the noise of airborne gamma spectrum. Based on the principle of statistics, the method is applied to segment regression fitting noise reduction of airborne gamma spectrum based on the distribution of energy channel addresses. It is concluded that the segmented LS-SVM noise reduction method can significantly reduce the statistical fluctuation of airborne gamma ray at the segmented region, and has good adaptability and generalization [7].

Nevertheless, the definition of optimal solution often has different interpretations according to the experimental object and the measured data. In this paper, noise reduction output of second, third and fourth order polynomials under different conditions is analyzed based on the least square method for gamma spectrum data. Data processing of various cases is compared and summarized, and the fitting effect of least squares polynomials is evaluated accurately. By using MATLAB platform for noise reduction fitting, different energy windows of gamma spectrum were respectively processed to evaluate the noise reduction effect of least square method in different energy windows, which provided important reference basis and practical analysis for the research on noise reduction technology of gamma spectrum.

2. Theory

2.1. Algorithms and principles of least square method

When the sum of squares of all parameter points and fitting points (assuming that the error of plane figure is the distance between data points and fitting points) is the minimum, that is, the sum of squared errors is the smallest. Then the energy spectrum curve fitted by the least square method can be obtained. The mathematical principle is as follows:

Assuming a given set of known data: \( \{ (x_i, y_i) \ (i=1,2,3,\ldots,m) \} \), if the mathematical model of fitting curve is \( y = f(x) \), then the distance from the ith error is \( f(x_i) - y_i \), so the sum of the squares of all points is \( \sum_{i=1}^{m} [f(x_i) - y_i]^2 \). The minimum value of \( \sum_{i=1}^{m} [f(x_i) - y_i]^2 \) and the corresponding parameters are obtained by calculating the zero point of derivative of \( \sum_{i=1}^{m} [f(x_i) - y_i]^2 \) as a function. Finally, the fitting function curve \( y = f(x) \) is obtained.

If there are known data points \( \{(x_i, y_i) \ (i=1,2,3,\ldots,m)\}, \) \( \phi \) denotes a class of functions consisting of polynomials of no more than \( m \), thus the solution of \( f_k(x_i) = \sum_{k=0}^{m} a_k x_i^k \notin \phi \) needs to satisfy:

\[
\sum_{i=1}^{m} [f(x_i) - y_i]^2 = \sum_{i=1}^{m} (\sum_{k=0}^{m} a_k x_i^k - y_i)^2 = \text{min}
\]

Formula (1) is a polynomial function, so its condition of minimum value existing and satisfying is that its corresponding partial derivative value is zero. Therefore, there are two equations:

\[
\frac{\partial I}{\partial a_j} = 2 \sum_{i=1}^{m} (\sum_{k=0}^{m} a_k x_i^k - y_i) x_i^j = 0, \quad j = 0, 1, \ldots, n \tag{2}
\]

\[
\sum_{k=0}^{m} (\sum_{i=1}^{m} x_i^{k+j}) a_k = \sum_{i=1}^{m} y_i x_i^j, \quad j = 0, 1, 2, \ldots, n \tag{3}
\]

The fitting function can be calculated as follows:

\[
f_k(x_i) = \sum_{k=0}^{m} a_k x_i^k, \quad (k=1,2,\ldots, m)
\]

\[
\sum_{k=0}^{m} a_k x_i^k, \quad (k>1)
\]
\[ a_k \quad (k=0, 1, 2...n) \] is the only solution to the equation.

2.2. Experimental data
In order to make the energy spectrum data more reliable, the Monte Carlo software Geant4 open source particle transport development platform is used to simulate the 40K energy peak data in the gamma spectrum. The experimental models are all taken from actual physics laboratory benches and equipment to ensure that the simulation data are highly consistent with the experimental results.

3. Experimental results and discussion
In this paper, aiming at the gamma spectrum data, the noise reduction effect of the least square method based on the second-order, third-order and fourth-order polynomials in different situations was compared and concluded, so as to accurately evaluate the optimal fitting effect of the least square polynomial. Gamma spectrum data was substituted into MATLAB software to obtain the fitting function parameters of various polynomials, and then the noise reduction parameters signal-to-noise ratio (SNR), RMS error (RMSE) and Pearson correlation coefficient (PCC) were calculated.

3.1. Second-order fitting function
The denoising effect parameters are obtained by denoising the address spectrum data selected by multiple gamma spectra (Table 1). It can be clearly seen that when the fitting function is \( f(x)=ax^2+b \) and \( f(x)=ax^2 \), RMSE is greater than 1.3, SNR is less than 3 and PCC is less than 0, so the fitting effect is very poor and it is not suitable to use this method for noise reduction. When the fitting function is \( f(x) =ax^2+bx+c \) and \( f(x) =ax^2+bx \), RMSE is less than 0.33, SNR is greater than 14, PCC is close to 0.9, and the fitting works well. Compared with the first two groups of methods, when the fitting function is \( f(x) =ax^2+bx+c \), RMSE is relatively small and PPC is relatively large. Although the signal-noise is relatively small, due to the small sample data sets adopted in this experiment, accidental error occurs. Conclusion: according to the following table, when the least square fitting of order 2 is performed, the noise reduction effect is most significant when the fitting function is \( f(x) =ax^2+bx+c \). (Figure 1)

![Figure 1. Comparison diagram of the fitting curve of the second-order fitting function and the noise reduction effect](image-url)
Table 1. The noise reduction effect when the fitting function is the second order

| Serial number | Function | RMSE (average value) | SNR/db (average value) | PCC (average value) |
|---------------|----------|----------------------|------------------------|---------------------|
| 1             | \(f(x) = ax^2 + bx + c\) | 0.3275               | 14.3162                | 0.8938              |
| 2             | \(f(x) = ax^2 + bx\)       | 0.3277               | 14.3503                | 0.8814              |
| 3             | \(f(x) = ax^2 + b\)        | 1.3853               | 2.5781                 | -0.1226             |
| 4             | \(f(x) = ax^2\)            | 1.4488               | 2.4688                 | -0.1128             |

3.2. Third-order fitting function

The denoising effect parameters are obtained by denoising the address spectrum data selected by multiple gamma spectra (Table 2). It can be clearly seen that when the fitting function is \(f(x)=ax^3+b\) and \(f(x)=ax^3\), RMSE is greater than 1.3, SNR is less than 2 and PCC is less than 0, so the fitting effect is very poor and it is not suitable to use this method for noise reduction. When the fitting function is \(f(x) = ax^3 + bx^2 + c\), \(f(x) = ax^3 + bx^2\), \(f(x) = ax^3 + bx + c\), and \(f(x) = ax^2 + bx\), the fitting effect is better. When the fitting function is \(f(x)=ax^3+bx^2+cx+d\) and \(f(x)=ax^3+bx^2+cx\) RMSE is less than 0.33, SNR is close to 15, PCC is greater than 0.9, and the fitting works well. Compared the two groups of methods with better results, when the fitting function is \(f(x)=ax^3+bx^2+cx+d\), the RMSE value is relatively small and the SNR is relatively high. Although the PPC is small, due to the small sample data set adopted in this experiment, accidental errors may occur. According to the above table, when the least square fitting of order 3 is performed, the noise reduction effect is most significant when the fitting function is \(f(x)=ax^3+bx^2+cx+d\). (Figure 2)
Table 2. The noise reduction effect when the fitting function is the third order

| Serial number | Function                        | RMSE (average value) | SNR/db (average value) | PCC (average value) |
|---------------|--------------------------------|----------------------|------------------------|---------------------|
| 1             | $f(x) = ax^3 + bx^2 + cx + d$ | 0.2841               | 15.0918                | 0.9042              |
| 2             | $f(x) = ax^3 + bx^2 + c$      | 0.2925               | 14.9619                | 0.9050              |
| 3             | $f(x) = ax^3 + bx^2 + c$      | 0.6396               | 10.4212                | 0.8377              |
| 4             | $f(x) = ax^3 + bx^2$          | 0.6438               | 10.40491               | 0.8370              |
| 5             | $f(x) = ax^3 + bx + c$        | 0.389                | 14.1432                | 0.8919              |
| 6             | $f(x) = ax^3 + bx$            | 0.3896               | 14.1348                | 0.8919              |
| 7             | $f(x) = ax^3 + b$             | 1.3261               | 1.6167                 | -0.2622             |
| 8             | $f(x) = ax^3$                 | 1.3509               | 1.4651                 | -0.2622             |

3.3. Fourth-order fitting function

The denoising effect parameters are obtained by denoising the address spectrum data selected by multiple gamma spectra (Table 3). It can be clearly seen that when the fitting function is $f(x) = ax^4 + bx^3 + c$, $f(x) = ax^4 + bx^3$, $f(x) = ax^4 + b$, and $f(x) = ax^4$, RMSE is greater than 0.7, SNR is less than 10, and PCC is less than 0.8, so the fitting effect is very poor and it is not suitable to use this method for noise reduction. When the fitting function is $f(x) = ax^4 + bx^3 + cx^2 + d$, $f(x) = ax^4 + bx^3 + cx^2$, $f(x) = ax^4 + bx + c$, and $f(x) = ax^4 + bx$, RMSE is small, SNR is large, PCC is great, and fitting effect is general. When the fitting function is $f(x) = ax^4 + bx^3 + cx^2 + dx + e$, $f(x) = ax^4 + bx^3 + cx^2 + dx = ax^4 + bx^3 + cx + d$, $f(x) = ax^4 + bx^3 + cx + d$, and $f(x) = ax^4 + bx^2 + cx$, RMSE is less than 0.3, SNR is greater than 15, PCC is greater than 0.9. Among the six methods with good noise reduction effect, when the fitting function is $f(x) = ax^4 + bx^3 + cx^2 + dx + e$, RMSE is the minimum, SNR is the maximum, PCC is the maximum, and the noise reduction effect is the best. (Figure 3)

Figure 3. Fitting function $f(x) = ax^4 + bx^3 + cx^2 + dx + e(a)$ (b), $f(x) = ax^4 + bx^3 + cx^2 + dx(a)$ (d), $f(x) = ax^4 + bx^3 + cx^2 + d(e)$ (f) and $f(x) = ax^4 + bx^3 + cx^2 (g)(h)$ fitting curve and contrast diagram of noise reduction effect.
Table 3. The noise reduction effect when the fitting function is the fourth order

| Serial number | Function | RMSE (average value) | SNR/db (average value) | PCC (average value) |
|---------------|----------|----------------------|------------------------|---------------------|
| 1             | \( f(x) = ax^4 + bx^3 + cx^2 + dx + e \) | 0.2511 | 15.4255 | 0.9119 |
| 2             | \( f(x) = ax^4 + bx^3 + cx^2 + dx \) | 0.2513 | 15.4220 | 0.9118 |
| 3             | \( f(x) = ax^4 + bx^3 + cx^2 + d \) | 0.3604 | 13.8960 | 0.9087 |
| 4             | \( f(x) = ax^4 + bx^2 + cx^2 \) | 0.3622 | 13.8960 | 0.9085 |
| 5             | \( f(x) = ax^4 + bx^3 + cx \) | 0.2606 | 15.2490 | 0.9092 |
| 6             | \( f(x) = ax^4 + bx^3 + c \) | 0.7947 | 6.5093 | 0.6700 |
| 7             | \( f(x) = ax^4 + bx^3 \) | 0.8944 | 6.4163 | 0.6650 |
| 8             | \( f(x) = ax^4 + bx^2 \) | 0.2746 | 15.1109 | 0.9042 |
| 9             | \( f(x) = ax^4 + bx^2 + cx + d \) | 0.2748 | 15.1084 | 0.9041 |
| 10            | \( f(x) = ax^4 + bx^2 + c \) | 0.7411 | 9.0944 | 0.7771 |
| 11            | \( f(x) = ax^4 + dx^3 + c \) | 0.4641 | 13.1474 | 0.8734 |
| 12            | \( f(x) = ax^4 + dx^3 \) | 1.2515 | 1.1422 | -0.2309 |
| 13            | \( f(x) = ax^4 + b \) | 1.2498 | 0.9571 | -0.3503 |

According to the results of the above program fitting noise reduction, we can know that the following functions have significant noise reduction effects, including the second order \( f(x) = ax^2 + bx + c \), the third order \( f(x) = ax^3 + bx^2 + cx + d \), and the fourth order \( f(x) = ax^4 + bx^3 + cx^2 + dx + e \). When the fitting function is equal to \( f(x) = ax^4 + bx^3 + cx^2 + dx + e \), the RMSE value is the minimum, the SNR is the maximum, PCC is the maximum, and the noise reduction effect is the best. Similarly, the analysis results of 233Th window and 238U window are consistent with the results of 40K window.

4. Conclusion
In this paper, the least square method is used to reduce the noise of gamma spectrum data, and MATLAB software is utilized to complete the development and design of the program. Moreover, according to the RMSE, PCC and SNR, the noise reduction effect of each fitting function is evaluated and analyzed, and finally the function with the most significant noise reduction effect is obtained by comparison. When the fitting function is the fourth order, the evaluation effect is obviously better than that of the second and third order. The results show that the least squares method has remarkable noise reduction effect in processing gamma spectrum, especially for high frequency channel data.

Acknowledgement
This work was funded by the National Natural Science Foundation of China (No.41774147), the National Key R&D Project (No.2017YFC0602100) and Science and technology activity project for overseas personnel in Sichuan province.
References

[1] Knežević, D., Jovančević, N., Krmar, M.1;Petrović, J. Modeling of neutron spectrum in the gamma spectroscopy measurements with Ge-detectors. NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH SECTION A-ACCELERATORS SPECTROMETERS DETECTORS AND ASSOCIATED EQUIPMENT. 2016,833:23-26.

[2] Fei Li, Zhixing Gu, Liangquan Ge, Hui Li, Xinyu Tang, Xinle Lang, Bo Hu. Review of recent gamma spectrum unfolding algorithms and their application. Results in Physics. 2019,13: 102211.

[3] Junsong Ren. The Study of γ Spectrum Analysis and Radioactive Nuclide Identification Method. Southwest university of science and technology.2017.

[4] Zaiyu Duan, Jianhua Chen, Guixin Zhang, Junjun Gong. Disposal of Smooth γ Spectrum on Matlab Process. Nuclear Power Engineering .2007,28(3):125-128.

[5] Lunhui LI, Jian-feng He, Qin Wang, Yuan Wang. Study of γ Energy Spectrum Denoising Based on Improved Wavelet Threshold Method. Atomic Energy Science and Technology.2016, 50(7):1279-1283.

[6] Dongsheng Shi, Yuming Di, Chunlin Zhou. Comparative study on γ energy spectrum denoise by fourier and wavelet transforms. Nuclear Electronics & Detection Technology.2006,6(26):134-137.

[7] Ming Hu. Based on Least Square Support Vector Machine Airborne Gamma Ray Energy Spectrum Segmentation Denoising Method Research. East China university of technology.2016

[8] Boris Ryabko.2013.A confidence-set approach to signal denoising [J].Statistical Methodology, 15(2013):115-120.

[9] Guobing Yu, Shan Jiang, Chuankai Chen. Study on Efficiency of Detection in Different Geometry Sample Cup in γ-ray Spectrometer Analyze. Journal of Anhui Agricultural Sciences.2009,37(24):11641-11642.

[10] Liang Chen. Research on the Nuclide Identification Algorithm and Digital Spectra Acquisition System. Tsinghua University.2009.