COMMUNICATIONS

COMPARISON OF ALGORITHMS FOR FITTING A GAUSSIAN FUNCTION USED IN TESTING SMART SENSORS

Elena Pastuchová — Michal Zákopčan [*]

This article offers useful algorithms for estimating the parameters of the Gaussian curve fitted to observed data, gained in testing of sensors. Fitting was realized using two algorithms Caruanas and Guos, which have been compared.

K e y w o r d s: smart sensors, Gaussian function, fitting algorithm, baseline correction

1 INTRODUCTION

Contaminants in the environment are closely monitored in different folders environment. The basic groups of these substances include heavy metals. Among the most harmful elements for humans and animals include: Cd, Hg, Pb and Cr as well as Zn, As, Cu, Ni, *etc*, which are in a little found in plants and soils, but by the accumulation they become more toxic. Detection of heavy metals in the water industry is of great importance, because their presence poses serious health risk. The presence and quantity of contaminants are detected by constantly developing special sensors. In research of new technological methods in the manufacturing process of smart sensors, each material has different properties and thus is suitable for employing in various spheres. Every new technology of manufacturing of sensors or using a specially treated material in the manufacture of sensors are needed to be tested for finding characteristics of every sensor. In order to determine the quality, accuracy or sensitivity of the sensors, it is necessary to perform measurements in laboratory conditions. Testing the sensors structures is carried out by cyclic voltammetry and square wave voltammetry subsequently. [1]. In the evaluation of characteristics of sensors, it is necessary to process experimental data and quantify the relationship between measured quantities. It is optimal to find a mathematical model that describes a set of data in a way that minimizes the difference between the model and the data, it means to find the best curve fitting observed data points. By testing, the sensor measures the voltage dependence of current from the electrode intended to measure heavy metals. The presence of heavy metals identifies significant increases in current at a given voltage. Fig. 1 shows the waveform in the case of zinc, cadmium and lead.
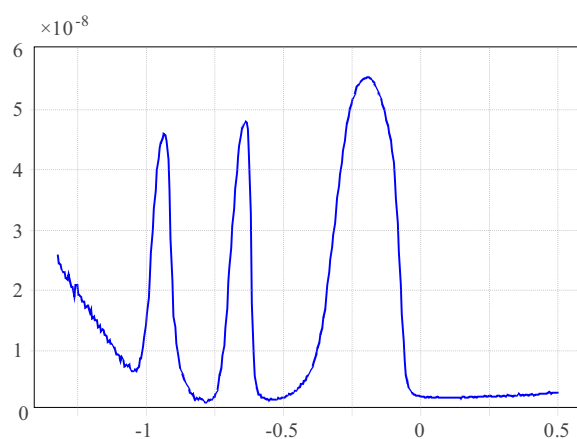

Fig. 1. Waveform in the case of zinc, cadmium and lead

This means that every peak on the graph represents the maximum of a current at given voltage at which the maximum metal detection is achieved. These peaks are necessary optimally approximated. We tried to use the fact that these peaks have Gaussian shape and we were looking for an optimal approximation algorithm.

2 CARUANAS ALGORITHM

The Gaussian function is a function in the form

$$f(x) = A \exp \frac{-(x - \mu)^2}{2\sigma^2} \qquad (1)$$

The graph of Gaussian is a characteristic symmetric bell curve, centered at the position $x = \mu$ with A being the height of the peak and σ determining the width of the bell. The focus of this paper is on the problem how to fit the observed data points by Gaussian function, which were obtained by testing sensors, and how to identify the parameters A, μ and σ precisely.

Institute of Informatics and Mathematics, Faculty of Electrical Engineering, Slovak University of Technology in Bratislava, Slovakia, elena.pastuchova@stuba.sk

The majority of Gaussian fit algorithm are based on two approaches: the first approach includes non-linear least-square optimization algorithm, the solution of such is to employ an iterative procedure like Newton-Raphson algorithm [2], the second approach is based on a fact that the Gaussian function is the exponential of a quadratic function. This function can be converted to a polynomial function by using a logarithmic transformation. A simple method of such conversion was proposed by Caruana *et al*, [3].

$$\ln f(x) = \ln A + \frac{-(x - \mu)^2}{2\sigma^2} \qquad (2)$$

after adjustment, one has

$$\ln f(x) = \ln A - \frac{\mu^2}{2\sigma^2} + \frac{2x\mu}{2\sigma^2} - \frac{x^2}{2\sigma^2} \qquad (3)$$

Modifying (3), we obtain a polynomial function

$$\ln f(x) = a + bx + cx^2 \qquad (4)$$

where $a = \ln A - \frac{\mu^2}{2\sigma^2}$, $b = \frac{\mu}{\sigma^2}$ and $c = -\frac{1}{2\sigma^2}$.

It is evident that (4) represents a parabola whose peak position is the same as that of the Gaussian function described in (1).

The error function is

$$S(x) = \ln f(x) - (a + bx + cx^2) \qquad (5)$$

Using the method of least square deals to the following linear system of equations

$$\begin{bmatrix} N & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum \ln f(x) \\ \sum x \ln f(x) \\ \sum x^2 \ln f(x) \end{bmatrix} \qquad (6)$$

where N is a number of observed data points and the symbol $\sum$ represents the sum of indexed variables $\sum_{n=1}^{N} (.)_n$.

After solving this system we can obtain parameters of the Gaussian function

$$\mu = \frac{-b}{2c}, \quad \sigma = \sqrt{\frac{-1}{2c}}, \qquad (7,8)$$

and

$$A = \exp(a - \frac{b^2}{4c}) \qquad (9)$$

Caruanas algorithm is computationally efficient, since it is noniterative. Main drawback of this algorithm is that its accuracy decreases dramatically in case of the presence of noise. Because this method uses the logarithm of data, it is clear that error creates if the observed data contains a zero value. To avoid these deficiencies we used Guos model, [4].

3 GUOS ALGHORITHM

In the case that there is an additive random noise η [4], the data, we observe, are not the ideal value $f(x)$ but

$$\hat{f}(x) = f(x) + \eta \qquad (10)$$

Consequently, the error function is

$$S = \ln \hat{f}(x) - (a + bx + cx^2) \qquad (11)$$

Expanding into Taylor series and omitting the high-order terms we obtain

$$S \approx \ln f(x) - (a + bx + cx^2) + \frac{\eta}{f(x)} \qquad (12)$$

and

$$E(S^2) = [\ln f(x) - (a + bx + cx^2)]^2 + \frac{\sigma_\eta^2}{f^2(x)} \qquad (13)$$

where σ_n^2 is standard deviation of the noise.

However the estimation using (13) is not accurate, increasingly dependent on the observed data points with low values rather than the big ones, and usually manual intervention must be carried out for thresholding the data in advance. It is therefore necessary to redefining the error function, using (5), as

$$\begin{aligned} \varepsilon &= f(x) \left[\ln(f(x) + \eta) - (a + bx + cx^2) \right] \\ &\approx f(x) \left[\ln f(x) - (a + bx + cx^2) \right] + \eta \end{aligned} \qquad (14)$$

Consequently

$$E(\varepsilon^2) = f^2(x) \left(\ln f(x) - a - bx - cx^2 \right)^2 + \sigma_\eta^2 \qquad (15)$$

Using method of least squares we achieve a linear system of

$$\begin{bmatrix} \sum \hat{y}^2 & \sum x\hat{y}^2 & \sum x^2\hat{y}^2 \\ \sum x\hat{y}^2 & \sum x^2\hat{y}^2 & \sum x^3\hat{y}^2 \\ \sum x^2\hat{y}^2 & \sum x^3\hat{y}^2 & \sum x^4\hat{y}^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum \hat{y}^2 \ln(\hat{y}) \\ \sum x\hat{y}^2 \ln(\hat{y}) \\ \sum x^2\hat{y}^2 \ln(\hat{y}) \end{bmatrix} \qquad (16)$$

where $\hat{y} = \hat{f}(x)$.

Solving (16) we obtain a, b, c and subsequently the Gaussian parameters in a similar manner as in part 2.
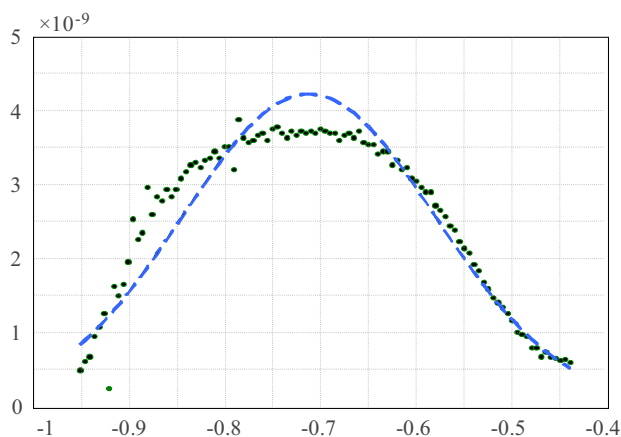
Fig. 2. Voltammogram of experimental data (dotted line) fitted by Caruanas algorithm
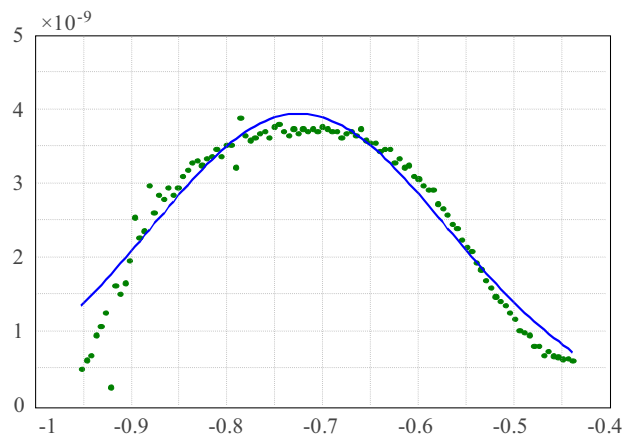

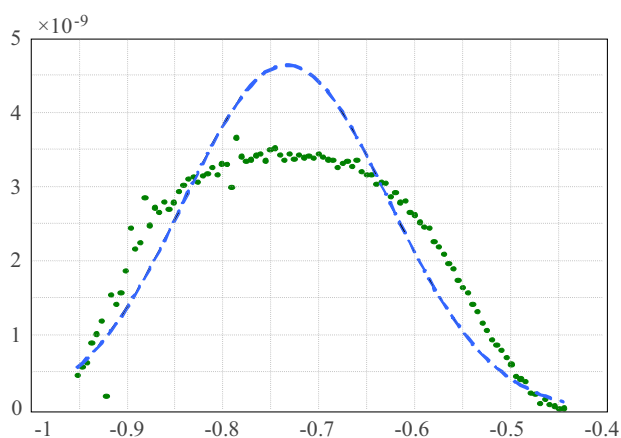**Fig. 3.** Voltammogram of experimental data (dotted line) fitted by Guos algorithm

Fig. 4. Voltammogram of experimental data using baseline correction fitted by Caruanas algorithm
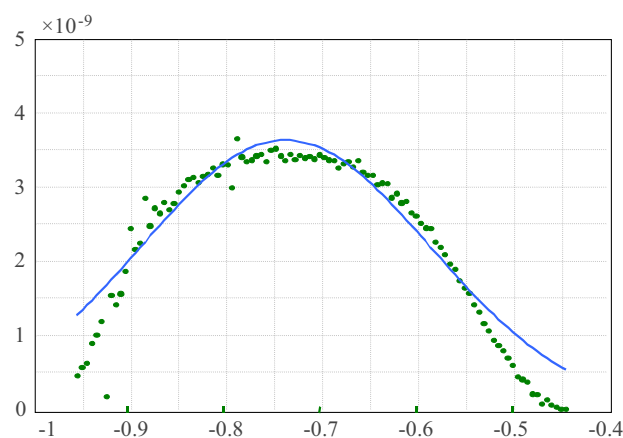

Fig. 5. Voltammogram of experimental data using baseline correction fitted by Guos algorithm

4 APPLICATION TO EXPERIMENTAL RESULTS

The parameters of Gaussian function are needed in establishing the calibration curve which is necessary to determine the detection limit. For estimation of these parameters from observed data points we used Caruanas algorithm. An example of this fitting is in Fig. 2, where $\mu = -0.7128, \sigma = 0.1331$ and $A = 4.2173 \times 10^{-9}$.

Because the presence of noise, as in Fig. 2, it is evident we decided to apply Guos algorithm. To compare algorithms of Caruana and Guo respectively, we use the same sample of experimental data. Fitting by Guos alghorithm we have received Gaussian curve in Fig. 3, where $\mu = -0.7254, \sigma = 0.1556$ and $A = 3.9264 \times 10^{-9}$.

In order to determine the kind of heavy metal in contaminated water, we need know as accurately as possible value of μ and parameter A, which is important for specifying the concentration of this metal. For these reasons we decide to correct these results. We took into consideration that it is necessary to correct the baseline. After using baseline correction and Caruanas alghorithm we received the fitting results as seen in Fig. 4, where $\mu = -0.7339, \sigma = 0.1074$ and $A = 4.6497 \times 10^{-9}$.

Using baseline correction and Guos alghorithm we obtained the fitting results as seen in Fig. 5, where where $\mu = -0.7359, \sigma = 0.1509$ and $A = 3.6489 \times 10^{-9}$.

5 CONCLUSION

We tried to focus on fitting a one dimensional Gaussian function in efforts to find optimal approximation algorithm. For this purpose we have compared Caruanas and Guos algorithms as the most appropriate methods. By fitting we found some disadvantages of these methods. While Caruana's algorithm fails to noisy data, its accuracy decreases dramatically, Guos technique is much less sensitive to random noise. In some samples, we had to make a baseline correction to obtain the actually measured values. More visual comparison of these two methods is shown in Fig. 6, where it is clear that Guos alghorithm approximates the data more accurate. Following experimental data and the calculation of the value of μ, heavy metal we identified in experiment was Cd. Both methods fail when the measured values are extremely sloping, which will be the subject of further research.
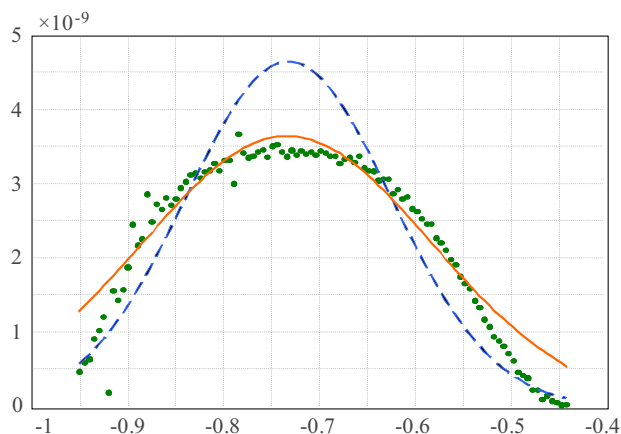
Fig. 6. Voltammogram of experimental data using baseline correction fitted by Caruanas algorithm and Guos alghorithm (Guo,s - solid line, Caruanas - dashed line)

Acknowledgement

References

[1] SZABADOS, L.—LÍŠKA, O.: Smart senzorové siete a príklady ich využitia (Smart Sensor Nets and Examples of their Application), In: http://www.sjf.tuke.sk/transferinovacii/pages/archiv/ transfer/7-2004/pdf/157-160.pdf (Accessed 2014-04-02).. (in Slovak)

[2] PRESS, W.—TEUKOLSKY, S.—VETTERLINK, W.—FLANNERY, B.: Numerical Recipes: The Art of Scientific Computing, 3rd ed., Cambridge Univ. Press, New York, 2007, pp. 733–836.

[3] CARUANA, R.—SEARLE, R.—HELLER, T.—SHUPACK, S.: Fast Algorithm for the Resolution of Spectra, Anal. Chem. **58** No. 6 (May 1986), 1162–1167.

[4] GUO, H.: A Simple Algorithm for Fitting a Gaussian Function, IEEE Signal Process. Mag. **28** No. 5 (2011), 134–137.

[5] DIAS PEREIRA, J. M.—POSTOLACHE, O.—SALGADO, R.—SILVA GIRAO, P.: Voltammetry Based Automated Instrument For In-situ and Online Measurement of Heavy Metals Concentration in Water, IEEE Transactions on Instrumentation and Measurement No. 1 (2007).

Elena Pastuchová, in 1978 graduated from the Faculty of Mathematics and Physics of the Comenius University, received her RNDr in 1990 and PhD degree in 2004. She is working at the Department of Mathematics of Institute of Computer Science and Mathematics of the Fakulty of Electrical Engineering and Information Technology, in Bratislava. Since 1990 she is involved in applied mathematics.

Michal Zákopčan was born in 1981 in Nitra, Slovakia. He graduated from the Faculty of Mathematics, Physics and Informatics of the Comenius University in 2005, received his PhD degree in 2009. He works at the Department of Mathematics of Institute of Computer Science and Mathematics of the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava. He deals with applied mathematics.