signal filtering, genetic algorithms,
spectra analysis, free radicals,
EPR spectroscopy, function approximation

Marcin BERNAS[1], Pawel RAMOS[2]

SEPARATION OF GROUPS OF FREE RADICALS FROM NOISED EPR SPECTRUM USING GENETIC ALGORITHM AND GRADIENT METHOD

Different groups of free radicals exist in biological material like animal tissues or plants parts. The processes like heating or cooling creates additional types of free radicals groups in this organic matter, due to changes in chemical bonds. The paper proposes a method to determine types and concentrations of different groups of free radicals in the matter processed in various temperatures. The method extracts the spectrum of free radicals using electron paramagnetic resonance with the microwave power of 2.2 mW. Then an automatic method to find a best possible fit using limited number of theoretical mathematical functions is proposed. The match is found using spectrum filtration, and a genetic algorithm implementation supported by a Gradient Method. The obtained results were compared against the samples prepared by an expert. Finally, some remarks were given and new possibilities for future research were proposed.

1. INTRODUCTION

Electron paramagnetic resonance spectroscopy (EPR) is an experimental method used to determine the properties of paramagnetic systems [10],[3] and [8]. Paramagnetic properties are exhibited by the samples containing free radicals with spin S = 1/2, biradicals with spin S = 1, paramagnetic ions, π electrons delocalized on multi-ring aromatic units [10],[3] and [8]. Oxygen molecules O_2 in the ground state with spin S =1 also exhibit paramagnetic properties [3]. Paramagnetic centers with unpaired electrons are a key factor used in medicine and pharmacy [3]. To find paramagnetic centers concentration and properties, EPR spectra can be used. Analysis of the spectra can give us important data like: amplitude (A), integral intensity (I), line width (Bpp) and g-factor [10],[3] and [8].

Structure of free radical groups in living matter is complex. Additionally, the thermal factor has impact on breaking chemical bonds in its cells and alters free radical structure. So far, the analysis of obtained spectrum is performed by an expert; however analysis of samples with rising microwave power over time usually takes days of expert time. The data on types and concentration of free radicals is useful both in medical treatment and analyzing the medicines, therefore, a need to create an automatic method to discern groups of free radicals emerged. There are many articles concerning EPR spectra analysis [5],[4], however no automatic solution for spectra separation for free radicals and data gathering was yet proposed.

[1]Institute of Informatics, University of Silesia, Bedzinska 39, 41-200 Sosnowiec, Poland, marcin.bernas@gmail.com.
[2]Department of Biophysics, Medical University of Silesia in Katowice, Jednosci 8, 41-200 Sosnowiec, Poland, pawelramos@sum.edu.pl.

2. PROPOSED METHOD

The proposed method was divided into five steps:

- acquiring EPR spectrum as a sample,
- filtering process to remove the noise,
- raw approximation of filtered signal using elemental functions and genetic algorithm,
- final approximation using gradient method,
- reading basic parameters from elemental functions.

The method steps were illustrated in Fig. 1.

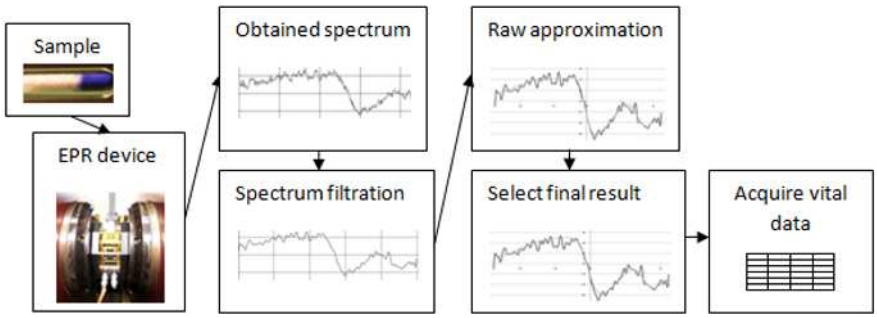

Fig. 1. Automatic spectrum analysis method block scheme.

First step consists of acquisition of the EPR spectra from biological material using magnetic modulation of 100 kHz. EPR spectra were measured at low microwave power of 2.2mW and high attenuation of 15dB. Microwave frequency (ν) was directly measured by an MCM101 recorder produced by EPRAD Company (CityplacePoznań, country-regionPoland).

The total free radical concentration (N) in the studied samples is determined according to a formula based on the reference sample:

$$N = N_u \frac{\frac{W_u A_u}{I_u}}{\frac{I}{W A_m}}, \tag{1}$$

where:

N_u - the number of paramagnetic center (1.2x10^{19}spin) in the ultramarine reference,

W and W_u - the power received for sample and the ultramarine,

A and A_u- the amplitudes of ruby signal for the sample and the ultramarine,

I and I_u- the integral intensities for the sample and ultramarine,

m - the mass of the sample.

The total free radical concentration is a value proportional to the integral intensity (I) of EPR spectrum [9]. The integral intensities were calculated by double integration of the first-derivative of EPR spectra.

3. IMPLEMENTATION

The analyzed spectra are given as a time series $f_r(x)$, $x=1..N$ representing values within domain (0, 10) with time step 0.02 (512 elements in one sample). The analyzed signal (presented in Fig. 2) is a complex one, therefore indirect analysis was proposed.

The method approximates f_r spectra using function f_p defined as a composition of elemental functions set and then performs the analysis. The previous expert research has shown [5],[4] that observed result is a derivative of the functions of absorption of free radicals. Therefore for further analysis the derivatives of basic functions were used.

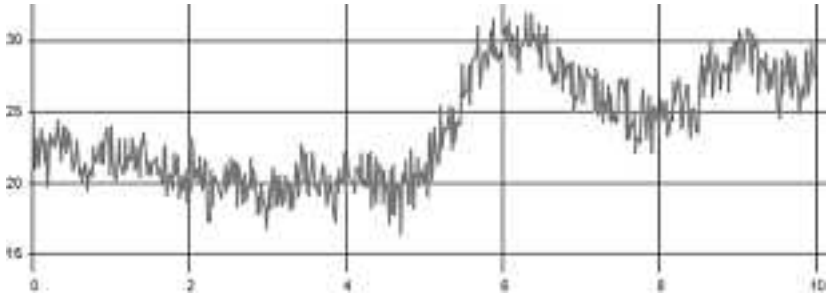

Fig. 2. Example of analyzed signal.

3.1. FILTERING PHASE AND FINDING THE REFERENCE POINT

Before the approximation procedure can take place, the noise from spectrum is removed and a reference point is defined (second step of the method). To clean the noise from the spectrum, several filtering methods were examined: average filters family [7], filters based on Fast Fourier Transformation (FFT) and Cosines Transformation (DCT) [1]. Average filters (especially moving average filter) remove the noise, unfortunately they are not applicable for this method due to the amplitude deformation in local minimum and maximum. The process of deformation is illustrated in Fig. 3.

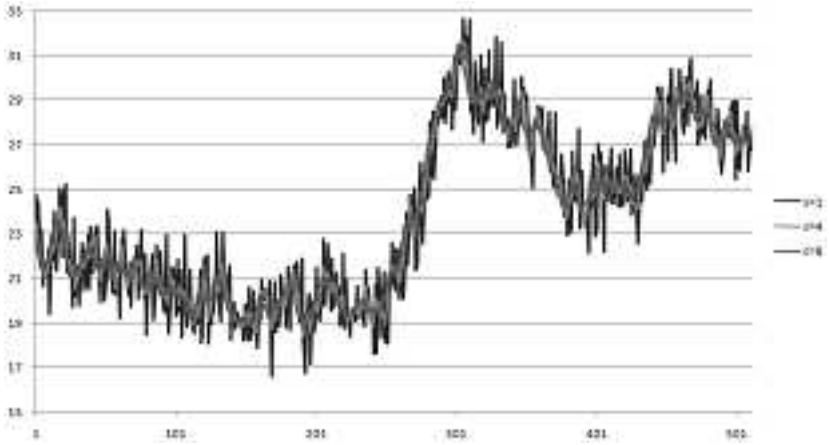

Fig. 3. Application of moving average filter.

Therefore the FFT-based filter was used. It removes the frequencies with the amplitude value below threshold 0.05 - the noise level in an EPR device.

A reference point is determined for the filtered spectra function f_r, according to the equation:

$$f_r{}' = f_r(x+dx)+dy, \tag{2}$$

where *dx* and *dy* shift is based on heuristic rules determined by an expert:

- the sum of values of samples separated by X axis is close to 0,
- the absolute value of sum of function from left side of Y axis plus the absolute value of sum of function from right side of Y should be maximized.

According to the heuristic rule, a simple linear optimization task was defined for *dx* and *dy* values:

$$\min_{dx,dy}\Big(\sum_{x=-\infty}^{\infty} f_r\left(x+dx\right)+dy\Big)$$

$$\max_{dx,dy}\Big(\sum_{x=-\infty}^{dx}\left|f_r\left(x+dx\right)+dy\right|+\sum_{x=dx}^{\infty}\left|f_r\left(x+dx\right)+dy\right|\Big)\ . \tag{3}$$

The equation 3 is unaffected by *dx* value, therefore *dy* value can be evaluated using linear approximation. The minimization and maximization tasks are performed using Simplex method [6]. The method finds a best match within defined boundaries $dx \in (-10,\ 10)$ and $dy \in (-40,\ 40)$. The initially transformed function f_r' is presented in Fig. 4.

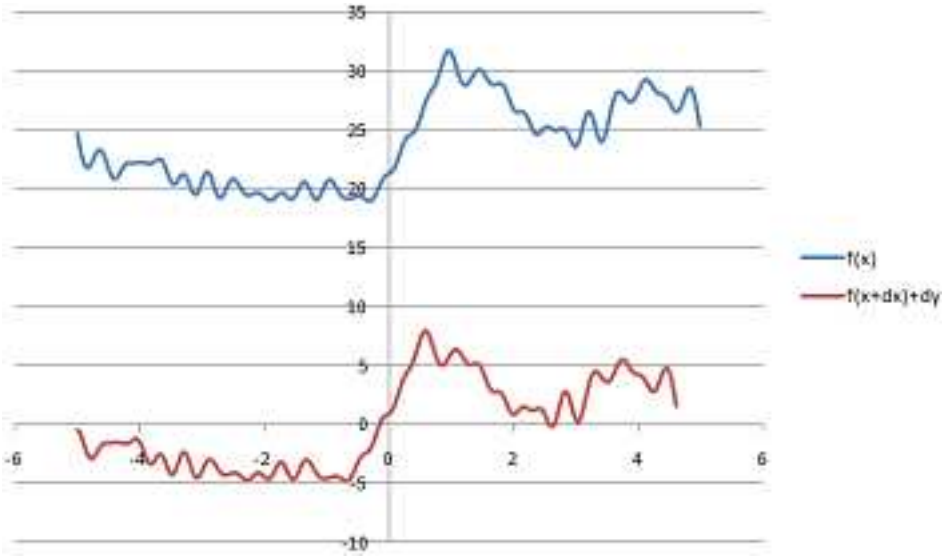

Fig. 4. Function shift according to dx, dy values.

Finally, the filtered and shifted function $f_r{}'$ is approximated in third and fourth step of the method.

3.2. THE FUNCTION APPROXIMATION

The function $f_r{}'$ is approximated by the function f_p. The function f_p is a composition of derivatives of continuous elementary functions. Based on a preliminary research, the Gaussian and Lorenz function was selected as offering the lowest error rate for the number of functions equal to or below 3 ($i \leq 3$). The function f_p was defined as follows:

$$f_p(x) = f_1'(x) \circ f_2'(x) \circ \cdots \circ f_i'(x) \circ \cdots \circ f_w'(x)$$

$$f_p(x) = p_{11} f_1'(x) + p_{21} f_2'(x) + \cdots + p_{i1} f_i'(x) + \cdots + p_{w1} f_w'(x), \tag{4}$$

$$f_i'(x) \equiv l_i'(x) \dot{\vee}\ f_i'(x) \equiv g_i'(x),\ \forall i$$

$$l_i'(x) = \frac{-2(x - p_{i2})\, p_{i3}}{\pi\left[(x - p_{i2})^2 + {p_{i3}}^2\right]^2}$$

$$g_i'(x) = \frac{-x + p_{i2}}{{p_{i3}}^3\sqrt{2\pi}} e^{\frac{-(x-p_{i2})^2}{2{p_{i3}}^2}}$$

where:

p_{ij} - j^{th} parameter (*j*=1..3) concerning i^{th} function, f_i'- derivative of i^{th} function, *w* - quantity of elemental functions, defined by an expert to 3 (*w*=3).

The i^{th} elemental Gaussian and Lorenz functions are described by $p_{ij},\ j = 1..3$ parameters, which are used to construct function f_p. The p_{i1} is a weight of the specific function in equation while remaining two parameters describe the function shape. The p_{i2} parameter defines the mean/location parameter while p_{i3} parameter defines the variation/scale parameter.

3.3. FINDING THE SOLUTION

To find the function approximation, an optimization task was defined which aims to minimize the approximation error between the real function f_r' and function approximation (f_p) modifying the p_{ij} parameters. The accuracy was measured using root mean squared error (RMSE). The error was calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{x=1}^{N}\left[f_r'(x) - f_p(x)\right]^2}, \tag{5}$$

where:

N - total number of the analyzed values.

To automatically find p_{ij} parameters, a genetic algorithm [1] and Conjugate Gradient Method [2] were implemented. The genetic algorithm was optimized for the given task. After preliminary research the initial population was set to 1000. The population above that value did not increase the algorithm efficiency. Each k^{th}, $k \in (1, 1000)$ individual describes candidate for f_p function. The p_{ij} parameters of k^{th} individual were defined as a P_k matrix:

$$P_k = \begin{bmatrix} p_{11} & \cdots & p_{31} \\ p_{12} & \ddots & p_{32} \\ p_{13} & \cdots & p_{33} \end{bmatrix}, \tag{6}$$

that takes initial values equally distributed within defined boundaries. Boundary values for P_k matrix were defined: p_{j1} parameters as the maximum signal value for a spectrum, $p_{j\ 2}$ parameters, which defines position of function centers, should not exceed the analyzed interval from -6 to 6 and p_{j3} parameters as the maximal range of approximated function (interval from 0 to 30).

Then, 100 epochs (calculated by minimizing RMSE) are run to find a rough estimation. The best solution from the previous generation is moved to the next one, while the remaining solutions undergo mutation and crossover. The mutation considers changing one of a $P_k$matrix elements by 10^{-2} of a parameter value range. The crossover is performed by exchanging the random elementary function between individuals (P_k). The probability of moving to the next epoch is calculated based on fitness function:

$$\min_{\mathrm{k}}\left(\sum_{\mathrm{x=1}}^{\mathrm{n}}\left[f_r(x) - f_p(x, P_k)\right]^2\right). \tag{7}$$

After 100 epochs, the 10 individuals (P_k) with the best match are further processed by Gradient Method to find a local minimum.

The final solution is selected according to eq. 5. The results of the method are presented in Fig. 5 for three functions $f_p = l_1' \circ l_2' \circ g_3'$ (two Lorenz and one Gauss) while Fig. 6 presents the elemental functions.

The final step of the method consists of reading parameters. For the EPR spectra, the parameters for the individual component lines were determined: g-factors, amplitudes (A), integral intensities (I), and line widths (ΔB_{pp}). Amplitude and integral intensity are dependent on paramagnetic centers concentration in the samples [9]. Line width reflects the magnetic properties of the samples [9]. The g-values are calculated from resonance condition according to the formula [9]:

$$g = \frac{hv}{\mu_B B_r} \tag{8}$$

where: h – Planck constant, ν – microwave frequency, μ_B – Bohr magneton, B_r – resonance magnetic field.

The achieved results were compared against the samples prepared by the expert. 95% of samples (100 test samples) were correctly approximated. In case of 5 samples the Gauss and Lorenz functions were swapped. This error is caused by the incomplete spectrum outside the spectra border. The obtained results from over 100 samples gave the average RMSE error equal to 0.55 against processed sample.

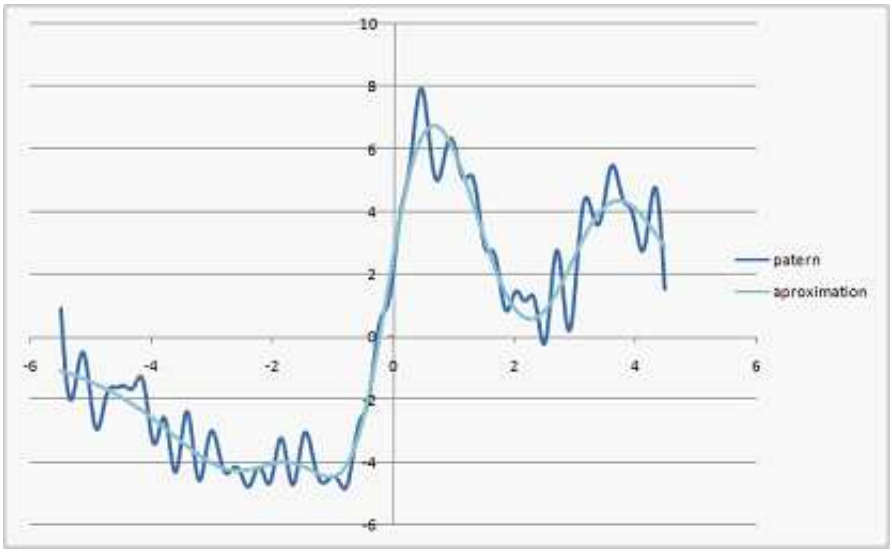

Fig. 5. The approximation of f_r function using f_p function.

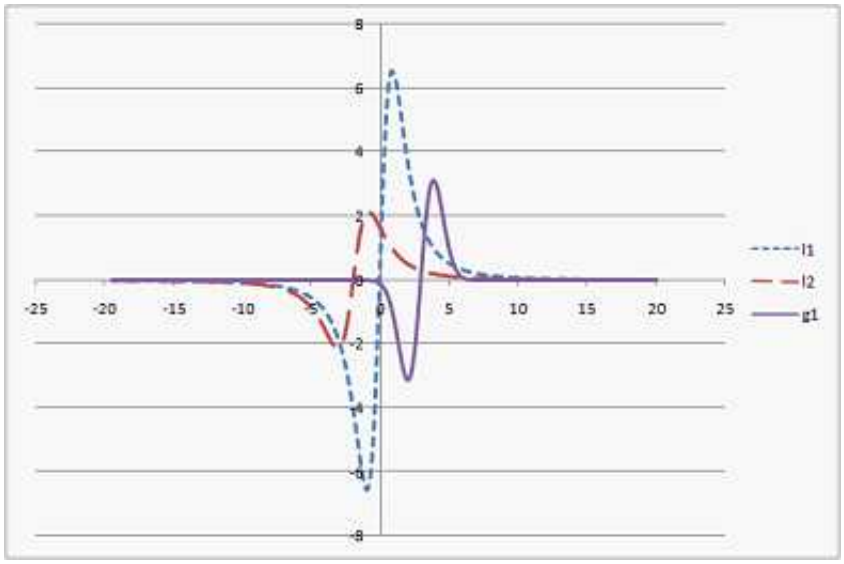

Fig. 6. The Gauss and Lorentz partial signals.

4. CONCLUSIONS

Electron paramagnetic resonance spectroscopy may be used to examine different groups of free radicals in biological matter. The experimental spectra were analyzed by the proposed algorithm using combination of six Gauss (G) and Lorentz (L) lines. The following combinations of theoretical (G, L) lines were tested: GG, LL, GL, GGG, LLL, GGL, and GLL. In most cases, when biological matter was concerned, the best results were achieved using GLL.

The proposed method proved to be efficient for 95% of researched samples and can be used for medicine studies. The original method using an expert is time-consuming and a research concerning thousands of spectra could not be performed within a short time period.

The next step is to improve efficiency of the procedure even further by approximation of the function outside the measured scope. Furthermore, the new heuristic method will be researched to decrease the computation complexity and improve the flexibility of the method by using a number of various elemental functions.

BIBLIOGRAPHY

[1] CHAN R., SUDHOFF S., An Evolutionary Computing Approach to Robust Design in the Presence of Uncertainties, IEEE Transactions on Evolutionary Computation, 2010, Vol. 14, No. 6, pp. 900 - 912.

[2] DONG J., JIAO B., CHEN L., A New Hybrid HS-DY Conjugate Gradient Method, Fourth International Joint Conference on Computational Sciences and Optimization (CSO), 2011, pp. 94 - 98.

[3] EATON G., EATON S., SALIKHOV K., Foundations of modern EPR, World Scientific, Singapore, 1998.

[4] RAMOS P., PILAWA B., KRZTOŃ A., LISZKA B., Free Radicals in the Thermally Sterilized Aminoglycoside Antibiotics, Pharmaceutica Analytica, 2012, Vol. 3, Issue 9, pp. 1-12.

[5] RZEPECKA-STOJKO A., PILAWA B., RAMOS P., STOJKO J., Antioxidative Properties of Bee Pollen Extracts Examined by EPR Spectroscopy, Journal of Apicultural Science, 2012, Vol. 56, No. 1, pp. 17-25.

[6] TAKAHAMA T., SAKAI S., Constrained optimization by applying the α constrained method to the nonlinear simplex method with mutations, IEEE Transactions on Evolutionary Computation, 2005, Vol. 9 , No. 5, pp. 437 - 451.

[7] TAN M., GUANGZHOU W., XU J., GUAN Z., ZHANG P., An Aggregation Approach to Short-Term Traffic Flow Prediction, IEEE Transactions on Intelligent Transportation Systems, 2009, Vol. 10 , No. 1, pp. 60 - 69.

[8] WANG D., Least squares-based recursive and iterative estimation for output error moving average systems using data filtering, Control Theory & Applications, 2011, Vol. 5, No. 14 , pp. 1648 - 1657.

[9] WANG S., INKOL R., RAJAN S., Patenaude, Franois, 2010, pp. 1424 - 1438.

[10] WERTZ J., BOLTON J., Electron Spin Resonance Theory and Practical Applications, New York, London, 1986.