# COMPARISON OF BASELINE ESTIMATION ALGORITHMS FOR CHROMATOGRAPHIC SIGNALS

Witold Ilewicz, Marcin Kowalczuk, Michał Niezabitowski, Dariusz Buchczik and Adam Gałuszka
Silesian University of Technology, Akademicka 2A Street, 44-100 Gliwice, Poland
{Witold.Ilewicz , Michal.Niezabitowski , Dariusz.Buchczik, Adam.Galuszka}@polsl.pl
Mar.Kowalczuk@gmail.com

***Abstract*— Baseline in signals is a relatively complicated problem in analysis of signals obtained in various analytical techniques such as chromatography and spectroscopy. In this article there are presented results of tests on four algorithms for a baseline estimation in chromatographic signals. Two of them are based on a polynomial fitting in a region of detected peaks. Another two algorithms, i.e. assymetric least squares method and iterative polynomial fitting with the automatic treshold, estimate the baseline under the whole signal and generally does not require peak detection. The tests of the algorithms are performed on both simulated and experimental signals.**

***Index Terms*— chromatographic signal, baseline removal**

## I. Introduction

FINDING and removing baseline is a very important procedure in many analytical techniques, for example in chromatography and spectroscopy, because the varying baseline can significantly influence on the measurement of peak areas. In chromatographic techniques, individual peaks observed in the chromatograms represent particular substances and the corresponding peak areas are proportional to the concentration of a given component. This is why, the differences in the estimated peak areas affect qualitative results of the analysis. Choice of the baseline correction algorithm is essential for correctly discriminating analyte signal from unresolved complex mixture signal [1], as it is a low-frequency signal and it cannot be estimated by the conventional frequency analysis. Universal algorithm for different types of spectroscopic data is presented in [2]. In [3] a smoothed spectrum is used for both baseline area recognition and modelling. In papers [4] - [6] using the asymmetric least squares method (LSM) for the baseline approximation is shown. Robust baseline estimation techniques are presented in [7], [8]. The attempts to automate the baseline correction are presented in [9] - [12]. Another method is presented in [13], where the polynomial fitting with the automatic threshold is introduced. Novel methods and algorithms are still proposed [14], [15].

In some cases, there is a problem of disturbances in the analytical signal having impulsive nature [16]. In this case, the baseline estimation method may be alternatively used to remove such the interference.

Baseline estimation is also applicable in other fields of technology. For example calibration process of the optical scanner mounted on the mobile vehicle requires the fitting plane determination, where the main goal of the procedure is to separate image features from the ground [17].

In the article four algorithms were choosen for tests. Two of them are relatively simple but another two are more complex. Comparison of them allows drawing relevant conclusions.

## II. Description of selected algorithms

### A. Approximation with first order polynomial (Alg.1)

Most basic algorithms for calculating the peak area find begining and ending of each peak and then draw a straight line between those two points. To determine start and end points of the peak we used method presented in [18]. This method does not take into consideration the baseline before and after the peak, so it is easy to implement but accuracy of it may be low.

### B. Approximation with second order polynomial (Alg. 2)

Second approach is bit more complicated. We can take few points before peak start point and few points after peak end and approximate the baseline with second order polynomial. Implementation of this method is harder but it should have more accurate result of calculations of the baseline.

### C. Asymetric LSM (Alg. 3)

Third method is able to completely eliminate the background in the whole signal. The algorithm described in [19] is an extension of idea first described in [20]. This method uses the asymmetric LSM and its goal is to smooth the signal. For purpose of the baseline calculation the objective function is modified by introducing weights [6]. With this approach the objective function is defined as:

$$Q = \sum_i v_i \left(y_i - z_i\right)^2 + \lambda \sum_i \left(\Delta^2 z_i\right)^2, \qquad (1)$$

where $y$ is the experimental signal, $z$ is the baseline approximation, $\lambda$ is a method parameter, and $\Delta$ is a differentiation operator. Weights $v$ are chosen in the asymmetric way:

$$\begin{array}{lll} v_i = p & \text{if} & y_i > z_i, \\ v_i = 1 - p & \text{if} & y_i < z_i, \end{array} \qquad (2)$$

where $p$ is number from 0 to 1. The equation is solved in an iterative way. For first approximation all weights are equal to 1 so the first baseline approximation can be calculated.

978-1-4799-8701-6/15/$31.00 ©2015 IEEE.

In consecutive iterations there is applied constant value of $p$, for example $p = 0.01$. The iterative procedure is repeated until the following stop criterion is fulfilled [13]:

$$\rho = \frac{||z_j - z_{j-1}||}{\|z_{j-1}\|} < \textbf{Stop threshold}, \tag{3}$$

where $z_j$ and $z_{j-1}$ are result of the baseline approximation in current and previous iteration.

*D. Baseline correction by improved iterative polynomial fitting with the automatic threshold (Alg. 4)*

This method was originally proposed for the estimation of the baseline in [13]. Baseline model can be written in a matrix form as

$$\mathbf{y} = \mathbf{X}\mathbf{a}, \tag{4}$$

where number of columns of the Vandermonde matrix $\mathbf{X}$ determines order of the polynomial to the baseline approximation. The values of the polynomial approximation to the baseline are

$$\widehat{\mathbf{y}} = \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T y}, \tag{5}$$

where $\widehat{\mathbf{y}}$ is an estimate of $\mathbf{y}$. Consequently, the algorithm can be written as follows:

1) set power $n$ of the polynomial;
2) create the polynomial fitting equation (4);
3) perform calculations of the polynomial fitting result $\widehat{\mathbf{y}}_k$ using eq. (5) in $k$ iteration;
4) make comparison of $\widehat{\mathbf{y}}_k$ and $\mathbf{y}_{k-1}$ in region of peaks. If $y_{k-1}(i) > \widehat{y}_k(i)$ then $y_k(i) = \widehat{y}_k\,(i)$;
5) if the stop criterion is satisfied, stop. In other case, take $\mathbf{y}_k$ as a new signal and go to the first step.

Stop criterion is satisfied when relative change in values of the polynomial parameters in consecutive iterations is smaller than the choosen stop value, for example 0.001.

## III. Tests and Comparison of Algorithms

Four algorithms (Alg. 1 - 4) were implemented using the Matlab environment. Algorithms were tested for many simulated and experimental signals. In the article representative results were shown for four simulated signals and one experimental signal.

Simulated signals were created as a combination of three components: three Gaussian peaks (first the narrowest one, last the widest one), normally distributed noise and low frequency baseline (Fig. 1).

Additionally as a test signal an experimental chromatographic signal was recorded using *Varian gas chromatograph type CP3800* with *FID* and *CP Sil 5CB* capillar column dedicated for alcohols. The analyzed mixture contained alcohols (from methanol do nonanol, big peaks) and was polluted with some other hydrocarbons (small peaks). The signal was sampled with sampling period 25 $ms$ and is shown in Fig. 4.

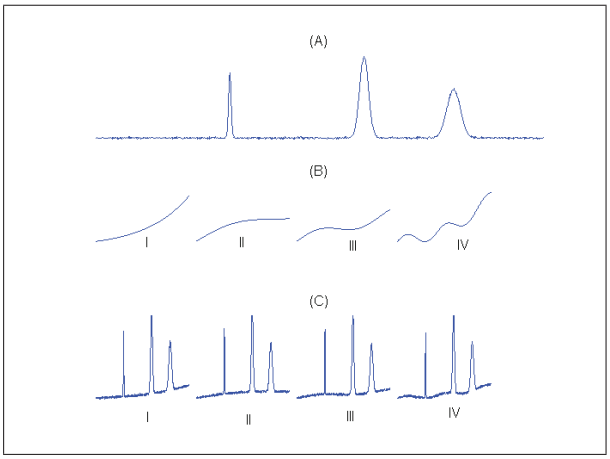


Fig. 1. Simulated test signals. (A) Three Gaussian peaks with noise. (B) Simulated baselines. (C) Simulated test signals I, II, III, IV as a sum of (A) and (B). (own source)

*A. Basic algorithms for estimation of the baseline under peak*

Results of the baseline estimation of signal I for Alg. 1 (Str. Line) and Alg. 2 (Pol. Fit.) and obtained peak areas are shown in Fig. 2, Fig. 3 and Table I.

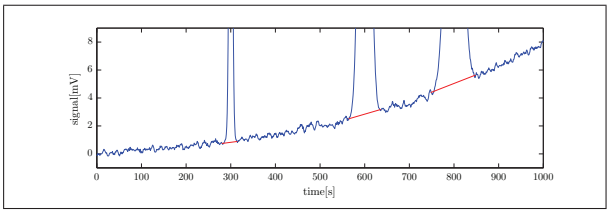


Fig. 2. Baseline under the peak as a straight line. (own source)

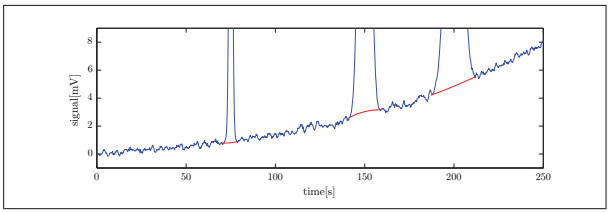


Fig. 3. Baseline under the peak as 2nd order polynomial. (own source)

Heading of Table I contains the numbers of peaks. In the second and third rows of the table, there are absolute and relative (in % of area of the whole signal) true peak areas. In the next rows there are presented results of calculation of the peak areas after estimation of the baselines under the peaks using Alg. 1 (Str. Line) and 2 (Pol. Fit.). For each peak there are calculated absolute and relative peak areas. $Diff(\%)$ row contains deviation of the relative peak area from its true value.

As it is shown in Table I better results were obtained for the algorithm 1 (straight line). The results of the baseline estimation and deviations of the peak areas from the true values in simulated signals II, III and IV for Alg. 1 and 2 are similar to the results for signal I, although in signals

TABLE I

PEAK AREAS IN SIMULATED SIGNAL I. (OWN SOURCE)

| No. | 1 | 2 | 3 |
|---|---|---|---|
| Original | 300.80 | 1253.30 | 1128.00 |
| % | 11.215 | 46.728 | 42.057 |
| Str. line | 300.08 | 1246.80 | 1105.40 |
| % | 11.31 | 47.01 | 41.68 |
| Diff(%) | **0.099** | **0.280** | **-0.379** |
| Pol. fit. | 302.00 | 1245.70 | 1089.00 |
| % | 11.45 | 47.24 | 41.30 |
| Diff(%) | 0.239 | 0.516 | -0.755 |

III and IV the baseline is slightly more complicated. Again better accuracy of the peak areas is obtained with the Alg. 1.

### *B. Baseline approximation with asymmetric LSM*

This algorithm allows to determine the baseline without knowing position of the peaks and does not require use of the peak finding algorithm. In [19] this method is described in detail and there are proposed few improvements and modifications of this algorithm that greatly reduce time and memory needed for calculations.

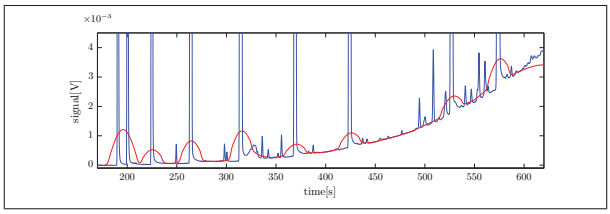


Fig. 4. Baseline approximation with unsufficient number of iterations. (own source)

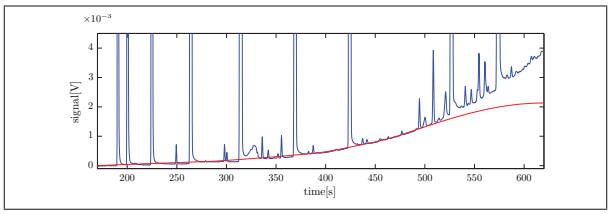


Fig. 5. Baseline approximation with too large number of iterations. (own source)

Proper setting of the number of iterations by the stop criterion is very important - results of the baseline estimation for too small and too large number of iterations is presented in Fig. 4 and Fig. 5. In order to obtain satisfactory results there must be found values of $\lambda$ and stop criterion $\rho$ which ensure proper approximation of the baseline. Unfortunately it is hard to find a strict dependence between those two variables. Generally with increase in $\lambda$ it can be observed bigger influence of the roughness coefficient at cost of getting lost of fit to the new approximation.

Exemplary calculation results of the baseline on simulated signals I and II are presented in Fig. 6 and Fig. 7. As one may see in Fig. 6, the approximated baselines deviates from the original one at the end of the signal.

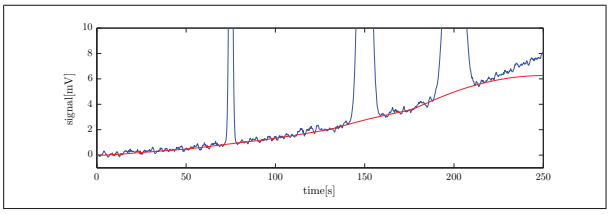


Fig. 6. Baseline approximation with $\lambda = 5$ and $\rho = 0.002$ (Obtained after 22 iterations and in 4.970 s.). (own source)

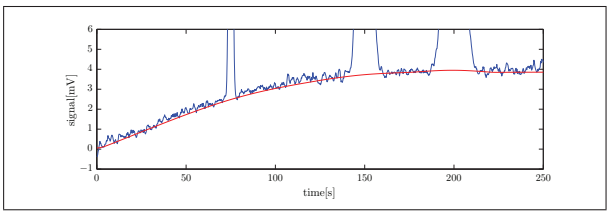


Fig. 7. Baseline approximation with $\lambda = 5$ and $\rho = 0.002$ (Obtained after 24 iterations and in 6.321 s.). (own source)

By increase in $\lambda$ we can reduce number of iterations, but if we set too small stop criterion $\rho$ it will start to move away from the signal along its whole length. Too large $\lambda$ will cause that approximation will be smoother and will not match the baseline. In Table II we may also see that results obtained with $\lambda = 5$ are the closest to the real one. The best results are bolded in the table.

TABLE II

PEAK AREAS IN SIMULATED SIGNAL I. (OWN SOURCE)

| No. | 1 | 2 | 3 |
|---|---|---|---|
| Original | 300.80 | 1253.30 | 1128.00 |
| % | 11.215 | 46.728 | 42.057 |
| $\lambda = 0.5$(%) | 11.39 | 46.88 | 41.73 |
| Diff(%) | 0.176 | 0.149 | -0.325 |
| $\lambda = 2$(%) | 11.39 | 46.94 | 41.68 |
| Diff(%) | 0.170 | 0.210 | -0.380 |
| $\lambda = 5$(%) | 11.34 | 46.88 | 41.78 |
| Diff(%) | **0.125** | **0.148** | **-0.273** |

TABLE III

PEAK AREAS IN SIMULATED SIGNAL II. (OWN SOURCE)

| No. | 1 | 2 | 3 |
|---|---|---|---|
| Original | 300.80 | 1253.30 | 1128.00 |
| % | 11.215 | 46.728 | 42.057 |
| $\lambda = 0.5$(%) | 11.38 | 46.85 | 41.77 |
| Diff(%) | 0.166 | **0.120** | -0.286 |
| $\lambda = 2$(%) | 11.35 | 46.92 | 41.73 |
| Diff(%) | 0.135 | 0.191 | -0.326 |
| $\lambda = 5$(%) | 11.33 | 46.88 | 41.79 |
| Diff(%) | **0.116** | 0.148 | **-0.263** |

Big advantage of the method is that locations of peaks in the chromatogram are needless. It means that searching for them we can leave for a further investigation of the chromatogram, what is a lot easier after removal of the baseline. As a drawback we can observe at the end of the signal some deviation of approximation from the signal.

### C. Baseline approximation with improved iterative polynomial fitting with the automatic threshold

The only variable which we can change to obtain different baseline is the degree of the fitting polynomial. Too small degree of the polynomial can cause baseline to be fitted in unsatisfactory way. Too big degree causes that the polynomial fits into peak as well as to the baseline. Effect of too small degree is presented in Fig. 8.

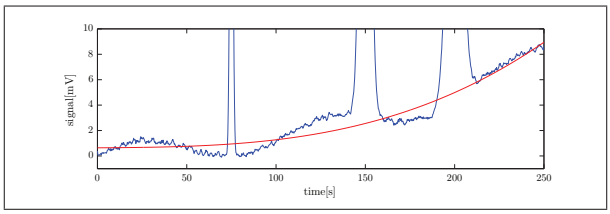


Fig. 8. Baseline approximation with polynomial fitting degree equal to 3. (own source)

As it can be seen in Fig. 8, small order of the polynomial could not handle such complicated baseline. Generally, results of the tests confirms that middle value of the polynomial order provides the best results. It requires more iterations than smaller values but in the end obtained results are better. Choosing too big value also results in some drawbacks. Too high order of the polynomial causes that the algorithm is trying to fit to the peaks instead of just to the baseline.

### D. Comparison of methods

Removal of the baseline is essential for the accuracy of quantitative analysis in chromatography. In Tables IV - VII there are presented best results obtained with all of the algorithms for each signal.

TABLE IV
COMPARISON OF RESULTS OBTAINED WITH USE OF DIFFERENT ALGORITHMS (SIGNAL I). (OWN SOURCE)

| Signal I | | | |
|---|---|---|---|
| No. peak | 1 | 2 | 3 |
| Original | 300.80 | 1253.30 | 1128.00 |
| % | 11.215 | 46.728 | 42.057 |
| Alg. 1 | 0.227 | 0.180 | -0.407 |
| Alg. 2 | 0.696 | 1.736 | -2.432 |
| Alg. 3 | 0.095 | 0.165 | -0.26 |
| Alg. 4 | **0.018** | **0.055** | **-0.072** |

TABLE V
COMPARISON OF RESULTS OBTAINED WITH USE OF DIFFERENT ALGORITHMS (SIGNAL II). (OWN SOURCE)

| Signal II | | | |
|---|---|---|---|
| No. peak | 1 | 2 | 3 |
| Original | 300.80 | 1253.30 | 1128.00 |
| % | 11.215 | 46.728 | 42.057 |
| Alg. 1 | 0.201 | 0.134 | -0.334 |
| Alg. 2 | 0.802 | 0.955 | -1.757 |
| Alg. 3 | 0.119 | **0.05** | -0.168 |
| Alg. 4 | **-0.02** | 0.107 | **-0.087** |

TABLE VI
COMPARISON OF RESULTS OBTAINED WITH USE OF DIFFERENT ALGORITHMS (SIGNAL III). (OWN SOURCE)

| Signal III | | | |
|---|---|---|---|
| No. peak | 1 | 2 | 3 |
| Original | 300.80 | 1253.30 | 1128.00 |
| % | 11.215 | 46.728 | 42.057 |
| Alg. 1 | 0.122 | 0.236 | -0.357 |
| Alg. 2 | 0.204 | 0.021 | **-0.183** |
| Alg. 3 | 0.339 | **-0.004** | -0.335 |
| Alg. 4 | **0.045** | 0.171 | -0.214 |

TABLE VII
COMPARISON OF RESULTS OBTAINED WITH USE OF DIFFERENT ALGORITHMS (SIGNAL IV). (OWN SOURCE)

| Signal IV | | | |
|---|---|---|---|
| No. peak | 1 | 2 | 3 |
| Original | 300.80 | 1253.30 | 1128.00 |
| % | 11.215 | 46.728 | 42.057 |
| Alg. 1 | 0.101 | 0.532 | -0.633 |
| Alg. 2 | 0.322 | **-0.165** | **-0.158** |
| Alg. 3 | -0.025 | 1 | -0.975 |
| Alg. 4 | **0.019** | 0.21 | -0.229 |

The results show that almost always the smallest errors of the peak area are obtained using Alg. 4. It can handle even very complicated baselines. Low memory consumption and short time of execution are advantages of this algorithm.

Asymmetric LSM for the simple signals also provides very accurate results. The problem of deviation from the baseline arises as the baseline is having more turns. We may see that for the last tested signal (signal IV) errors of peak areas are the biggest. That algorithm does not require any information about the peak placement in the signal so the peak finding algorithm is not necessary. Drawback of that algorithm is that it requires large amount of memory during the execution.

For simpler baselines Alg. 1 has small deviations from the real value but together with increase of complexity of the baseline differences become bigger. Alg. 2 acts in a opposite manner. For simple signals it fits to the baseline depending more on the noise than on the signal itself. As the baseline of the signals is getting more complicated Alg. 2 provides better results.

### E. Comparison of methods for the real signal

In case of the real signal we may observe that the shape of peaks differ from the Gaussian peak. As we can see they are non-symmetrical. This phenomenon is called peak tailing and appears frequently in chromatograms obtained in GC. In that case the algorithm for peak finding has great influence on look of the estimated baseline in the region of peaks. There are also in the real signal nine big peaks with number of small peaks, which may pose difficulties for the tested algorithms. Also with the real signal it is hardly possible to compare the accuracy of the methods because the true values of peak area are unknown. Areas of nine biggest peaks in the real signal, described in table as Oryginal in that case were calculated using STAR software working with a gas

chromatograph Varian CP3800. Results for Alg.1 and Alg. 2 are presented in Fig. 9, 10 and in Table VIII (first four peaks only).

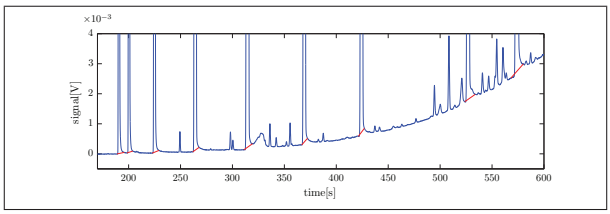


Fig. 9. Finding baseline under the peak with straight line. (own source)

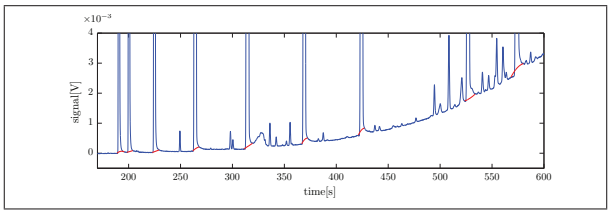


Fig. 10. Finding baseline under the peak with polynomial fitting. (own source)

TABLE VIII

PEAK AREAS IN REAL SIGNAL I. (OWN SOURCE)

| No. | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Original | 353767 | 426491 | 590911 | 971741 |
| % | 5.723 | 6.899 | 9.559 | 15.719 |
| Str. line | 14.144 | 17.050 | 23.624 | 38.863 |
| % | 5.724 | 6.900 | 9.560 | 15.727 |
| Diff(%) | **0.00117** | **0.00077** | 0.00143 | **0.00795** |
| Pol. fit. | 14.136 | 17.054 | 23.622 | 38.873 |
| % | 5.721 | 6.902 | 9.560 | 15.732 |
| Diff(%) | -0.00195 | 0.00252 | **0.00082** | 0.01231 |

Baseline approximation for the real signal with Alg. 3 (assymetric LSM) method are presented in Fig. 11 and in Table IX (first four peaks only).

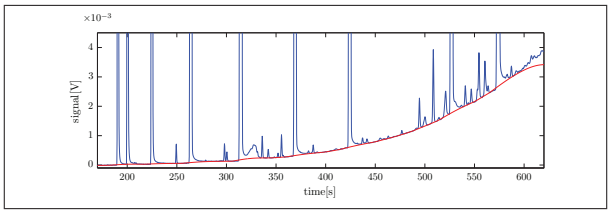


Fig. 11. Baseline approximation with $\lambda = 0.5$ and $\rho = 0.00035$ (Obtained after 81 iterations and in 54.130 s.). (own source)

Similarly to the simulated signals, we can observe that the approximated baseline deviates from the original signal. Different values of $\lambda$ were checked and it occurred in each case. Generally, the best results are for the smallest value of $\lambda$. Baseline for this value is visible in Fig. 11 and it has the smallest deviation - line runs closest to the real baseline (does not have positive deviations along signal).

Results for Alg. 4 (iterative polynomial fitting with automatic threshold) are presented in Fig. 12 and in Table X (first four peaks only).

TABLE IX

PEAK AREAS IN REAL SIGNAL I. (OWN SOURCE)

| No. | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Original | 353767 | 426491 | 590911 | 971741 |
| % | 5.723 | 6.899 | 9.559 | 15.719 |
| $\lambda = 0.5(\%)$ | 5.7205 | 6.898 | 9.5594 | 15.725 |
| Diff(%) | **-0.00219** | **-0.00111** | **0.00056** | **0.00569** |
| $\lambda = 2(\%)$ | 5.7182 | 6.8969 | 9.5601 | 15.727 |
| Diff(%) | -0.00449 | -0.00221 | 0.00126 | 0.00769 |
| $\lambda = 5(\%)$ | 5.7156 | 6.8941 | 9.5575 | 15.73 |
| Diff(%) | -0.00709 | -0.00501 | -0.00134 | 0.01069 |

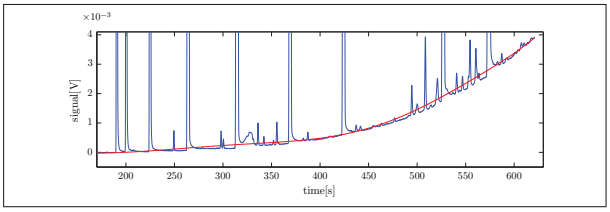


Fig. 12. Baseline approximation with polynomial fitting degree equal to 8. (own source)

Baseline approximation for the real signals is not so good as for the simulated ones. This is probably caused by irregularity of the baseline that depends on many factors. Other reason is that not all of the peaks are detected by used algorithm so the baseline is not replaced in those areas. It can be visible in Fig. 12 where in regions between 330 second and 370 or 500 and 580 we may see a lot of small peaks. Their amplitude is significantly smaller then the detected ones but have influence on the approximated baseline. Results in Table X confirm that tests on the simulated signals complies with results for the real signals. Approximation with middle degree of the polynomial fitting provides the best results.

## IV. CONCLUSIONS

From all of the tests performed for the purpose of this work it is hard to state which of the algorithms is the best for determination of the chromatograms baseline, as every numerical methods have drawbacks. Firstly Alg. 1 and 2, which allow only to find the baseline under peak but they are very simple. In some cases drawing only a straight line between beginning and end of the peak is satisfactory. However it may happen that start or end of peak will be detected with small accuracy and the routed line may differ from the original baseline. Alg. 2 turned out to be too sensitive to the noise and the results were strongly corrupted. Using more points of the real baseline in the polynomial approximation might improve results.

Other two algorithms have capability to determine the baseline along the whole signal. Depending only on chromatograms that were created in simulations one may say that the improved iterative polynomial fitting with the automatic threshold algorithm provides the most accurate results (Alg. 4). But unfortunately in the real chromatograms there occur many obstacles and the algorithm still needs some improvements to provide fully satisfactory results. Especially

TABLE X

PEAK AREAS IN REAL SIGNAL I. (OWN SOURCE)

| No. | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Original | 353767 | 426491 | 590911 | 971741 |
| % | 5.723 | 6.899 | 9.559 | 15.719 |
| $\lambda = 3(\%)$ | 5.7211 | 6.9012 | 9.5597 | 15.732 |
| Diff(%) | -0.00159 | **0.00209** | **0.00086** | 0.01269 |
| $\lambda = 8(\%)$ | 5.7261 | 6.9054 | 9.5603 | 15.727 |
| Diff(%) | 0.00341 | 0.00629 | 0.00146 | **0.00769** |
| $\lambda = 15(\%)$ | 5.7239 | 6.9036 | 9.5609 | 15.73 |
| Diff(%) | **0.00121** | 0.00449 | 0.00206 | 0.01069 |

important is to have well detected peaks along the whole signal.

Alg. 3 (assymetric LSM) does not require peak detection for estimation of the baseline too, but its accuracy is not satisfactory in case of more complex baselines. Moreover, the memory and time consumption for this algorithm is significantly larger than in the others. This algorithm also needs further improvements.

Performing many tests on the simulated and real signals leads to the conclusion that the baseline removal is very complicated topic in chromatography preprocessing. There exists many approaches and articles concerning this problem, but still there is wide range of possibilities in creating new solutions concerning estimation of background in chromatographic signal. As long as someone will not create perfect algorithm we may expect that new ideas will arise.

## V. ACKNOWLEDGEMENT

This work has been partially supported by Institute of Automatic Control BK grants no. BK-227/RAu1/2015 in the year 2015. The calculations were performed with the use of IT infrastructure of GeCONiI Upper Silesian Centre for Computational Science and Engineering (NCBiR grant no POIG.02.03.01-24-099/13).

## REFERENCES

[1] S. Samanipour, P. Dimitriou-Christidis, J. Gros, A. Grange and J. S. Arey, "Analyte quantification with comprehensive two-dimensional gas chromatography: Assessment of methods for baseline correction, peak delineation, and matrix effect elimination for real samples", *Journal of Chromatography A*, 1375 123-139 (2015).

[2] D.A. Barkauskas and D.M. Rocke, "A general-purpose baseline estimation algorithm for spectroscopic data", *Analytica Chimica Acta*, 657, 191-197 (2010).

[3] S. Golotvin and A.Williams, "Improved baseline recognition and modeling of FT NMR spectra", *Journal of Magnetic Resonance*, 146(1), 122-125 (2000).

[4] J. Peng, S. Peng, A. Jiang, J. Wei, Ch. Li and J. Tan, "Asymmetric least squares for multiple spectra baseline correction", *Analytica Chimica Acta*, 683(1) 63-68 (2010).

[5] P.G. Stevenson, X.A. Conlan, N.W. Barnett, "Evaluation of the asymmetric least squares baseline algorithm through the accuracy of statistical peak moments", *Journal of Chromatography A*, 1284, 107-111 (2013).

[6] B. Walczak and M. Daszykowski, "Use and abuse of chemometrics in chromatography", *Trac-Trends in Analytical Chemistry*, 25(11), 1081-1096 (2006).

[7] A.F. Ruckstuhl, M.P. Jacobson, R.W. Field and J.A. Dodd, "Baseline subtraction using robust local regression estimation", *Journal of Quantitative Spectroscopy and Radiative Transfer*, 68, 179-93 (2001).

[8] X. Ning, I. W. Selesnick and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (BEADS)", *Chemometrics and Intelligent Laboratory Systems*, 139, 156–167 (2014).

[9] Ł. Górski, F. Ciepiela and M. Jakubowska, "Automatic baseline correction in voltammetry", *Electrochimica Acta*, 136, 195-203 (2014).

[10] P. Yaroshchyk and J.E. Eberhardt, "Automatic correction of continuum background in Laser-induced Breakdown Spectroscopy using a model-free algorithm", *Spectrochimica Acta Part B-Atomic Spectroscopy*, 99, 138-149 (2014).

[11] F. Xian, Y.E. Corilo, Ch.L. Hendrickson and A.G. Marshall, "Baseline correction of absorption-mode Fourier transform ion cyclotron resonance mass spectra", *International Journal of Mass Spectrometry*, 325-327, 67-72 (2012).

[12] Q. Bao, J. Feng, F. Chen, W. Mao, Z. Liu, K. Liu and Ch. Liu, "A new automatic baseline correction method based on iterative method", *Journal of Magnetic Resonance*, 218, 35-43 (2012).

[13] F. Gan, G. Ruan and J. Mo, "Baseline correction by improved iterative polynomial fitting with automatic threshold", *Chemometrics and Intelligent Laboratory Systems*, 82(1-2), 59-65 (2006).

[14] X. Liu, Z. Zhang, Y. Liang, P.F. M. Sousa, Y. Yun and L. Yuc, "Baseline correction of high resolution spectral profile data based on exponential smoothing", *Chemometrics and Intelligent Laboratory Systems*, 139, 97–108 (2014).

[15] A. Fasano and V. Villani, "Baseline wander removal for bioelectrical signals by quadratic variation reduction", *Signal Processing*, 99, 48-57 (2014).

[16] J. Wiora, "Improvement of measurement results based on scattered data in cases where averaging is ineffective", *Sensors and Actuators B: Chemical*, 201, 475–481 (2014).

[17] S. Budzan, "Fusion of visual and range images for object extraction", *Computer Vision and Graphics, Lecture Notes in Computer Science*, 8671, 108-115 (2014).

[18] T. O'Haver, "Peak finding and measurement", http://terpconnect.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm

[19] P.H.C. Eilers, "A perfect smoother", Analytical Chemistry, 75(14), 3631-3636 (2003).

[20] E.T. Whittaker, "On a new method of graduation", *Proceedings of the Edinburgh Mathematical Society*, 41(1), 63-75 (1922).