Resolving of Overlapping Asymmetrical Chromatographic Peaks by Using Wavelet-Transform and Gram-Charlier Peak Model

A V Bochkarev¹

¹Department of information and measuring technologies, SamSTU, Molodogvardeyskaya 244, Samara 443100, Russian Federation

E-mail: bochkarevpubliciit@gmail.com

Abstract. The paper describes a method for resolving overlapping asymmetric peaks that make up a chromatogram. The presented method uses the Gram-Charlier model in the form of the first three terms of the Gram-Charlier series as a basis. Using the wavelet transform, the parameters of this model are determined, which is used to describe a single or overlapping chromatographic peak. Hermitian wavelets of the first four orders are used in the computation of the wavelet transform. To speed up the computation of multiple wavelet transforms, the possibility of coding a signal using the Chebyshev-Hermite functions is considered in order to further restore the set of wavelet transforms simultaneously. According to the presented method, the parameters of the peaks are determined by analytical expressions without using the numerical approximation of the chromatogram by the peak model, which avoids the disadvantages of the numerical approach. The resulting method is used to resolve overlapping asymmetric peaks. The advantage of the method over others is shown by calculating the area of each of the resolved peaks. Keywords: overlapping peak resolving, wavelet transform, chromatography, peak area calculation, chromatographic data processing

1. Introduction

Analytical instrumentation is an intensively developing branch of measurement technology designed to study the composition and properties of substances. The development of this branch goes along the path of improving the characteristics of analytical devices and introducing new methods for analyzing measurement results. Among the tasks of analytical instrumentation functional transformation, filtering, detection and resolving overlapping chromatographic peaks, the elimination of the distorting effect of the device's spread function can be allocated [1].

The improvement of computational processing technologies for measuring information for signals of complex configuration is associated with the development of applied mathematical methods that underlie these technologies, with the introduction of new software and algorithmic data processing tools [2]. Thus, the actual problem is the development of theoretical and applied approaches that allow synthesizing compact and fast computational algorithms for estimating the desired signal, aimed at realizing all the potential possibilities of analytical devices.

Wavelet analysis is one of the most powerful and flexible means of research and digital processing of signals. For chromatographic signals it can be used for denoising, compression, spectral analysis [3],...
detecting apexes positions of overlapping peaks in low resolution chromatogram, resolving overlapping symmetrical peaks [4] and others.

The wavelet transform is implemented by integrating of a signal multiplied with a shifted and scaled wavelet, which makes calculations laborious if wavelets of complex shape are used or calculating of wavelet transform with many different wavelets are needed. To simplify the calculations, the signal can be encoded in the basis of the Chebyshev-Hermite functions (also named Hermite functions in some references) and the wavelet transforms can be restored using the algorithm described in the paper [5]. This encoding is based on decomposing the signal into a series by Chebyshev-Hermite functions [5-7]. These functions are based on Chebyshev-Hermite polynomials (also named Hermite polynomials in some references) and both of them are widely used in different branches of computational science: for design of filter banks [8], to determine optical parameters of laser beam [9], to key factors recognition of biometrical images [10] and other areas [11–13]. Simultaneous localization of these functions in both frequency and time space makes the method sufficiently resistant to instrumental errors [14–15]. In addition, encoding-decoding in the Chebyshev-Hermite function basis has a smoothing property [8].

At the moment, when processing signals from analytical instruments, there is no unified approach to the resolve overlapping chromatographic peaks, which is further complicated by the presence of asymmetry [16], and therefore the issue of developing new computational algorithms aimed at solving this problem is urgent. This paper described the method for resolving overlapping asymmetric chromatographic peaks by calculating its wavelet transforms of various orders.

2. Theoretical basis

2.1. Asymmetrical peak model selection

The first task that needs to be solved when developing a method for resolving overlapping chromatographic peaks is the choice of a model for these peaks. For simplicity, symmetric Gaussian function is often chosen as a model, and because of its symmetry the methods developed on its basis can't provide high accuracy of processing real chromatograms consisting of asymmetric peaks.

The most well-known models were considered: exponential modified Gaussian function [17], generalized exponential function [18], Fraser-Suzuki model [19], lognormal distribution function [20], Gram-Charlier series [21]. All these models except the last can be used in signal processing only by using numerical methods, such as the nonlinear least squares method. Such methods are more labor intensive and their accuracy deteriorates with decreasing SNR, increasing the number of overlapping peaks, and with incorrectly chosen initial conditions. It is not possible to analytically determine the parameters of these models, but for the Gram-Charlier model, the parameters of the peaks can be determined analytically through the wavelet transform without the described disadvantages [4], therefore this model was chosen for the developed method.

2.2. Gram-Charlier series

Gram-Charlier series are given by expression [21]:

$$F(x) = A \cdot \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \sum_{k=3}^{N} \frac{a_k}{k!} \cdot He_m \left( \frac{x-\mu}{\sigma} \right)$$

where \( He_m \left( \frac{x-\mu}{\sigma} \right) \) is the Hermite polynomial [22].

For practical use only the first three terms of (1) are usually taken:

$$F(x) = A \cdot \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \left[ 1 + \frac{\gamma_1}{6} He_3 \left( \frac{x-\mu}{\sigma} \right) + \frac{\gamma_2}{24} He_4 \left( \frac{x-\mu}{\sigma} \right) \right]$$

(2)
This expression has as parameters the amplitude $A$, a shift along the time axis $\mu$, asymmetry $\gamma_1$ (apex displacement within the localization region), and kurtosis $\gamma_2$ (apex sharpening). In contrast to the Gaussian function in model (2) at $\gamma_1 \neq 0$ the position of the apex is not equal to $\mu$ and at $\gamma_1 \neq 0$ or at $\gamma_2 \neq 0$ the apex height is not equal to $A$.

2.3. Wavelet transform with Hermitian wavelets

As it is well known wavelet transform of signal $f(x)$ is determined by the expression:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x) \cdot \psi\left(\frac{x-b}{a}\right) dx,$$

where $b$ is shift coefficient, $a$ is scale coefficient, $\psi\left(\frac{x-b}{a}\right)$ is a used wavelet.

There is a known method for resolving overlapping symmetric chromatographic peaks given by the Gaussian function, built on the wavelet transform with Hermitian wavelets $[4]$. Since model (2) at $\gamma_1 = 0$ and $\gamma_2 = 0$ turns into a Gaussian function, the mentioned method can be considered as a special use case of the Gram-Charlier model, and for asymmetric peaks it is necessary to find a general solution.

Hermitian wavelets are a series of wavelets obtained from the product of an exponent and the Hermite polynomial of order $m$:

$$\psi_m\left(\frac{x-b}{a}\right) = -\exp\left[-\frac{(x-b)^2}{2a^2}\right] \text{He}_m\left(\frac{x-b}{a}\right).$$

3. Resolving asymmetrical peaks by wavelet transform

3.1. Method for resolving overlapping asymmetrical chromatographic peaks by using wavelet-transform and Gram-Charlier peak model

By applying a wavelet transform with Hermitian wavelets to the Gram-Charlier model at $m = 1, 2, 3, 4$ and $b = \mu$ the following expressions can be obtained:

$$W_1(a,\mu) = \sqrt{\frac{\pi}{2}} a^{3/2} \sigma^4 \gamma_1 A,$$

$$W_2(a,\mu) = \sqrt{\frac{\pi}{2}} a^{5/2} A \sigma \left(8a^4 + 16a^2 \sigma^2 + (8 + 5\gamma_2) \sigma^4\right),$$

$$W_3(a,\mu) = \sqrt{\frac{\pi}{2}} 5a^{7/2} \gamma_1 A \sigma^4,$$

$$W_4(a,\mu) = \sqrt{\frac{\pi}{2}} a^{9/2} A \sigma \left(24a^4 + 48a^2 \sigma^2 + (24 + 35\gamma_2) \sigma^4\right),$$

$$4(\sigma^2)^{7/2}.$$
where \( W_m(a, \mu) \) is wavelet transform with \( \psi_n \left( \frac{x-b}{a} \right) \) Hermitian wavelet. To simplify the calculation of four orders of the wavelet transform at once, the previously mentioned method of reconstructing wavelet transforms from the signal coding coefficients in the basis of the Chebyshev-Hermite functions are used [5].

The wavelet transform is calculated at \( b = \mu \), since in this case the resulting expressions are greatly simplified. From expressions (3)-(6), estimates of the parameters of the Gram-Charlier model can be obtained:

\[
\hat{\sigma} = a \left( -\frac{5}{W_2(a, \mu) / W_4(a, \mu) - 1} \right)^{1/2},
\]

\[
\hat{\gamma}_2 = \frac{32 + 96 \mu}{5 + 168 \mu},
\]

\[
\hat{\gamma}_1 = \frac{a^{3/2} \sigma^3}{2W_4(a, \mu)(a^2 + \sigma^2)^{3/2}},
\]

\[
\hat{\lambda} = \sqrt{\frac{a^{5/2}}{2} \left( \frac{8a^4 + 16a^2 \sigma^2 + (8 + 5\hat{\gamma}_2) \sigma^4}{4(a^2 + \sigma^2)^{7/2}} \right)}.
\]

To calculate the parameter estimates from expressions (7) - (10), it is necessary to accurately determine \( \mu \), which cannot be done analytically by wavelet or other transform. To solve this problem, it is necessary to take into account two circumstances: 1) the local maxima of even wavelet transforms with Hermitian wavelets coincides with the position of the maxima of the overlapped peaks [4]; 2) for \( \gamma_1 > 0 \), the apex is shifted to the left of \( \mu \), and for \( \gamma_1 < 0 \), it shifts to the right [18]. Consequently, it is possible to determine \( \mu \) numerically, assuming the position of the \( W_2(a, b) \) maximum as the initial value and using the sum of the squares of the differences between the values of the chromatogram and the peak calculated by (2) with the substitution of (7) - (10) and the current estimate of \( \mu \) as recovery accuracy. In this paper the dichotomy method is used to search optimal \( \mu \).

The presented method of resolving overlapped asymmetrical chromatographic peaks can be described by the following sequence of actions applied to signal \( f(x) \):

- calculate the (3)-(6) for the fragment in which the last peak is localized (from the minimum between the tops of the last two peaks to the end of the signal);
- using the dichotomy method, find \( \mu \) by assessing the quality of restoration by the fragment selected at the previous step;
- restore the last peak \( p(x) \) according to model (2) and estimates (7) - (10), calculate the difference signal \( f'(x) = f(x) - p(x) \);
- repeat the previous steps replacing \( f(x) \) with \( f'(x) \), respectively changing the localization area.

The last step is repeated for all peaks remaining in the signal after subtracting the last one. If the reconstruction result is unsatisfactory, the described procedures should be repeated (until the desired accuracy is obtained) for individual difference signals \( f'(x) \) - in the case of two overlapped peaks, this
signal always is $f'(x) = f(x) - p(x)$ from above, if there are more than two peaks, $f'(x)$ is calculated for each peak as the difference between $f(x)$ and reconstructed earlier nearest peaks (one to the left and one to the right).

It is suggested to start with the last peak, since the studied real signals have an asymmetry with a gamma greater than zero, which corresponds to the mu position to the right of the peak maximum. In such signals, the last peak is usually the least superimposed on other peaks in the region where mu is located (from the top to the end). Since the wavelet transforms (3) - (6) are calculated for a single signal at the point mu, the absence of overlaps with other peaks in this region allows one to determine the peak parameters with high accuracy using (7) - (10). Since the method is implemented iteratively, this feature can be ignored, but more iterations may be required to obtain optimal results.

3.2. Resolving peaks by presented method

To test the performance of the presented method, a synthesized signal was investigated, consisting of the sum of three overlapped peaks, the parameters of each of which are reliably known. To resolve the peaks, the signal was processed according to the presented method and by wavelet method with Gaussian peak model described in [4]. The comparison results are shown in Figure 1.

![Figure 1](image)

**Figure 1.** The figure shows the result of the separation of peaks by the presented method (a) and the wavelet method with Gaussian peak model (b)

The final goal of any method of chromatographic signals processing is to reliably determine the informative parameters - the area and position of the apex of each peak. To assess the reliability of the resolving results by the proposed method, the area was calculated for the obtained peaks. For comparison, the area of the peaks was also calculated using the wavelet method with Gaussian peak model and classical methods – perpendicular drop and triangulation [23]. Table 1 shows the relative errors of the calculated areas. The position of the apexes when using all methods was determined without error, therefore, it is not indicated in the table.

| Resolving method                  | Area calculation error, % |
|-----------------------------------|----------------------------|
|                                   | Peak 1 | Peak 2 | Peak 3 |
| Perpendicular drop                | 7,25   | 4,63   | 10,24  |
| Triangulation                     | 5,81   | 3,1    | 6,42   |
| Wavelet with Gaussian peak model  | 12,3   | 12,6   | 31,63  |
| Presented method                  | 0,94   | 2,24   | 2,16   |

4. Result discussion and conclusion

As can be seen from Figure 1, approximation by the Gaussian function does not allow to reliably reconstruct the shape of asymmetric peaks. According to the results presented in Table 1, the presented
method has the smallest error among all those considered. These results show the possibility of using the presented method when processing chromatograms containing asymmetric peaks. The advantage of the method is the increased accuracy of processing asymmetric chromatograms, as well as the elimination of the disadvantages of numerical methods for approximating overlapping peaks.

5. References

[1] Mattrey F, Makarov A, Regalado E, Bernardoni F, Figus M, Hicks M, Zheng J, Wang L, Schafer W, Antonucci V et al. 2017 TrAC Trends in Analytical Chemistry 95 2

[2] Zolotov Y A 2017 Journal of Chromatography and Separation Techniques 8:2 1-3

[3] Medhat M E 2015 Journal of Applied and Computational Mathematics 4:4 2-6

[4] Sayfullin R T and Naumov A A 2016 Bulletin of the Samara Scientific Center of the Russian Academy of Sciences 1-2 412-415

[5] Sayfullin R T and Bochkarev A V 2019 Vestnik of the Samara State Technical University. Series: technical sciences 2(62) 113-124

[6] Sayfullin R T and Bochkarev A V 2019 Vestnik of the Samara State Technical University. Series: technical sciences 1(61) 68-81

[7] Sayfullin R T and Bochkarev A V 2019 Proceedings of the XI Scientific Conference “Mathematical modeling and boundary value problems” 2 137-139

[8] Balakin D A and Shytkov V V 2014 Journal of Radio Electronics 9 1-15

[9] Kotljar V V, Kovaljov A A and Porfir’ev A P 2019 Computer optics №3 Vol.43 356-367

[10] Paveleva E A and Krylov A S 2010 Informatics and its applications №1 Vol.4 79-82

[11] Bayakovsky Y M, Zhirkov A O, Korchagin D N, Krylov A S and Lukin A S 2001 Neural network analysis and comparison of time-frequency vectors based on the short-term spectral representation and adaptive Hermite transform Preprint IPM 087

[12] Estudillo-Romero A and Escalante-Ramirez B 2008 CIARP 2008 5197 86–93

[13] Mamaev N V, Lukin A S, Yurin D V, Glazkova M A and Sinitsin V E 2013 Proceedings of the 23rd International Conference on Computer Graphics and Vision GraphiCon’2013 254-258

[14] Martens J B 1990 IEEE Trans. Signal Process. 38:9 1595-1606

[15] Martens J B 1990 IEEE Trans. Signal Process. 38:9 1607-1618

[16] Dubrovkin J 2015 IJETCAS 11(1) 1-8

[17] Grushka E 1972 Anal. Chem. 44 1733-1738

[18] Vaidya R A and Hester R D 1984 J. Chromatogr. 287 231-244

[19] Fraser R D B and Suzuki E 1969 Shapes Anal. Chem. 41 37-39

[20] Siano D B 1972 J. Chem. Educ. 49(11) 755

[21] Olivé J and Grimalt J O 1991 Analytica Chimica Acta 249:2 337-348

[22] Abramowitz M and Stegun I A 1972 Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Applied Mathematics Series 55. Tenth Printing (New York: United States Department of Commerce, National Bureau of Standards) 775

[23] Vandeginste B G M and Rutan S C 1998 Data Handling in Science and Technology 21. Data Analysis and Signal Processing in Chromatography (Amsterdam: Elsevier) 197-199

Acknowledgments

The reported study was funded by RFBR according to the research project № 20-31-90014.