ELSEVIER

Journal of Chromatography A, 930 (2001) 53–60

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Application of a new mathematical function for describing chromatographic peaks

T.L. Pap*, Zs. Pápai

Institute of Analytical Chemistry, University of Veszprém, P.O. Box 158, H-8201 Veszprém, Hungary

Received 3 April 2001; received in revised form 27 July 2001; accepted 27 July 2001

Abstract

Curve fitting seems to be one of the best methods for the evaluation of chromatographic signals. As it is known, in this case mathematical function is fitted to digitized measured points. The most important task is to find the best mathematical function, which corresponds perfectly to the peak shape, and then to determine the parameters of the equation using a computerized least-squares method of approximation. In this work, a new mathematical function was sought for with the purpose of describing different chromatographic signals and it was fitted to the digitized measured points. The fitted curve is suitable for a quick evaluation of chromatographic information, noise filtering and correction of baseline drift. The fitting of gas chromatographic and high-performance liquid chromatographic signals were completed. The mathematical function, the generated chromatographic curves, the application of the function for describing real signals and the fitting process will be demonstrated in this study.

Keywords: Curve fitting; Peak asymmetry; Signal processing

1. Introduction

The computerized chromatogram evaluation offers the possibility of peak shape analysis and it is not limited to the calculation of retention time and peak area. Peak shape analysis helps not only the qualitative and quantitative analysis but we can draw conclusions about the operation of the chromatographic system and change the operating parameters. The peak shape is characterized by the respective statistical moments, the peak excess and skew calculated from them, the asymmetry factor calculated from peak widths at various peak heights and standard deviation of the peak [1,2].

*Corresponding author. Fax: +36-88-421-869.
E-mail address: pap@almos.vein.hu (T.L. Pap).

1.1. Mathematical description of a single peak

The above mentioned characteristics are used in many cases as parameters of the mathematical functions describing peak shapes. The first attempts were simulated by the Gauss function and its simply modified form as published by Baundisch [3]:

$$y = h_{\max} \exp\left\{\frac{(t - t_R)^2}{\sigma t}\right\} \quad (1)$$

where $h_{\max}$ and t_R are the maximum peak height and retention time, respectively, t is the time and σ is the standard deviation. The density function of the Poisson distribution was also used for describing chromatographic peak shapes; good results with its application were obtained especially by Degen [4]. He fitted the Poisson function to the digital peaks of

0021-9673/01/$ – see front matter
PII: S0021-9673(01)01163-3

benzene, naphtalene and anthracene, which were separated by liquid chromatography using a LiChrosorb RP-8 column and an acetonitrile–water mixture as the mobile phase. The relative difference was 3–5%.

These functions describe well the symmetrical chromatographic peaks but they are unsatisfactory for asymmetrical peaks. There are a lot of methods for describing skewed chromatographic peaks. One of the most popular models used in the literature is the exponentially modified Gaussian function (EMG), which is the result of the convolution of a Gaussian function and an exponential decay function:

$$y(t) = h(t)*f(t) \tag{2}$$

where $y(t)$ is the convoluted function, which describes the chromatographic peak, $h(t)$ is the function describing a pure Gaussian peak and $f(t)$ is the exponential decay function:

$$h(t) = h_{\max} \exp\left\{ -\frac{(t - t_R)^2}{2\sigma^2} \right\} \tag{3}$$

$$f(t) = \frac{1}{\tau} \exp\{ -t/\tau \} \tag{4}$$

Here the peak parameters are the same as in Eq. (1), σ is the peak width and τ is a constant that quantifies the decay time of the system.

The EMG function was and is still used nowadays by several authors for fitting chromatographic peaks. Anderson [5], Foley [6], Davis [7] and Hanggi [8] have examined theoretically the shape of the EMG function, the effect of baseline errors, the errors created by fitting and the possibilities of its application. Foley also published a BASIC program of the EMG function calculations [6]. Using the EMG function every transition function from Gauss curve ($\tau = 0$) to the ''clean'' exponential function ($\sigma = 0$) is also suitable for fitting to the measured digital signals. The τ/σ ratio is an asymmetry factor that characterizes the peak shape deviation from the ideal symmetrical peak.

In spite of the wide range of application, the EMG function has also some disadvantages: the τ and σ parameters can be determined only with fitting, and the starting parameters of the fitting sometimes are badly estimated; the parameters don't give information about the skew and excess; at small values of τ, in the region of 0, the calculation may become uncertain. Since the parameters of the EMG model are not directly related to the properties of the real peaks, and their evaluation requires the use of approximation, the practical application of the model in the simulation of asymmetrical peaks and optimization of the resolution in a chromatographic separation is somewhat problematic.

A family of models is proposed for the description of skewed chromatographic peaks based on the modification of the standard deviation of a pure Gaussian peak. Torres-Lapasió [9] and his co-workers replaced the constant value of the standard deviation of the Gaussian peak with a polynomial function. Thus skewed peaks were fitted with good accuracy including those showing a large asymmetry, with a deformation either to the right or to the left. Another case for describing asymmetrical peak shapes is the function resulted from the product of a Gauss function and a Hermite polynomial.

New modified forms of the Gauss function were used by Grubner [10] and Li [11] where the Gaussian is approximated by two-step functions.

1.2. Resolution of overlapped peaks

In chromatographic practice, the separation of two or more components in spite of the change of chromatographic circumstances (temperature, column length, column load and flow velocity, etc.) is often incomplete and more or less overlapping peaks appear on the chromatogram.

Geometrical methods such as perpendicular drop at the valley, triangular correction and tangent skimming of the overlapped tailing peaks have been most widely used. These methods were examined in detail by Woerlee [12] in his paper. The error of the resolution caused by the perpendicular drop method at the valley was examined by Novak [13] and Kishimoto [14] using different tailing ratio and peak shape parameters. According to their establishment the error depends considerably on the asymmetry and the difference of peak heights of the overlapped chromatographic peaks. Mori [15,16] worked out a calculation method for the accurate determination of overlapped chromatographic peak heights in the case

of two or three overlapped peaks. He determined single peak heights solving mathematical equations after determination of the characteristic points of the overlapped peaks.

Curve fitting is often used for the resolution of overlapped peaks. In that case one can apply equations to the description of single peaks but it must certainly to be completed by finding the adequate number of peaks.

The Gauss function was used by Metzger [17], Braswell [18], Rosenbaum [19] and Thomas [20] for the resolution of gas chromatographic peaks using curve fitting methods. Grimalt [21] tested several mathematical distribution functions for chromatographic peak resolutions such as the log-normal function, the Gamma, the Weibull and the modified Gauss function. He found different optimal functions for the resolution of different asymmetrical peaks and determined the accuracy of the resolution using the least square residuals and the relative error of the area under the curve. D'Allura [22] completed the resolution of liquid chromatographic peaks by computer. According to his algorithm, he calculated the parameters of the mathematical functions of the overlapped peaks from the single chromatographic peaks of the overlapping components injected in different concentrations using curve fitting.

Crilly [23–25] applied the Jansson deconvolution method for the resolution of gas chromatographic peaks. This method is an iterative nonlinear deconvolution technique to sharpen the overlapped peaks. For the computerized iteration, it is necessary to know only the impulse reply function and the peak height maximum. During the iteration, the area of the real overlapped peaks corresponds to the sum of peak areas obtained from deconvolution.

Using factor analysis, Maeder [26] and Lacey [27] resolved overlapped chromatographic peaks. As preliminary knowledge they had at disposition of the infrared spectrums of single components.

A good and substantial summary of curve resolution methods was given by Fritsch [28] and Felinger [29]. They presented the characteristics of some methods, the a priori information necessary for the methods, the limits of the application and the employment of the obtained results in tabular form.

The EMG model is also used for the resolution of overlapped peaks in the recent literature [30–32].

2. Experimental

2.1. Measuring conditions

During the experiments, signals were collected using two separation methods.

2.1.1. Gas chromatography

The analyses were carried out on a GCHF-18-3-1 gas chromatograph, which was connected to an INTEL 8085 microprocessor. This system collected the signals and stored the digital chromatograms. The microprocessor performed some calculations like a simple filtering with a moving window, peak searching and determination of retention time, then the signals were sent to an IBM compatible computer of high capacity.

The experimental conditions were as follows: stationary phase = OV-101, mesh range = 80/100, column = 1000 × 4.0 mm, temperature = 110°C, detector = heat conductivity, mobile phase = hydrogen, flow-velocity = 10 ml/min.

2.1.2. Liquid chromatography

The measurements were carried out with a Merck-Hitachi liquid chromatograph and an UV detector. The signals were sent through an A/D converter to an IBM compatible computer, then were evaluated with the NELSON program.

The experimental conditions were as follows: stationary phase = LiChrosorb RP-18, column = 250 × 3.0 mm, temperature = 25°C, detector = UV (280 nm), mobile phase = $CH_3OH–H_2O$ (60:40, v/v) + 0.001 *M* $HClO_4$, flow-velocity = 0.6 ml/min, sample concentration = 0.0025 *M*, injection volume = 20 μl.

3. Results and discussion

3.1. Mathematical function for describing chromatographic peak shapes

One especially difficult but effective method of modern signal processing is the approach to fitting a mathematical function to the analytical signal. It is important to find the best mathematical function that corresponds perfectly to the signal shape. The chro-

matographic peaks can be very different due to the physical–chemical processes in the column, the heterogenity of the stationary phase surphase, the heterogeneity of the column packing, column overload, extra-column or instrumental effects, therefore, it is impossible to establish a common equation. The equation of the function can be always selected to the chromatographic peak. In the literature generally one can find functions, which are suitable for describing only symmetrical or slightly asymmetrical peaks. In the chromatographic practice we often encounter strongly skewed peaks, whose fittings cannot be solved with the above mentioned equations. In this work a new mathematical function was looked for, which is suitable for describing both symmetrical and asymmetrical chromatographic peak shapes and was fitted to the digitized measuring points. The fitted curve is suitable for a quick acquisition of chromatographic information, noise filtering and correction of baseline drift.

The examined mathematical function is as follows:

$$f(t)=\begin{cases}0, & \text{if } t<M-\dfrac{D(4-a^2)}{2a}\\ H\exp\left\{\left(\dfrac{4}{a^2}-1\right)\cdot\left[\ln\left(1+\dfrac{2a(t-M)}{D(4-a^2)}\right)-\dfrac{2a(t-M)}{D(4-a^2)}\right]\right\}\end{cases} \quad (5)$$

where M is the peak maximum (s), H is the peak height ($H>0$), a is an asymmetry factor ($0<a<2$) and D is the standard deviation ($D>0$) (s).

The baseline of the chromatogram can be simulated by adding a linear (or quadratic) equation to Eq. (5).

The applicability of mathematical functions should be decided by examining simulated peak shapes. The peak shape calculated by using Eq. (5) is influenced first of all by the asymmetry factor (a). The asymmetry factor as function variable should take values between 0 and 2, so it is suitable to analyze simulated peak shapes in this domain at various peak widths (D).

In Figs. 1 – 2 a simulated symmetrical ($a=10^{-6}$) and a strongly asymmetrical ($a=1.999999$) peak with large peak widths ($D=0.1$) are illustrated.

In the next two figures (Figs. 3–4), the simulated shapes of narrow (needle) peaks are shown. Comparing the asymmetrical peaks of different widths, one can conclude that they show similar characteristics.

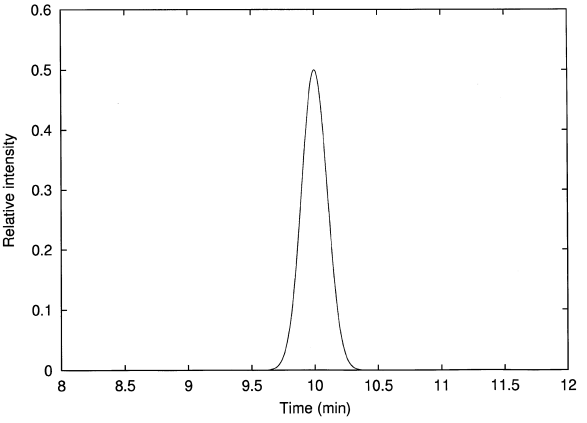

Fig. 1. Simulated symmetrical peak ($a=10^{-6}$) by Eq. (5).

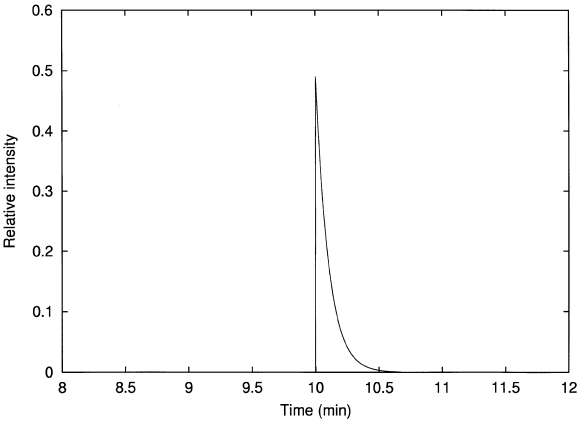

Fig. 2. Simulated asymmetrical peak shape ($a=1.999999$) by Eq. (5).

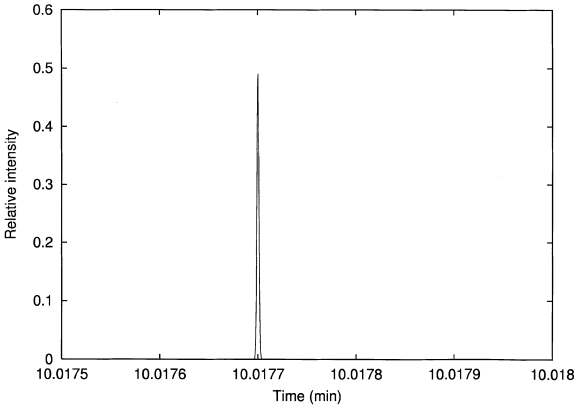

Fig. 3. Simulated symmetrical peak shape ($a=10^{-6}$; $D=10^{-6}$) by Eq. (5).

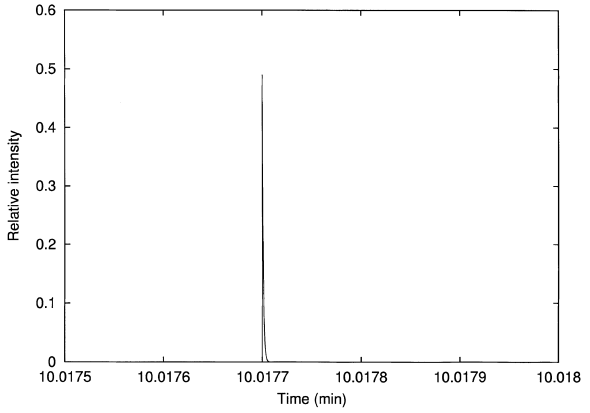

Fig. 4. Simulated asymmetrical peak shape ($a=1.999999$; $D=10^{-6}$) by Eq. (5).

The asymmetrical peaks have a prompt ascending and an exponential descending part; naturally using a medium asymmetry factor one should simulate peak shapes with smaller asymmetry.

In next figure (Fig. 5) the variety of peak shapes of different asymmetry factors is observable. Increasing the value of asymmetry factor (a) the peak shape becomes more and more asymmetrical and this asymmetry is already observable at a value of $a=0.5$. The value $a=1.5$ results in a sharply ascending and a slowly descending tailing peak. The further increase of the asymmetry factor gives a nearly exponential function.

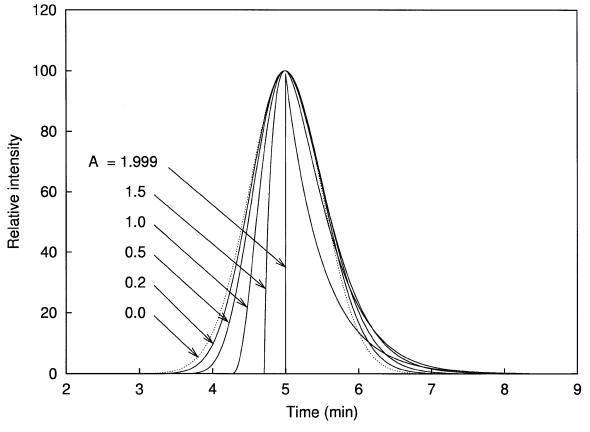

Fig. 5. The variety of peak shapes in function of asymmetry factor; dotted line=symmetrical curve ($a=0$); solid line=asymmetrical peak shapes of $a=0.2$; 0.5; 1.0; 1.5; 1.999. ($M=5.0$ min; $H=100.0$; $D=0.5$).

The main advantage of the present function is that it can be integrated in a closed form. Thus, the slow numerical integration can be avoided in case of peak area calculation:

$$\int f(t)\,\mathrm{d}t = \frac{\Gamma\left(\frac{4}{a^2}\right)HaD}{2\exp\left\{\left(\frac{4}{a^2}-1\right)\cdot\left[\ln\left(\frac{4}{a^2}-1\right)-1\right]\right\}} \qquad (6)$$

where the $\Gamma(z)$ mathematical function is defined by the Euler integral [33]:

$$\Gamma(z)=\int_0^\infty t^{z-1}\,\mathrm{e}^{-t}\,\mathrm{d}t \qquad (7)$$

There are important relations between the parameters and the moments for the chromatographic evaluation, which is another advantage of the function:

$$m_1 = M + \frac{aD}{2} \quad \text{the first statistical moment} \qquad (8)$$

$$\mu_2 = D^2 \quad \text{the second central moment} \qquad (9)$$

$$\mu_3 = aD^3 \quad \text{the third central moment} \qquad (10)$$

$$\mu_4 = 3D^2 + \frac{3a^2D^4}{2} \quad \text{the fourth central moment} \qquad (11)$$

$$S = \frac{\mu_4}{D^4} - 3 \quad \text{skewness} \qquad (12)$$

$$E = \frac{\mu_3}{D^3} = a \quad \text{excess} \qquad (13)$$

3.2. Application of the function for describing single peaks

The presented mathematical function is well suitable for describing real chromatographic peaks. First of all it can fit relatively slowly descending peaks, therefore wider gas and liquid chromatographic peak shapes. Next we will illustrate some particular examples for function application in chromatography (Figs. 6 and 7).

The function is also suitable for fitting both

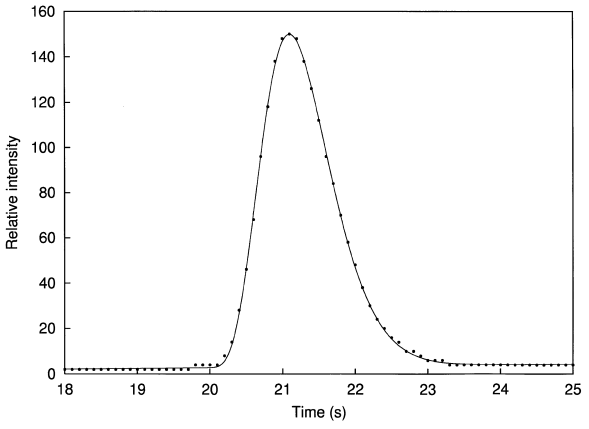

Fig. 6. Fitted curve of a gas chromatographic peak (cyclohexane). (dotted line = measured data; solid line = fitted curve; $M = 21.099$ s; $H = 146.77$; $a = 0.7294$; $D = 0.5113$; the fitted parameters of the linear baseline: $w = 0.3204$; $b = -3.619$).

symmetrical and sharply ascending, slowly descending tailing peak shapes.

3.3. Resolution of overlapped peaks with curve fitting

The resolution of overlapped peaks with curve fitting can be correct only if the initial parameter values are available as preliminary information. In case of symmetrical peaks, it is sufficient to know the peak maximum positions and the approximate values of the peak heights from the digital overlapped chromatogram as a priori information, and to start from these values. If the overlapped single peaks are not symmetrical curves, the previous information means the relative exact knowledge of the single component peak shape (moments, asymmetry factor, skewness). In case of overlapped peaks, the number of the parameters increases as the number of fitted peaks is increased. In a relatively simple case when the aim is to resolve two overlapped peaks using Eq. (5) 8+3 parameters for the resolution are needed. The three extra parameters are the number of correction parameters of the possible baseline (quadratic equation).

For curve fitting, the least-squares method was used, the change of the parameters always occurred by the steepest descend method (Fig. 8). The value of M and H from the function parameters can be determined exactly, while the determination of a and D is more doubtful.

The next figure shows the difference between the simulated overlapped curve fitted by Eq. (5) and the measured digitized data (Fig. 9). The differences between the calculated function and the measured data are the smallest in the neighborhood of the peak maximum, and they are the largest where the signal takes off from and returns to the baseline.

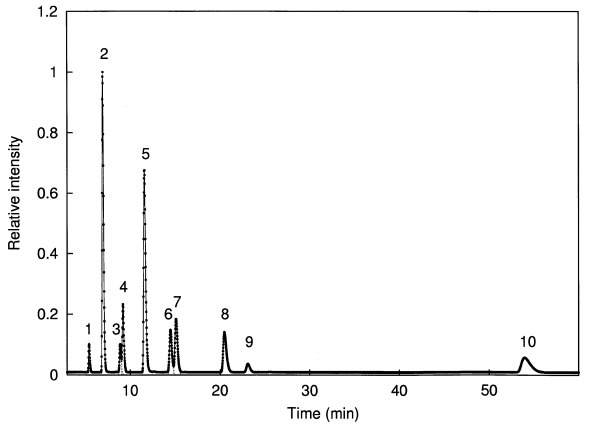

Fig. 7. Chromatogram of the resolution of phenol and different phenol derivatives using HPLC. The compounds of the peaks are the following: (1) phenol, (2) 4-nitrophenol, (3) 2-methylphenol, (4) 2-chlorophenol, (5) 2-nitrophenol, (6) 4-ethylphenol, (7) 2,4-dimethylphenol, (8) 2,3-dichlorophenol, (9) 2,3,6-trimethylphenol, (10) 2,3,4-trichlorophenol.

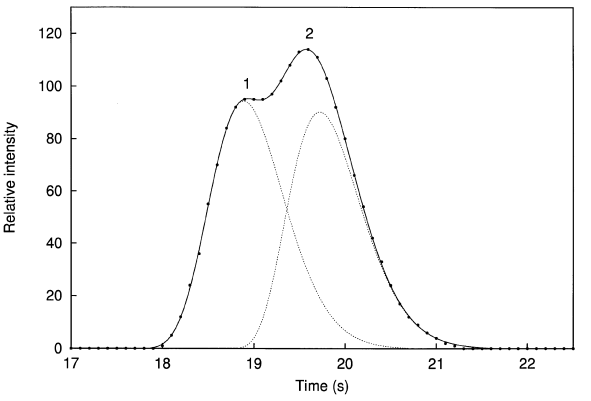

Fig. 8. Fitted curve of gas chromatographic peaks: (1) *n*-penthane; (2) *n*-hexane (dot = measured data; solid line = fitted sum; dotted line = single peak; $M_1 = 18.889$ s; $H_1 = 94.387$; $a_1 = 0.4873$; $D_1 = 0.4179$; $M_2 = 19.724$ s; $H_2 = 90.262$; $a_2 = 0.6497$; $D_2 = 0.4168$).

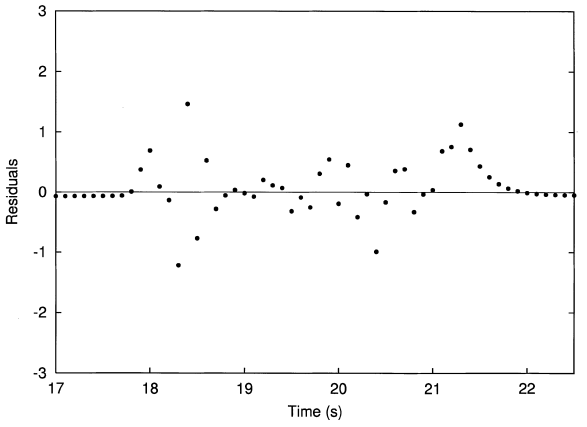

Fig. 9. The differences between the calculated and measured digitized data of the curve fitting (%) (Fig. 8).

The knowledge of the function $f(t)$ parameters makes the peak area calculation possible, which is more accurate, faster and simpler than by the trapezoidal method. As the results of the two different calculation methods of integration are compared, the differences are less than 0.01%. The peak area can be calculated with Eq. (6).

The practical application of function $f(t)$ is very advantageous because the fitting to the chromatographic digitized data is precise and using the parameters, which are the results of the fitting, the peak area calculation is fast and accurate.

Fig. 7 shows an example for the resolution, in which digitized data are fitted using Eq. (5) in high-performance liquid chromatography. The results of the chromatogram evaluation and the data obtained from the curve fitting are summarized in Table 1.

The function $f(t)$ is also well applicable for fitting overlapped peaks, and it is suitable to calculate the total area, integrating $f(t)$ using the fitting parameters.

4. Conclusion

The presented mathematical function is suitable with good accuracy for describing chromatographic (both gas and liquid) peaks. The differences between the points of the fitted curve and the measured digitized data is the smallest by using the method for the least mean square.

The function may be used with good results for the resolution of overlapped chromatographic peaks and determination of baseline drift; thus, correction of baseline and noise filtering can be perfectly realized.

The main advantages of the function (5):

The function can be integrated, thus the slow numerical integration can be avoided in case of peak area calculation, which requires considerably longer calculating time in case of many digitized points. The closed form of the integral proposed by us always is faster and more exact than the approaching numerical integral value.

Table 1
Fitting the peaks of phenol and phenol derivatives resolved with HPLC (Fig. 7)

No.	Component	$t_R = M$ (min)	H	$\sigma = D$ (min)	a	A (min)
1	Phenol	5.422	0.089	0.0647	0.775	0.0136
2	4-Nitrophenol	6.946	0.979	0.0819	0.815	0.1868
3	2-Methylphenol	8.881	0.089	0.0912	0.701	0.0193
4	2-Chlorophenol	9.191	0.218	0.0942	0.693	0.0490
5	2-Nitrophenol	11.59	0.655	0.1151	0.880	0.1733
6	4-Ethylphenol	14.48	0.136	0.1317	0.646	0.0430
7	2,4-Dimethylphenol	15.13	0.172	0.1368	0.500	0.0574
8	2,3-Dichlorophenol	20.49	0.131	0.1895	0.774	0.0574
9	2,3,6-Trimethylphenol	23.09	0.027	0.1890	0.532	0.0125
10	2,3,4-Trichlorophenol	54.00	0.048	0.5372	0.949	0.0583

Measuring conditions: LiChrosorb RP-18, 250 mm×3.0 mm column, 25°C, UV 280 nm detector, CH_3OH–H_2O (60/40, v/v)+0.001 M $HCLO_4$, 0.6 ml/min flow-velocity, 0.25×10^{-2} M concentration.

The calculated integral has a smaller error of peak area calculation because it is calculated from the continuous function.

The fitted function is suitable in the overwhelming majority of chromatographic peaks, which are symmetrical or slightly asymmetrical ones.

By means of the function, one can produce quite variable slightly asymmetrical peak shapes, which are shown in Fig. 5.

There is a direct relationship between the fitted parameters (M, H, a, D) of the function and the chromatographic characteristics (the first statistical; the second, third, fourth central moments; the skewness and excess), which are presented by Eqs. (8)–(13).

To determine the parameters of the proposed mathematical function, curve fitting is not necessary. All of the function's parameters can be read directly from the chromatographic peaks or can be calculated using Eqs. (8)–(10).

It is simpler if the asymmetry parameter is constituted only from one value and not from two like in other models. Using factor τ/σ, the number of the calculated parameter increases, which is not negligible during curve fitting, especially when the chromatogram includes many peaks.

This article with the enumeration of the advantages and disadvantages does not desire to range the models. Every model has it advantages and disadvantages. Our opinion is that one cannot establish a general model for describing chromatographic peaks because the mathematical functions represent only one kind of peak shape, while the real chromatograms have endless number of types as the result of very complex physico–chemical processes in the column. The chromatographer's duty is to find and apply the best model, which corresponds perfectly to the chromatographic peak.

References

[1] J. Foley, J.G. Dorsey, Anal. Chem. 55 (1983) 730.
[2] W.W. Yau, S.W. Rementer, J.M. Boyajian, J.J. De Stefano, J.F. Graff, K.B. Lim, J.J. Kirkland, J. Chromatogr. 630 (1993) 69.
[3] J. Baudisch, H.D. Papendick, V. Schlöder, Chromatographia 3 (1970) 469.
[4] J. Degen, R. Look, R. Wernicke, Labor Praxis 9 (1985) 317.
[5] D.J. Anderson, R.R. Walters, J. Chrom. Sci. 22 (1984) 353.
[6] J.P. Foley, J.G. Dorsey, J. Chrom. Sci. 22 (1984) 40.
[7] J.M. Davis, J.C. Giddings, J. Chrom. 289 (1984) 277.
[8] D. Hanggi, Anal. Chem. 57 (1985) 2394.
[9] J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Alvarez Coque, Anal. Chem. 69 (1997) 3822.
[10] O. Grubner, Anal. Chem. 43 (1971) 1934.
[11] J. Li, Anal Chem. 69 (1997) 4452.
[12] E.F.G. Woerlee, J.C. Mol, J. Chrom. Sci. 18 (1980) 258.
[13] J. Novak, K. Petrovic, S. Wicar, J. Chrom. 55 (1971) 221.
[14] K. Kishimoto, H. Miyauchi, S. Musha, J. Chrom. Sci. 10 (1972) 220.
[15] Y. Mori, J. Chrom. 66 (1972) 9.
[16] Y. Mori, J. Chrom. 70 (1972) 31.
[17] H.-D. Metzger, Chromatographia 3 (1970) 64.
[18] E. Braswell, Anal. Biochem. 44 (1971) 58.
[19] M. Rosenbaum, V. Hancil, R. Komers, J. Chrom. 246 (1982) 1.
[20] H. Thomas, Anal. Biochem. 120 (1982) 101.
[21] J. Grimalt, H. Iturriaga, Anal. Chim. Acta 139 (1982) 155.
[22] N.J. D'Allura, R.S. Juvet Jr., J. Chrom. 239 (1982) 439.
[23] P.B. Crilly, J. Chemometrics 1 (1987) 79.
[24] P.B. Crilly, J. Chemometrics 1 (1987) 175.
[25] P.B. Crilly, J. Res. Nat. Bur. Stand. 93 (1988) 413.
[26] M. Maeder, A.D. Zuberbuehler, Anal. Chim. Acta 181 (1986) 287.
[27] R.F. Lacey, Anal. Chem. 58 (1986) 1404.
[28] S. Fritsch, G. Veress, Magy. Kém. Lapja 35 (1980) 178.
[29] A. Felinger, Data Analysis and Signal Processing in Chromatography, Elsevier, Amsterdam, 1998.
[30] C.P. Cai, N.S. Wu, Chromatographia 31 (1991) 595.
[31] M.S. Jeansonne, J.P. Foley, J. Chrom. Sci. 29 (1991) 258.
[32] D.Y. Youn, S.J. Yun, K.-H. Jung, J. Chrom. 591 (1992) 19.
[33] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in Fortran, The Art of Scientific Computing, 2nd ed., Cambridge University Press, 1992.