Сорбционные и хроматографические процессы. 2024. Т. 24, № 6. С. 944-955.
Sorbtsionnye i khromatograficheskie protsessy. 2024. Vol. 24, No 6. pp. 944-955.

ISSN 1680-0613

Original article

Fast calculation of exponentially modified functions

Yuri A. Kalambet[1,2]✉, Yuri P. Kozmin[1]

[1]Ampersand Ltd., Moscow, Russian Federation, kalambet@ampersand.ru✉
[2]Frumkin Institute of Physical Chemistry and Electrochemistry, Moscow, Russia

Abstract. There is a great need for modelling experimental data represented as asymmetric peaks, for example those detected by chromatography. One of the most important models for these peaks is Exponentially Modified Gaussian (EMG) function. In statistics, EMG distribution describes the probability density of the sum or difference of two random variables, one of which has a normal distribution, and the other has an exponential distribution. Drawbacks of this distribution are I) rather complicated set of formulas used for its computation and II) lack of formulas that can be used for calculation of the density of the sum of one normal and more than one exponentially distributed variables. In this study a general method for rapidly calculating exponentially modified functions using the exponentially weighted moving average (EWMA) algorithm has been investigated. The algorithm allows very simple and fast way to calculate an approximate estimate of the exponential modification of Gaussian or any other function with required precision, as well as to make a double, triple, and more exponential modifications. New formulas relating the time constant of the exponential modification τ and the coefficient of the EWMA algorithm α are proposed and accuracy of these formulas depending on experimental data rate are evaluated.

Keywords: asymmetric peak, exponential modification, exponential moving average, exponentially weighted moving average, exponentially modified Gaussian, EMG, EWMA, EMA.

Acknowledgments: the work is partially (50%) supported within the framework of the project of the Ministry of Education and Science for the development of domestic scientific instrumentation, topic code: "Chromatograph"

For citation: Kalambet Yu.A., Kozmin Yu.P. Fast calculation of exponentially modified functions. *Sorbtsionnye i khromatograficheskie protsessy*. 2024. 24(6): 944-955. https://doi.org/10.17308/sorpchrom.2024.24/12571

ОРИГИНАЛЬНЫЕ СТАТЬИ

Научная статья
УДК 543
doi: 10.17308/sorpchrom.2024.24/12571

Быстрое вычисление экспоненциально модифицированных функций

Юрий Анатольевич Каламбет[1,2]✉, Юрий Петрович Козьмин[1]

[1]ООО «Амперсенд», Москва, Россия, kalambet@ampersand.ru✉
[2]Институт физической химии и электрохимии им. А.Н. Фрумкина РАН, Москва, Россия

Аннотация. Существует насущная потребность в моделировании экспериментальных данных, представленных в виде асимметричных пиков, например, детектируемых с помощью хроматографии. Одной из наиболее важных моделей для этих пиков является экспоненциально модифицированная гауссова функция (ЭМГ). В статистике ЭМГ-распределение описывает плотность вероятностей суммы или разности двух случайных величин, одна из которых имеет нормальное распределение, а вторая – экспоненциальное. Расчёт этого распределения затруднён по причине: i) достаточно сложного набора формул, используемых для его вычисления и II) отсутствие формул, которые можно использовать для вычисления плотности суммы одной нормальной и более чем одной экспоненциально распределенной случайной величины. В данном исследовании был исследован общий метод быстрого вычисления экспоненциально модифицированных функций с использованием алгоритма экспоненциально взвешенной скользящей средней (EWMA). Алгоритм позволяет очень простым и быстрым способом вычислить приблизительную оценку экспоненциальной модификации гауссовской или любой другой функции с

требуемой точностью, а также вычислить функции с многократной экспоненциальной модификацией. Предложены новые формулы, связывающие постоянную времени экспоненциальной модификации τ и коэффициент α алгоритма EWMA, а также оценена погрешность этих формул в зависимости от числа измерений в экспериментальных данных.

Ключевые слова: асимметричный пик; экспоненциальная модификация; экспоненциально взвешенная скользящая средняя; экспоненциальная скользящая средняя; экспоненциально модифицированная гауссиана; ЭМГ; EWMA; EMA

Благодарности: работа частично (50%) поддержана в рамках проекта Министерства образования и науки по развитию отечественного научного приборостроения, код темы: «Хроматограф».

Для цитирования: Каламбет Ю.А., Козьмин Ю.П. Быстрое вычисление экспоненциально модифицированных функций // *Сорбционные и хроматографические процессы*. 2024. Т. 24, № 6. С. 944-955. https://doi.org/10.17308/sorpchrom.2024.24/12571

Introduction

The purpose of this work is to help researchers modeling experimental data in the construction of models that take into account the exponential modification of the simulated functions. Such modeling faces difficulties, since there are no explicit analytical expressions that allow calculating the exponentially modified form for most functions, with the happy exception of EMG [1-7]. In statistics, EMG describes the probability density of the sum or difference of two variables, one of which has a normal distribution, and the other has an exponential distribution. Even in the case of EMG, calculating a function over the entire range of its parameters is not an easy task due to computational problems [4].

We propose to solve the problem of modeling exponential modification using its digital analogue – EWMA. To solve the modeling problem, we needed to find formulas for the relationship between the parameters of continuous and discrete exponential distributions that work at a small data rate of experimental points, when the time constant of the exponential modification is comparable to the step of the discrete representation of the function.

Exponential Modification. One of the most often reasons of asymmetry in nature are processes of decomposition. Mathematical formulation (or relaxation) of process can be described by exponent. This process accompanies many other processes, and when two processes co-exist and one of them is a decay, the resulting shape of the curve, describing some property of combined process, is a convolution of two functions, one describing original process, and another describing decay process. The result of convolution is called exponentially modified function. In the case, when the curve, describing original process is symmetric, after convilution it becomes asymmetric due to decay. Simple example of the relaxation process in electronics is RC filtering, used to suppress noise. Our paper is devoted to the easy ways of computation of digital tables describing exponentially modified functions with required accuracy.

Exponential modification is implemented by an operation on the functions of a real variable called a convolution (or smoothing). Namely, a convolution of two functions f(x) and g(x) is a function

$$F(x) = (f \cdot g)(x) = \int_0^\infty f(y)\, g(x-y)dy \quad (1)$$

A convolution is called an exponential modification if one of the functions is a falling branch of the exponent (exponential distribution):

$$g(x) = \begin{cases} \frac{1}{\tau} e^{-x/\tau}, x \geq 0 \\ 0,\ x < 0 \end{cases} \quad (2)$$

Convolutions of this kind are the basis of Laplace transform. Laplace transform can be used to prove, that the exponential distribution g(x) is a solution of the differential equation

$$(f \cdot g)(x) + \tau \frac{d(f*g)(x)}{dx} = f(x). \quad (3)$$

This equation performs deconvolution (calculation of original function f(x) from the convolution F(x)=(f·g)(x)); it has a consequence that the maximum of the exponentially modified peak F(x) is always on the

graph of the non-modified peak f(x), since at the maximum dF(x)/dx=0.

Exponentially modified Gaussian. While interpretation of experimental data in the field of chromatography there is a great need to interpret peak patterns of mass spectra, chromatographic elution profiles, etc. There are many functions that simulate the shape of experimental peaks [8,9]. The most widely used function is the one, in which exponential modification of the Gaussian signal is involved (EMG) [1-7]. The physical model behind this shape in chromatography can be represented as an ideal chromatographic column yielding gaussian peak, and a mixing chamber at the outlet of the column, simulating the imperfection of the separation system. Gaussian peak shape can be represented by

$$f(t) = H_G \cdot exp(-(\mu_G - t)^2/2\sigma_G^2) \quad (4)$$

where H_G is the peak height, μ_G is the apex position, σ_G is the standard deviation. EMG is the result of an exponential modification of Gaussian. For computation purposes, EMG can be written in the form of several formulas [4]:

$$T = \frac{t-\mu_G}{\sigma_G};\ S = \frac{\sigma_G}{\tau};\ z = \frac{1}{\sqrt{2}}(S - T) \quad (5)$$

$$F(t) = H_G S\sqrt{\frac{\pi}{2}} \cdot e^{\left(\frac{S^2}{2} - TS\right)} \cdot erfc(z) \quad z > 0 \ (6)$$

$$F(t) = H_G \cdot e^{\frac{-T^2}{2}} \cdot S \cdot \sqrt{\frac{\pi}{2}} \cdot erfcx(z), \quad z \le 0 \ (7)$$

$$F(t) \approx \frac{H_G \cdot e^{\frac{-T^2}{2}}}{\left(1 - \frac{T}{S}\right)}, \quad z > 6.71 \cdot 10^7 \quad (8)$$

where T is the reduced time, τ is the time constant of the modifying exponent, S is the dimensionless value, inverse of the reduced τ; erfcx() is a scaled complementary error function [4,10]. The formula for calculating the EMG value is chosen so that there is no overflow of the computer representation of numbers; depending on the parameter *t*, different points of the same peak can be calculated using different formulas [4].

If more than one relaxation processes distort the shape of the peak, i.e. the signal is generated by the sum of one normally and several exponentially distributed random variables, it will be described by multiply exponentially modified Gaussian (mEMG) [11,12]. mEMG also describes a more general distribution of the sum of several normally distributed and several exponentially distributed random variables, since the sum of several normally distributed random variables is described by a normal distribution. The set of formulas 5-8 used to calculate EMG[4,13] can be extended to carry out double, triple, etc. exponential modification [11], although the formulas of mEMG become much more complex and new areas of instability of computation appear in them.

Exponentially weighted moving average. The discrete analogue of convolution is a weighted moving average in which integration turns into summation:

$$Y_k = \sum_{i=0}^{\infty} w_i y_{k-i} \quad (9)$$

where the capital letter Y_k indicates the smoothed value at the position with the index k, w_i is the weight of the point with the index (k-i), y_{k-i} is the non-smoothed ("raw") value of the signal. Usually, the sum of the weights equals one.

The exponentially weighted moving average (EWMA) can be represented as a weighted average (formula 9) with weights having a geometric distribution:

$$w_i = \alpha\,(1-\alpha)^i$$

0>α>1, which is a discrete analogue of the exponential distribution (formula 2).

EWMA can be recursively calculated as the weighted sum of the last measured and last smoothed measurements:

$$Y_0 = y_0;\ Y_k = \alpha y_k + (1-\alpha)Y_{k-1}, \quad (10)$$

Such EWMA calculation requires very low number of simple computer operations (summations and multiplications), proportional to the number of points in the data array. EWMA is often used in microprocessors to smooth the signal; the noise reduction coefficient K is calculated as the ratio of the variances of the random component of the error before smoothing and after smoothing by EWMA [14,15] and equals

$$K = D[y_i]/D[Y_i] = 1/\sum w_i^2 = (2-\alpha)/\alpha.$$

Table 1. Cumulants of exponential and geometric distributions
Таблица 1. Полуинварианты (кумулянты) экспоненциального и геометрического распределений

Cumulant order	EWMA/Geometric distribution	Exponential distribution	Gaussian	EMG
M0	h	1	$H_G \cdot \sigma_G \cdot (2\pi)^{1/2}$	$H_G \cdot \sigma_G \cdot (2\pi)^{1/2}$
M1	$h(1-\alpha)/\alpha$	τ	μ_G	$\mu_G+\tau$
M2	$h^2(1-\alpha)/\alpha^2$	τ^2	σ^2	$\sigma_G^2+\tau^2$
M3	$2h^3(1-\alpha)(1-\alpha/2)/\alpha^3$	$2\tau^3$	0	$2\tau^3$
k4=M4-3·M2²	$6h^4(1-\alpha)(1-\alpha+\alpha^2/6)/\alpha^4$	$6\tau^4$	0	$6\tau^4$

Berthod [16] used a transform equivalent to EWMA formula 10 to produce discrete exponential modifications of functions. We will try to compare the EWMA having a parameter α and an exponential distribution with the parameter τ and make an estimate of the modeling errors. There is no single rule establishing a correspondence between the specified parameters of these distributions, we will select rules for such a correspondence using the statistical moments of the functions and distributions used in the simulation.

Statistical moments. Zeroth moment - peak area

$$M0 = \int_0^\infty f(x)\,dx \approx h \cdot \sum_{i=0}^{N} f(x_i) \quad (11)$$

The first moment in chromatography is called average retention time

$$M1 = \frac{1}{M0}\int_0^\infty x \cdot f(x)\,dx \approx \frac{h}{M0}\sum_{i=0}^{N}(x_i \cdot f(x_i)) \quad (12)$$

Moments starting with the second are usually calculated centrally with respect to position of the first moment. The second central moment is the variance of the peak:

$$M2 = \sigma^2 = \frac{1}{M0}\int_0^\infty (x - M1)^2 \cdot f(x)\,dx \approx \frac{h}{M0}\sum_{i=0}^{N}((x_i - M1)^2 \cdot f(x_i)), \quad (13)$$

the standard deviation σ is the square root of the variance.

Other central moments

$$Mn = \frac{1}{M0}\int_0^\infty (x - M1)^n \cdot f(x)\,dx \approx \frac{h}{M0}\sum_{i=0}^{N}((x_i - M1)^n \cdot f(x_i)) \quad (14)$$

can be additionally normalized to σ^n

$$Mn' = Mn/\sigma^n$$

The statistical moments of convolution have several useful properties that can be used in modeling. The zeroth convolution moment of two functions equals to the product of their zero moments

$$M0(f \cdot g) = M0(f) \cdot M0(g); \quad (15)$$

Statistical moments higher than zeroth can be cumulants (M1-M3) or be part of polynomial cumulants (M4 and higher):

$$M1(f \cdot g) = M1(f) + M1(g); \quad (16)$$

$$M2(f \cdot g) = M2(f) + M2(g); \quad (17)$$

$$M3(f \cdot g) = M3(f) + M3(g). \quad (18)$$

The fourth cumulant k_4 is a polynomial of the moments:

$$k_4 = M4 - 3 \cdot M2^2, \quad (19)$$

$$k_4(f \cdot g) = k_4(f) + k_4(g). \quad (20)$$

Table 1 contains first four cumulants of geometric, exponential, Gaussian and EMG distributions. Parameter h stays for sampling interval of the measurement array, it has nothing in common with H_G, height of the Gaussian peak.

In the case of experimental data, the higher the order of the moment, the greater the error of its estimation, so cumulants based on moments higher the third order should be used with caution, only after making sure that the error of their calculation is small enough to estimate target parameters.

Results and discussion

Simulation of experimental data. The analog signal generated by the detector is converted into a sequential set of numbers (digitized) by a chip called an analog-to-digital converter (ADC) in most cases regularly over time. The time between successive measurements *h* is called the measurement interval, the inverse is called the measurement frequency. The ADC measures the sig-

nal over a certain proportion of the measurement interval (measurement time), this proportion is called the duty cycle. The duty cycle can be expressed as a simple fraction or as a percentage. A duty cycle of 0.0 (0%) means that the measurement corresponds to the detector response value at the time of measurement, a duty cycle of 1.0 (100%) means that the number generated by the ADC is proportional to the integral of the signal during the entire measurement interval. An intermediate duty cycle means that the ADC integrates the signal for part of the time, and then it is busy with other work (for example, measuring another signal). An ADC with a duty cycle of 1.0 is called an integrating ADC. In chromatography the duty cycle can be very small for scanning mass spectrometric (MS) detectors and large for a traditional chromatographic detector if its signal is measured by an integrating ADC.

The points corresponding to the measurements in the absence of noise will fall on the signal waveform for an ADC with zero duty cycle, and on a waveform convolved with the hardware function (for example, with a rectangle with width equal to the measurement interval) in the case of an integrating ADC. The influence of integrating ADCs should be combined with the study of the effect of noise on peak modeling. In this work we will simulate measurements with zero duty cycle, since in such a model there is only one convolution and, accordingly, the model is simple.

We define modelling error E as maximum absolute deviation of the model signal ordinate from “true” signal divided by maximum signal in the range of modelling abscissas. To describe experimental data, the single-precision floating-point format is almost always enough [17], since the uncertainty of the mantissa of this number is approximately $E \leq 10^{-7} = 10^{-5}\%$, and taking into account the sign, corresponds to most precise ADC available (24-bit). Actual detector noises are much higher than ADC noise, and experimental peaks to be simulated typically have a signal-to-noise ratio of 10 to 1000, which greatly raises the model acceptability thresholds to $E \leq 10^{-3} = 0.1\%$.

Peak moments calculated from discrete digitized points do not have to be exactly equal to the moments of the original "analog" peak. In the paper [18] it was shown that if the peak and its moments are integrated by the trapezoidal rule, the error in estimating moments decreases abnormally quickly (faster than exponentially) with a decrease in sampling interval h. In the case of Gaussian peak, the area error of 0.1% (10^{-3}) is achieved at a sampling frequency of $\sigma_G/h \approx 0.62$ and falls by at least an order of magnitude with an increase in frequency by 0.1. Therefore, it can be expected that to obtain a relative error of area 10^{-7}, a measurement frequency of $\sigma_G/h \geq 1.1$ will be required. The error in estimating moments higher than zeroth also falls quite quickly, a model experiment on Gaussian modeling using Excel showed that the error of all five cumulants from Table 1 does not exceed 10^{-7} at $\sigma_G/h \geq 1.2$. Thus, it is to be expected that at a measurement frequency higher than $\sigma_G/h \geq 1.2$ points the errors in estimating Gaussian moments related to the discreteness of the data presentation can be neglected, and the calculated moments of the discrete representation of Gaussian will insignificantly differ from the "continuous" moments.

In the presence of a random measurement error, the moments may fail to provide a model that is optimal in the sense of the least squares; rather, moments should be used to calculate the first approximation of the peak shape, completing the process with conventional least squares optimization [19]. One of the main factors affecting the moment values is the position of the baseline [20-25].

Calculation of EMG using EWMA. Since the EMG formula has four parameters, it is necessary to specify them directly or indirectly to model the EMG. We can calculate EMG parameters from the known statistical moments of the peak:

$$\tau = (M3/2)^{1/3}, \quad (21)$$

$$\sigma_G = (M2 - \tau^2)^{1/2}, \quad (22)$$

$$\mu_G = M1 - \tau, \quad (23)$$

$$H_G = M0/\sigma_G(2\pi)^{1/2}, \quad (24)$$

and formulas 6 to 8 are used to calculate the values of the function at the points of the digitization grid. If necessary, the statistical moments of EMG are calculated using the formulas from Table 1. Algorithm 1 provides a way to calculate discrete approximations of EMG using EWMA and known statistical moments of the distribution.

Algorithm 1. EMG approximation using EWMA from known statistical moments M0-M3.

1. The value of the constant α is estimated. Calculation options will be discussed below.
2. The position of Gaussian is calculated using the first moment of the geometric distribution, Table 1: $\mu_G = M1_{EMG} - M1_{EWMA} = M1_{EMG} - (1-\alpha)/\alpha$
3. The Variance of Gaussian is calculated $\sigma_G^2 = M2_{EMG} - M2_{EWMA} = M2_{EMG} - (1-\alpha)/\alpha^2$
4. The height of Gaussian is calculated $H_G = M0_{EMG}/((2\pi)^{1/2}\sigma_G)$
5. Data array is generated with values of Gaussian function using estimated parameters and assuming digitization step h
6. Data array is smoothed using EWMA with parameter α.

Estimation of α using known τ

Method α1. This estimate is based on a comparison of the first moments of geometric and exponential distributions. Let's construct an exponent passing through all the points of the geometric distribution, assuming that numerical integration occurs according to the midpoint rule. In order to combine the exponent with the rectangles of the midpoint rectangle rule, the origin of the exponent must be half a step to the left of the first point of the geometric distribution, and as a result, the estimate of $M1_{exp} = \tau$ will be half a step of the grid greater than $M1_{EWMA}$:

$$\tau = M1_{EWMA} + 0.5h = \frac{h}{\alpha 1}\left(1 - \frac{\alpha 1}{2}\right) \quad (25)$$

The solution of this equation gives the value $\alpha 1 = h/(\tau + 0.5h) = 1/(\tau/h + 0.5)$. The EWMA parameter α can accept values from 0.0 to 1.0, from which the requirement for the frequency of the digitization grid immediately appears when modeling EMG: the pitch of the grid should be such that $\tau/h > 0.5$. The second and third moments of the resulting distribution differ from the given ones.

Method α2. This estimate is based on the equalization of the second moments of geometric and exponential distributions from Table 1. α2 is calculated by solving a quadratic equation $\tau^2 = h^2(1-\alpha 2)/\alpha 2^2$.

The root of the equation with a positive sign is taken:

$$\alpha 2 = 2/(1 + (4(\tau/h)^2 + 1)^{1/2}), \quad (26)$$

calculated as the ratio of the free term of the equation to the negative root. The values of the convolution moments M0-M2 coincide exactly, M3 differs from the specified one. There is no limit to the value of τ, α2 for any τ is in the range from 0 to 1. Method α2 gives a correct estimate of $\alpha 2 = 1$ at $\tau = 0$.

Method α3. This estimate is based on the equalization of the third moments of the geometric and exponential distributions and exponential distributions from Table 1. α3 is calculated by solving a cubic equation $M3 = 2\tau^3 = h^3(1-\alpha 3)(2-\alpha 3)/\alpha 3^3$. At $M3 = 2\tau^3 > 0.0963h^3$ ($\tau > 0.3638h$) the equation has a single real solution.

Method α4. This estimate is based on the replacement of the EWMA coefficients formula 9 by an exponent [16]. α4 is calculated from the ratio of the values of the "analog" exponent with a constant time τ at points separated from each other by a distance h along the abscissa axis

$$\alpha 4 = 1 - \exp\left(-\frac{h}{\tau}\right) \quad (27)$$

As will be shown later, methods α3 and α4 provide very close estimates of α.

Unlike the α1 and α2 methods, the secong moment of digital Gaussian in the α3 and α4 methods differ from second moment of the original "analog" Gaussian, but three (α4) or all four (α3) EMG moments M0-M3 after EWMA convolution are exactly equal to those specified by original “analog” function.

Approximation error. The error of EMG peak construction using EWMA was estimated as the maximum absolute value of the difference between EMG constructed using EWMA and EMG constructed by formula 6,

Table 2. Threshhold values of acceptable data rates for EMG modelling, pts/(min(σ,τ)) for different models

Таблица 2. Пороговые значения допустимых частот данных для моделирования ЭМГ, pts/(min(σ,τ)), для различных моделей

Model	E=0.10% with fitting M2	E=0.10% without fitting M2	E=0.00001% with fitting M2
α1	8	-	-
α2	5	5	-
α3	2	6	11
α4	2	6	11

divided by the height of the peak. Single formula 6 can be used instead of the set of formulas 6-8, as maximum modelling error is always in the region of EMG, calculated by this formula. Reasonable settings of the digitization grid pitch depend on the required error, we investigated the range from 2 to 30 points per σ_G and τ (Supplementary material). Results of error analysis are summarized in Table 2 which shows that the error of the modeling with the α2 method is preferrable to all other methods when applied with M2 assumed to be a known value, without M2 (step 3 of Algorithm 1) fitting. The reason seems to be in the fact, that α2 method generates correct M2 after EWMA smoothing by the way how α is calculated. The errors of the α3 and α4 methods with M2 adjustment are two orders of magnitude less than the error of α1 and α2 methods. The difference between α3 and α4 methods is significant only below 4 pts/τ, α3 performing better than α4. Besides, formula 27 of α4 is simpler for programming than formula of the single root of third-order equation and for this reason can be considered preferable. Errors of α2 method drop down with increasing data rate, but do not reach 10^{-5}% at 30pts per σ and τ data rate. In general, the fitting process looks like approximating peak by adjusting peak moments, which, as expected, decreases approximation error.

Improving accuracy by oversampling. Experimental data may be measured with the data rate lower than required by method α2. In practice, these restrictions do not create big difficulties, since step h can be easily reduced: instead of modelling N-point array with experimental step h, one can make a model using m times more points, building m·N-point intermediate array with the step h/m and construct N-point model for comparison with experimental data by extracting every m-th point from the intermediate array.

Distance of the model data from the beginning of the array. If the relative (normalized to maximum) values of the function being smoothed by EWMA at the extreme points of the data array do not exceed the target measurement error and the function can be considered monotonic before and after the apex point, then the error caused by the absence of values outside the array is insignificant.

In general case, the error of calculating EWMA increases because EWMA uses only values within the data array. If values outside the array are not higher than maximal value of the signal within the array, the "memory distance" of the EWMA for an error of 10^{-5}% is approximately $7\cdot\ln10\approx17\tau\approx17/\alpha$, i.e. for points with indices of 17τ/h from the ends of the array, the values of the function outside the array cease to be a problem with an error of 10^{-5}%. With a target error of 0.1%, a distance of $3\cdot\ln10\approx7\tau\approx7/\alpha$ is sufficient. In the case of negative τ and inverse direction of smoothing, distance should be measured from the end of the array.

Multiple times exponentially modified Gaussian (mEMG). There may be situations where the process is described by the sum of a normally distributed random variable and two or more random variables with an exponential distribution [11,12]. The probability

distribution for such a process will be described by twice or more times exponentially modified Gaussian distribution.

Multiple modification with one positive time constant is equivalent to a one-time modification by Erlang distribution, which is generated by the sum of several exponentially distributed variables. According to the central limit theorem of probability theory, as the number of convolutions with exponential distribution increases, the result of the convolution will tend to Gaussian, albeit very slowly compared to the moving arithmetic average. Discrete analog of Erlang distribution is a negative binomial distribution. If decay time constants τ are different, continuous distribution is called Hypoexponential distribution, and it can be modelled by multiple applications of EWMA with different coefficient α as described it this paper.

The method described above for calculating exponentially modified functions is perfect for generating mEMG. As the simplest example of the generation of the peak shape (perhaps not very vital, but visual), we will take an n-fold modification of Gaussian with the same time constant exponentials. Statistical moments of multiple convolutions of exponential distribution will look like

$$M3=2\sum\tau_i^3=2n\tau^3 \quad (28)$$

$$M2=\sum\sigma_i^2=\sigma_G^2+n\tau^2 \quad (29)$$

$$M1=\sum M1_i=\mu_G+n\tau \quad (30)$$

$$M0=H\cdot\sigma_G\cdot(2\pi)^{1/2} \quad (31)$$

Conditions for the existence of a solution of the above system of equations:

$$\sigma_G^2=M2-n\tau^2\geq 0$$

whence

$$n\leq M2^3/(M3/2)^2=(2/M3')^2$$

where $M3'=M3/M2^{3/2}$ is a normalized third moment of the distribution.

Let us consider modelling of experimental peak by mEMG with equal τ using statistical moments. The lower the initial coefficient of skewness of the experimental peak M3', the more variants of its description by mEMG exist. If initial asymmetry $M3'>\sqrt{2}$, then the number of variants drops to one, i.e. there is no alternative to a once-modified Gaussian. At M3'>2, the peak with given moments cannot be described by the EMG or mEMG. When choosing a value of n within the permissible limits, we get a solution

$$\tau=(M3/2n)^{1/3},$$

$$\sigma_G=(M2-n\tau^2)^{1/2},$$

$$\mu_G=M1-n\tau,$$

$$H_G=M0/\sigma_G(2\pi)^{1/2}.$$

This solution is constructed similar to Algorithm 1:

Algorithm 2. Approximation of mEMG using EWMA.

1. The value of constants τ and α is estimated.

2. Position of the Gaussian is calculated using M1 formula of geometric distribution, Table 1 (μG=M 1-M 1EWMA=M1-n(1-α)/α).

3. The variance of Gaussian σG²=M2-M2EWMA=M2--n(1-α)/α² is calculated.

4. The height of Gaussian HG = M0/((2π)1/2σG is calculated.

5. An array of data with Gaussian values is generated.

6. EWMA with the parameter α is applied to the resulting data array n times.

We can try to evaluate the difference in the shape of the mEMG peaks with the same moments up to the third and a different number of exponential modifications and hence fourth moments. Let's approximate the EMG, built according to the "classical" formulas, by Gaussian, multiple times modified by EWMA. The first four statistical moments M0, M1, M2, M3 are taken as parameters that must correspond to the peaks. When constructing peaks, the following parameters were used: n=1 ... 7; M0=100.00; M1=250.00; M2=450.00; M3=6750.00; $\sigma_G=\tau_1=15.00$.

The difference between the curves is clearly demonstrated in Figure 1, which shows the solution difference profiles and the "true" curve calculated by formulas 5-8. Approximation with a single modification of EWMA almost accurately reproduces the

Сорбционные и хроматографические процессы. 2024. Т. 24, № 6. С. 944-955.
Sorbtsionnye i khromatograficheskie protsessy. 2024. Vol. 24, No 6. pp. 944-955.

ISSN 1680-0613

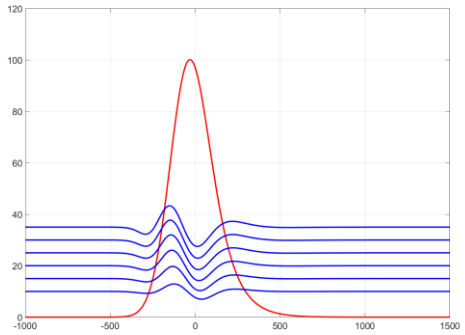

Fig. 1. The initial peak of EMG and six curves describing its difference from EWMA mEMG of approximations with identical moments M0 ... M3 and different multiplicity of exponential modification, from 2 to 7. Model α2 is used.

Рис. 1. Исходный пик ЭМГ и шесть кривых, описывающих его отличие от EWMA аппроксимаций mEMG с идентичными моментами M0 ... M3 и различной кратностью экспоненциальной модификации, от 2 до 7. В иллюстрации используется модель α2

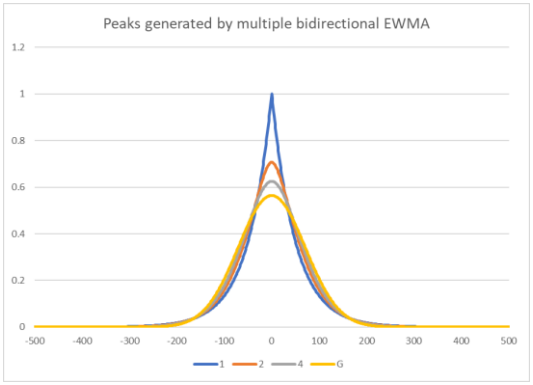

Fig. 2. Symmetrically relaxed delta function M0=100, M1=0. In all cases, after smoothing M2=5000. Smoothing was performed in one combined EWMA pass in the forward and reverse directions (curve 1, τ=50), in two passes (curve 2, τ=50/√2), in four passes (curve 4, τ = 25) and Gaussian (curve G).

Рис. 2. Симметрично-релаксированная дельта-функция M0=100, M1=0. Во всех случаях после сглаживания M2=5000. Сглаживание проводилось за один комбинированный проход EWMA в прямом и обратном направлении (кривая 1, τ=50), за два прохода (кривая 2, τ=50/√2), за четыре прохода (кривая 4, τ = 25) и по Гауссу (кривая G).

original EMG, as the multiplicity of modification increases (the higher the multiplicity of modification, the higher the curve is shifted), the amplitude of differences increases.

Since we have added another variable – the multiplicity of modification n, we can add to the system of equations 28-31 another equation to determine its value, namely the equation for the fourth cumulant of EMG: $k4=6\sum\tau_i^4=6n\tau^4$. If the value of k4 is known with sufficient accuracy, then it is easy to calculate n:

$\tau=k4/(3\cdot M3)$.

$n=M3/2\tau^3=27\cdot M3^4/(2\cdot k4^3)$. The more smoothings, the lower is excess kurtosis $k4/M2^2$.

According to our estimates, in most practical cases of experimental data, it is very difficult to calculate the fourth moment with sufficient accuracy, therefore one should be extremely careful in modelling by multiple EMG smoothings.

Exponential modification with negative τ. In the case of EMG negative τ corresponds to subtraction of exponentially distributed variable from the normally distributed one and can be modelled by the passage of the EWMA in the direction from the end of the data array to its beginning:

$$Y_N = y_N;\ Y_k = \alpha y_k + (1-\alpha)Y_{k+1}. \quad (31)$$

The absolute values of all moments of the geometric distribution remain the same, in the functional representation the direction of the convolution changes. Formulas 6-8 continue to work at a negative value of τ. Using both negative and positive modifications generates a whole class of peak-like distributions, including symmetric ones, which can be called symmetrically relaxed.

Figure 2 shows graphs of the Dirac delta function smoothed forward and backward one, two, and four times, overlayed with Gaussian. All functions are built to have the same variance; model α2 is used. Distribution with one pass in each direction is an approximation of Laplace distribution [26]. As

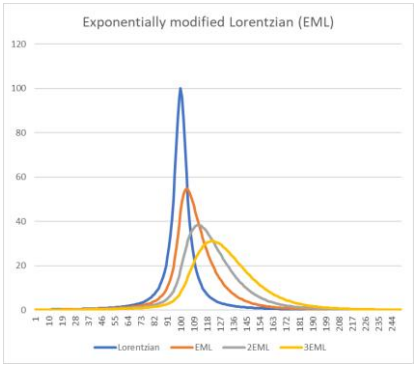

Fig. 3. Y=100/(1+(x-100)²/25) Lorentzian, modified with EWMA 1, 2 and 3 times; α=0.09.
Рис. 3. Лоренциана Y=100/(1+(x-100)²/25), модифицированная EWMA один, два и три раза, α=0.09

number of smoothings increase, distribution tends to become closer to Gaussian in full accordance with the central limit theorem.

Exponentially modified Cauchi distribution (Lorentzian). Multiple modification by the EWMA can be made using any set of time constants with any sign and it can be applied not only to Gaussian, but also to any function. In particular, EWMA can be applied to the Lorentz function [16]. Such a peaks can occur, for example, in spectroscopy, with exponential modification originated from hardware distortion of the line. There is no analytical formula for exponentially modified Lorentzian (EML) peak shape, but it can be used in practice, modeling the exponential modification of Lorentzian using EWMA. Unlike EMG, moments cannot be used to estimate EML parameters, since the Cauchy distribution is a classic example of a function in which moments other than zeroth cannot be computed because the corresponding integrals diverge. According to Equation 3, the maximum of the function F(x) is on the curve f(x) (Fig. 3).

Reversibility of convolution and peak sharpening. The exponential convolution operation is reversible, the inverse operation is called deconvolution and is described by Equation 3. This equation is a fairly obvious property that follows from the application of Laplace transformations to convolution, and in chromatographic literature we first came across it in the Ashley paper [27], where even an analog electrical circuit that implements the deconvolution was described.

The EWMA operation is also reversible, the reverse operation allows one to restore the original function:

$$y[i]= Y[i] + (1 - \alpha)\cdot(Y[i]- Y[i-1])/\alpha = Y[i]+\tau\cdot Y'_R[i] \quad (32)$$

$Y'_R[i]=(Y[i]- Y[i-1])/h$– (right) estimate of the first derivative; $\tau=h(1 - \alpha)/\alpha$.

In the past, Equation 3 was repeatedly "rediscovered" [28-31]. Unfortunately, in one chromatogram, peaks with different asymmetries may occur and even overlap, and the application of deconvolution with one constant τ to the entire chromatogram may create problems in the interpretation of data. It is dangerous to implement digitally evaluated derivatives in the equation 3, as they introduce quite high level of uncertainty in the reconstructed profile due to instrument noise and finite differences errors. Peak fitting provides much higher precision of reconstruction of peak parameters, and peak models, generated using EWMA algorithm, may help in understanding processes, generating these peaks.

Computational properties of the algorithm. Traditional EMG formulas 5-8 include multiplication of exp() to erfc() or erfcx() functions. Replacement of error functions by EWMA requires only two multiplications and one addition per point and thus should decrease overall computation time.

Conclusions

We hope, that the proposed in this paper way to calculate exponentially modified

functions may find wide application in practice of analysis of asymmetric overlapping peaks or for out-of-range peak reconstruction. The proposed algorithm allows very fast computation of peak shapes with reasonable accuracy, includes straightforward support of multiple exponential modifications with different τ.

Supplementary Materials

The following supporting information can be downloaded at: HYPERLINK "https://zenodo.org/record/7525896#.Y77py_7P2yo""Fast calculation of exponentially modified functions" spreadsheet | Zenodo

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work presented in this paper.

References

1. Sternberg J.C. Extracolumn contributions to chromatographic band broadening. in *Advances in Chromatography* (eds. Giddings, J.C. & Keller, R.A.). 1966: 205-270 (Marcel Dekker).

2. Grushka E. Characterization of exponentially modified Gaussian peaks in chromatography. *Anal. Chem.* 1972; 44: 1733-1738.

3. Delley R. Series for the Exponentially Modified Gaussian Peak Shape. *Anal. Chem.* 1985; 57: 388.

4. Kalambet Yu.A., Kozmin Yu.P., Mikhailova K.V., Nagaev I.Yu., Tikhonov P.N. Reconstruction of chromatographic peaks using the exponentially modified Gaussian function. *J. Chemom.* 2011; 25: 352-356.

5. Delley R. The peak width of nearly Gaussian peaks. *Chromatographia*. 1984; 18: 374-382.

6. Jeansonne M.S., Foley J.P. Review of the Exponentially Modified Gaussian (EMG) Function Since 1983. *J. Chromatogr. Sci.* 1991; 29: 258-266.

7. McWilliam I.G., Bolton H.C. Instrumental peak distortion. I. Relaxation time effects. *Anal. Chem.* 1969; 41: 1755-1762.

8. Di Marco V.B., Bombi G.G. Mathematical functions for the representation of chromatographic peaks. *J. Chromatogr. A.* 2001*;* 931: 1-30.

9. Romanenko S.V., Stromberg A.G. Classification of mathematical models of peak-shaped analytical signals. *J. Anal. Chem.* 2000; 55: 1024-1028.

10. Oldham, K. B. An algorithm for the exp x2 erfc x function. *J. Electroanal. Chem. Interfacial Electrochem.* 1982; 136: 175-177.

11. Delley R. Modifying the Gaussian Peak Shape with More Than One Time Constant. *Anal. Chem.* 1986; 58: 2344-2346.

12. Purushothaman S. Hyper-EMG: A new probability distribution function composed of Exponentially Modified Gaussian distributions to analyze asymmetric peak shapes in high-resolution time-of-flight mass spectrometry. *Int. J. Mass Spectrom.* 2017; 421: 245-254.

13. Foley J.P., Dorsey J.G. A Review of the Exponentially Modified Gaussian (EMG) Function: Evaluation and Subsequent Calculation of Universal Data. *J. Chromatogr. Sci.* 1984; 22; 40-46.

14. Kalambet Yu.A. Optimization of parameters of linear smoothing applied to chromatographic peaks. *Nauchnoe Prib.* 2019; 29: 51-65.

15. Ziegler H. Properties of Digital Smoothing Polynomial (Dispo) Filters. *Appl. Spectrosc.* 1981; 35: 88-92.

16. Berthod A. Mathematical Series for Signal Modeling Using Exponentially Modified Functions. *Anal. Chem.* 1991; 63: 1879-1884.

17. *IEEE Std 754™-2019 (Revision of IEEE Std 754-2008) IEEE Standard for Floating-Point Arithmetic*. (2019). https://doi.org/10.1109/IEEESTD.2019.8766229

18. Kalambet Yu.A., Kozmin Yu.P., Samokhin A. Comparison of integration rules in the case of very narrow

chromatographic peaks. *Chemom. Intell. Lab. Syst.* 2018; 179: 22-30.

19. Papai Z., Pap T.L. Determination of chromatographic peak parameters by non-linear curve fitting using statistical moments. *Analyst* 2002; 127: 494-498.

20. Gritti F., Guiochon G. Accurate measurements of peak variances: Importance of this accuracy in the determination of the true corrected plate heights of chromatographic columns. *J. Chromatogr. A* 2011; 1218: 4452-4461.

21. Anderson D.J., Walters R.R. Effect of Baseline Errors on the Calculation of Statistical Moments of Tailed Chromatographic Peaks. *J. Chromatogr. Sci.* 1984; 22: 353-359.

22. Gao H., Stevenson P.G., Gritti F., Guiochon G. Investigations on the calculation of the third moments of elution peaks. I: Composite signals generated by adding up a mathematical function and experimental noise. *J. Chromatogr. A* 2012; 1222: 81-89.

23. Stevenson P.G., Conlan X.A., Barnett N.W. Evaluation of the asymmetric least squares baseline algorithm through the accuracy of statistical peak moments. *J. Chromatogr. A* 2013; 1284: 107-111.

24. Howerton S.B., Lee C., McGuffin V.L. Additivity of statistical moments in the exponentially modified Gaussian model of chromatography. *Anal. Chim. Acta* 2003; 478: 99-110.

25. Eikens D.I., Carr P.W. Application of the Equation of Error Propagation to Obtaining Nonstochastic Estimates for the Reproducibility of Chromatographic Peak Moments. *Anal. Chem.* 1989; 61: 1058-1062.

26. Kotz S., Kozubowski T.J., Podgórski K. *The Laplace Distribution and Generalizations*. *The Laplace Distribution and Generalizations* (Birkhäuser Boston, 2001). https://doi.org/10.1007/978-1-4612-0173-1

27. Ashley J.W., Reilley C.N. De-Tailing and Sharpening of Response Peaks in Gas Chromatography. *Anal. Chem.* 1965; 37: 626-630.

28. Johansson M., Berglund M., Baxter D.C. Improving accuracy in the quantitation of overlapping, asymmetric, chromatographie peaks by deconvolution: theory and application to coupled gas chromatography atomic absorption spectrometry. *Spectrochim. Acta Part B At. Spectrosc.* 1993; 48; 1393-1409.

29. Gilmutdinov A.K., Shlyakhtina O.M. Correlation between analytical signal and rate of sample atomization in electrothermal atomic-absorption spectrometry. *Spectrochim. Acta Part B At. Spectrosc.* 1991; 46: 1121-1141.

30. Felinger A. Deconvolution of Overlapping Skewed Peaks. *Anal. Chem.* 1994; 66: 3066-3072.

31. Wahab M.F., O'Haver T.C., Gritti F., Hellinghausen G., Armstrong D.W. Increasing chromatographic resolution of analytical signals using derivative enhancement approach. *Talanta* 2019; 192: 492-499.

Информация об авторах / Information about the authors

Ю.А Каламбет – к.ф.-м.н., учредитель и генеральный директор ООО «Амперсенд»; инженер Институт физической химии и электрохимии им. А.Н. Фрумкина РАН, Москва, Россия

Ю.П. Козьмин – учредитель и внештатный научный сотрудник ООО «Амперсенд», Москва, Россия

Yu.A. Kalambet – Ph.D., Founder and CEO of Ampersand LLC; Engineer, Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, Moscow, Russia

Yu.P. Kozmin – the founder and freelance Researcher of Ampersand LLC, Moscow, Russia

Статья поступила в редакцию 05.11.2024; одобрена после рецензирования 27.11.2024; принята к публикации 04.12.2024.
The article was submitted 05.11.2024; approved after reviewing 27.11.2024; accepted for publication 04.12.2024.