



Comparison of integration rules in the case of very narrow chromatographic peaks

Yuri Kalambet^{a,*}, Yuri Kozmin^b, Andrey Samokhin^c

^a Ampersand Ltd., Kurchatov sq.2 bld. 2, Moscow, 123182, Russia

^b Shemyakin-Ovchinnikov Institute of bioorganic Chemistry RAS, Ulitsa Miklukho-Maklaya, 16/10, GSP-7, 117997, Moscow, Russia

^c Chemistry Department, Lomonosov Moscow State University, 1-3 Leninskiye Gory, GSP-1, 119991, Moscow, Russia



ARTICLE INFO

Keywords:

Narrow chromatographic peak
Data sampling error
Simpson's rule
Trapezoidal rule
Integration
Euler-Maclaurin formula

ABSTRACT

Theory of peak integration is revised for very narrow peaks. It is shown, that Trapezoidal rule area is efficient estimate of full peak area with extraordinary low error. Simpson's rule is less efficient in full area integration. Theoretical conclusions are illustrated by digital simulation and processing of experimental data. It was shown that for Gaussian peak Trapezoidal rule requires 0.62 points per standard deviation (2.5 points per peak width at baseline) to achieve integration error of only 0.1%, while Simpson's rule requires 1.8 times higher data rates. Asymmetric peaks require higher data rates as well. Reasons of poor behavior of Simpson's rule are discussed; averaged Simpson's rules are constructed, these rules coincide with those based on Euler-Maclaurin formula. Euler-Maclaurin rules can reduce error in the case of partial peak integration. Higher peak moments (average retention time, dispersion, skewness, etc.) also exhibit extraordinary low errors and can potentially be used for evaluation of peak shape.

1. Introduction

Extension of data rate range in the direction of low data rates is an important capability that can be very useful in the case of fast chromatography, hyphenated techniques, chromatography–mass spectrometry data processing. These techniques sometimes produce data with quite little number of points per peak, and capability to extract useful information from these data can significantly help researchers.

In this article we focused on the theory of data processing in the case of very low data rates, typically considered as unacceptable due to insufficient number of points per peak [1]. Main attention is paid to peak area, which is the major metrological characteristic of the peak.

The task of evaluation of sufficient data rate in chromatographic analysis started to be discussed in early 1970's [2–9] after appearance of computer data processing in chromatography. Authors paid great attention to influence of noise level, in most studies proper determination of peak height, width and asymmetry factor was required. Approaches to the problem of area integration were usually based on:

1) theoretical conclusions made using Fourier transform and information theory: according to [6] 0.9 pts/ σ is needed to achieve <0.1% of integration error;

2) digital modeling experiments: according to [7] 0.5 pts/ σ is needed to achieve <1% of integration error, lower error limits were not achieved due to noise simulation. Seeley [9] studied dependence of peak parameters on duty cycle using rectangle rule and confirmed value of 0.5 pts/ σ for small duty cycles;

3) theoretical conclusions made using textbook error formulas for integration rules: Trapezoidal integration rule requires 14 pts/ σ to achieve <0.1% of integration error [8]; Simpson's rule for the same accuracy requires 1.7 pts/ σ [2] or 2.5 pts/ σ [8].

All data rate requirements correspond to Gaussian peak. In papers that used textbook error formulas [2,8] no digital modelling was made; estimates were based entirely on theoretical considerations. Requirement of such a large number of points per peak in third approach contradicts results of first two approaches [3,4,6], and our own estimates. We decided to revise argumentation used in textbooks, especially for the case of peak-like function. This is especially important, as some of papers insist on using Simpson's rule in chromatographic integration software [8].

* Corresponding author.

E-mail addresses: kalambet@ampersand.ru (Y. Kalambet), y.kozmin@ampersand.ru (Y. Kozmin).

2. Theory

2.1. Notations and formulas

Variable x stays for elution time, volume, distance or other independent retention parameter.

Peak is a real analytic (in the math sense) function $f(x)$ of one real variable x , such that the function itself and all its derivatives can be considered equal to zero outside of finite interval $x \in (a, b)$. Mathematical term “Analytic” means, that we can use Taylor series for analysis of the function.

Our definition of peak is too simple from the rigorous mathematical point of view, but perfectly fits the case of experimental data processing. It allows us to avoid using $o()$, $O()$ and \sum , presenting just the ideas of proofs in the simplest way. Exact zero outside the interval (a, b) is not a must, but in practice of data processing all experimental data are produced by some analog-to-digital converters (ADCs), that output integers as a result of conversion. Signal in the region, where peak function $|f(x)|$ (with subtracted baseline) becomes smaller, than ADC conversion unit (or baseline noise), can be considered as zero together with all statistically significant derivatives. Experimental data processing should be arranged so, that derivatives, which cannot be measured with sufficient accuracy, can be neglected.

Data grid (frame). We assume, that function $f(x)$ is measured or calculated at discrete set of $N+1$ points $\{x_0, x_1, x_2, \dots, x_N\}$ with equidistant sampling period (step) h :

$$x_i = a + i \cdot h + \varepsilon; i = 0 \dots N; h = \frac{b-a}{N}; -h/2 < \varepsilon < h/2$$

where ε is a digitization grid shift, which is a random real number uniformly distributed on the interval $-h/2 < \varepsilon < h/2$ (probability density equals to $P(\varepsilon) = 1/h$ inside this interval and $P(\varepsilon) = 0$ outside it). The reason of introducing term ε is in the lack of advance knowledge about the position of the peak apex with respect to grid (e.g. due to variability of chromatographic retention time from run to run).

Exponentially Modified Gaussian (EMG) function [10–13].

$$f(x) = h_G \cdot e^{-\frac{(\mu_G - x)^2}{2\sigma_G^2}} \cdot \frac{\sigma_G}{\tau} \cdot \sqrt{\frac{\pi}{2}} \cdot \operatorname{erfcx}\left(\frac{1}{\sqrt{2}} \left(\frac{\mu_G - x}{\sigma_G} + \frac{\sigma_G}{\tau}\right)\right) \quad (1)$$

where h_G is height, μ_G – position of the apex, σ_G – standard deviation of unmodified Gaussian; τ – time constant of modifying exponent; $\operatorname{erfcx}()$ – scaled complementary error function [14]. Dispersion of EMG σ^2 can be calculated [10–13] as $\sigma^2 = \sigma_G^2 + \tau^2$

Euler-MacLaurin formula [15]

$$\int_{x_0}^{x_N} f(x) dx = h \left(\sum_{i=0}^N f(x_i) - \frac{f(x_0) + f(x_N)}{2} \right) + \sum_{k=1}^m h^{2k} \frac{B_{2k}}{(2k)!} (f^{(2k-1)}(x_0) - f^{(2k-1)}(x_N)) + R_m \quad (2)$$

where B_{2k} is a Bernoulli number ($B_2 = 1/6$; $B_4 = -1/30$; ...), $2m$ is maximal derivative order used in calculation and R_m is a remainder term, evaluating contribution of derivatives, higher than $2m$. We present Euler-Maclaurin formula not exactly like in textbook, but solved for integral part.

Peak moments.

Zeroth peak moment is peak area

$$M0 = \int_{-\infty}^{\infty} f(x) dx \approx h \cdot \sum_{i=0}^N f(x_i) \quad (3)$$

First moment is average retention time (unfortunately, it is rarely used in chromatography)

$$M1 = \frac{1}{M0} \int_{-\infty}^{\infty} x \cdot f(x) dx \approx \frac{1}{M0} \sum_{i=0}^N (x_i \cdot f(x_i)) \quad (4)$$

Second central moment is a dispersion of the peak (standard deviation σ is a square root of dispersion):

$$M2 = \sigma^2 = \frac{1}{M0} \int_{-\infty}^{\infty} (x - M1)^2 \cdot f(x) dx \approx \frac{1}{M0} \sum_{i=0}^N ((x_i - M1)^2 \cdot f(x_i)) \quad (5)$$

Other moments usually are presented not only central, but also normalized to σ^n .

$$Mn = \frac{1}{M0 \cdot \sigma^n} \int_{-\infty}^{\infty} (x - M1)^n \cdot f(x) dx \approx \frac{1}{M0 \cdot \sigma^n} \sum_{i=0}^N ((x_i - M1)^n \cdot f(x_i)) \quad (6)$$

Instead of third moment, it is convenient to use estimate of parameter τ of EMG peak function with the same third moment [10–13]:

$$\tau = \sigma(M3/2)^{1/3} \quad (7)$$

For simplicity of presentation digital formulas of moments correspond to Midpoint Rectangle integration rule.

Phase shift $\varphi = \varepsilon/h$.

Duty cycle – fraction of sampling period where function (signal) is averaged during measurement. Can be expressed as a fraction of one or in percent. All considerations of this paper correspond to instantaneous measurements with duty cycle of 0.0. For integrating ADC with duty cycle 1.0, peak area is defined by the sum of measurements by default and even one-point peak in the absence of noise will have exactly measured area. Duty cycle for the first dimension in the 2-D chromatography is usually close to 1.0, while duty cycles for fast scanning UV detectors or single quadrupole GC-MS are close to 0.0.

Data rate $\nu = \sigma/h$.

2.2. Integration rules

The task of integration is to estimate area – definite integral of function $f(x)$ on (a, b) .

All composite integration rules can be represented by a single formula:

$$A = h \cdot \sum_{i=0}^N w(x_i) f(x_i) \quad (8)$$

where A is area, $w(x_i)$ – weight coefficients. Rectangle, Trapezoidal, Simpson's and other composite integration rules differ from each other by the set of coefficients $w(x_i)$.

2.2.1. Rectangle and Trapezoidal rules give identical peak areas

Let us set limits of summation in formula 8 so, that $f(x_1) = f(x_N) = 0$ according to our definition of peak. For the Rectangle rule, all coefficients are ones $w() = \{1, 1, 1, \dots, 1, 1, 1\}$. Weight coefficients of Trapezoidal rule are $w() = \{1, 2, 2, \dots, 2, 2, 1\}/2$. As peak function is equal to zero on the boundaries of the integration interval, areas of the peak calculated using Rectangle and Trapezoidal rules are exactly equal.

In general case we should note that integration limits for composite Midpoint Rectangle rule are from position $x_0 - h/2$ to position $x_N + h/2$, and integration limits for Trapezoidal rule are from x_0 to x_N . If we adjust integration limits for Midpoint Rectangle rule to those of Trapezoidal rule by throwing away half of first and last rectangles, weight coefficients of two rules will exactly coincide; their common weight formula is that of Trapezoidal rule.

2.2.2. Simpson's composite integration rule provides two estimates

Simpson's 1/3 (further named just Simpson's) rule utilizes parabola built for three successive points (Fig. 1). It has coefficients $w() = \{1, 4, 1\}/3$ for three successive nodes (elementary rule) and $w() = \{1, 4, 2, 4, \dots, 4, 2, 4, 1\}/3$ for odd number of nodes (composite rule) [15]. In the case of peak, we can get two different estimates of area, shifting first point of integration frame by one point. These two implementations of Simpson's

Download English Version:

<https://daneshyari.com/en/article/7561810>

Download Persian Version:

<https://daneshyari.com/article/7561810>

[Daneshyari.com](https://daneshyari.com)