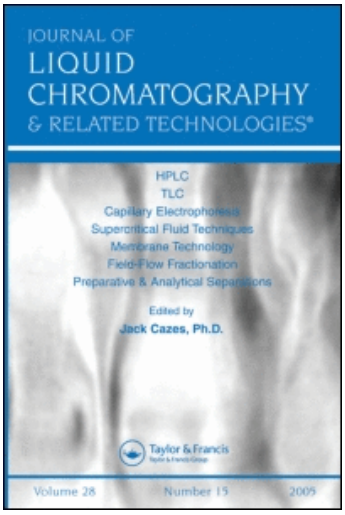


Journal of Liquid Chromatography & Related Technologies
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

Numerical Integration of Complex Chromatograms Using Fitted Gaussian Functions

Matti O. Koskinen[a]; Leena K. Koskinen[b]
[a] Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland [b] Department of Clinical Chemistry, Tampere University Hospital, Tampere, Finland

To cite this Article Koskinen, Matti O. and Koskinen, Leena K.(1993) 'Numerical Integration of Complex Chromatograms Using Fitted Gaussian Functions', Journal of Liquid Chromatography & Related Technologies, 16: 15, 3171 — 3184
To link to this Article: DOI: 10.1080/10826079308019641
URL: http://dx.doi.org/10.1080/10826079308019641

JOURNAL OF LIQUID CHROMATOGRAPHY, 16(15), 3171-3184 (1993)

NUMERICAL INTEGRATION OF COMPLEX CHROMATOGRAMS USING FITTED GAUSSIAN FUNCTIONS

MATTI O. KOSKINEN[1] AND LEENA K. KOSKINEN[2]
[1]*Department of Clinical Physiology*
[2]*Department of Clinical Chemistry*
Tampere University Hospital
P.O. Box 2000
SF-33521 Tampere, Finland

ABSTRACT

A microcomputer program is presented for the analysis and integration of peaks in chromatography using numerical gaussian fittings and repetitive peelings. The method is based on a IBM compatible PC microcomputer and Basic language. Data is transported to the computer from primary output curves of a chromatography device by a coordinate reader. This makes the system applicable to any chromatography apparatus. The program itself is adaptive to almost any IBM PC compatible microcomputer. The method was developed especially for the analysis of complex chromatograms with severely overlapped peaks. It was tested by using spectra of hemoglobin, obtained from cation exchange liquid chromatography.

INTRODUCTION

The integration of peaks in complicated chromatogram is a common problem with controversial solutions. Overlapped peaks give rise to difficulties in all types of chromatography. A straightforward way to analyse this kind of curves is to cut

them in to different bands simply with sharp upright lines as is done in most commercial devices. The fluctuating baseline, another frequent problem in chromatography, is also often solved by unsatisfactory methods. Several methods have been published in the literature, where computer programs have been designed to detect peaks and integrate their areas (1-9). In some of them, attention has been drawn to correction of the drawbacks mentioned above (3,4,9). Many of the programs control highly automated real time systems, the chromatography as well as integration (1,5,6). These types of systems do not permit the operator to contribute to individual integration procedures.

In this work we developed a method for integration of complex chromatograms. The program is interactive at various stages permitting the discretion of operator in problematic cases.

METHODS

General Principle

The chromatograms are analysed mathematically by fitting multiple gaussian curves to the measured raw data. Integration can then be carried out numerically by using the mathematical formulae of gaussian curves. The analysing computer program is written in PowerBasic (10) for a personal microcomputer. The measuring and analysing parts of the system are separate, which can be considered a drawback in routine work. On the other hand it makes it possible to use the analysing program for almost any kind of measuring apparatus.

Input

The chromatogram curve is entered to the program manually from original integrator sheet using a coordinate reader and a stylus. The baseline of the chromatogram is given to the system by pointing the start and end points of the curve at the zero level. Time scale is fixed by pointing two maxima and entering their retention time values accordingly. The path of the curve is then given by

picking up several values on the curve. The analysing program will interpolate between this sparse set of points by using a spline function. This makes manual input easy: distance between points should only be small where the curvature is high. Input data have not to be evenly spaced in x-direction. The resolution of coordinate reader is better than 10 points per mm. In the case of too small a size of the initial curve, zooming with a common copying machine can be used.

Spline Fit

A spline function is used to smooth and interpolate the primary input data points. The spline is a continuing set of second order polynomials going exactly through all the measured points (11). The computer uses this secondary set of spline smoothed data points for further analysis. This later curve is evenly spaced in the x-direction. The curve is reconstituted to 1000 discrete points that seems to be dense enough. Spatial resolution can be adjusted by selection of zoom in the primary sheet and the number of points in the spline curve.

The discrete entered data points and the smooth spline curve are shown on the computer screen (Fig.1). Peaks of chromatogram can be selected for gaussian fitting and peeling processes using cursor and keyboard.

Gaussian Fit

The program fits a gaussian function to the given part of the curve using a least square method. The program has been written by using methods of common linear algebra with matrix operations described in detail in Appendix. The essential subroutines for matrix inversion have been translated from FORTRAN (12) to PowerBasic. Double precision numbers have to be used in some parts of computing.

Peeling off the Curves

The peeling method is shown schematically in Figure 2. The program automatically selects the highest peak of actual curve to be fitted and presumes

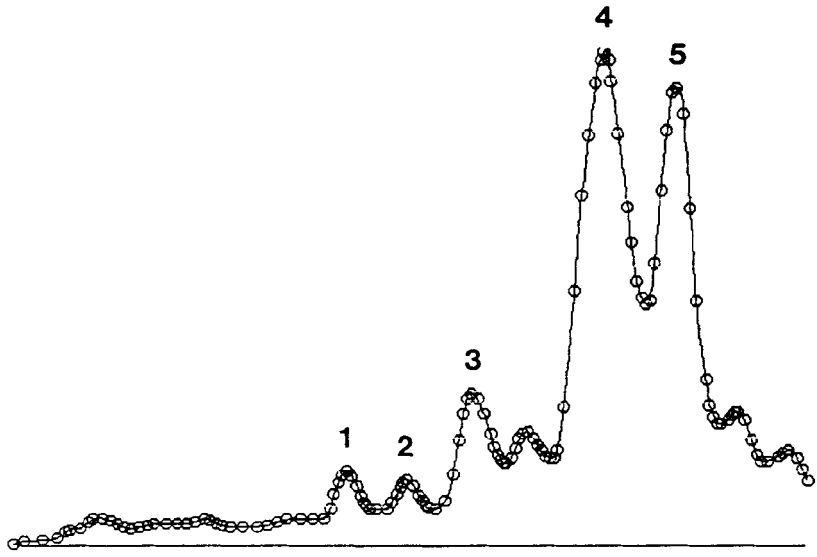


FIGURE 1.

Entered data points and spline smoothing. Chromatogram: fast fractions of normal blood hemoglobin after incubation in glucose and acetaldehyde. The peaks are: 1, HbA_{1b}; 2, HbF; 3, acetaldehyde adduct; 4, labile HbA_{1c}; 5, stabile HbA_{1c}.

left and right limits for fitting. The operator may accept or change the settings and the next gaussian fitting will be carried out, and so on. The residual curve is always shown on the screen and will be the basic curve for the next step of fitting. It is easy to see how much of the peak area there is left to be peeled off.

In this way the set of overlapped peaks can be interactively separated to their component parts. If the fitting is unsatisfactory, the operator can always cancel the last fit, change fitting limits and try again.

Integration

Areas of peaks are calculated numerically with trapezoidal method, because the gaussian function can not be integrated analytically in the mathematical sense.

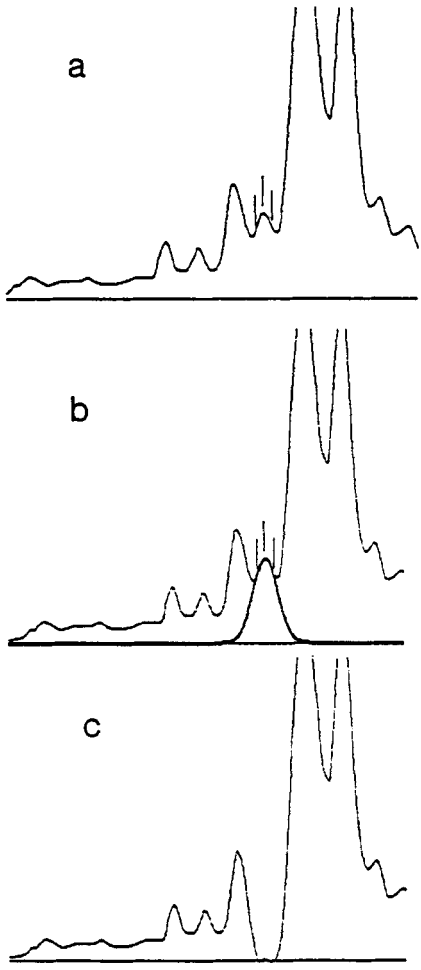
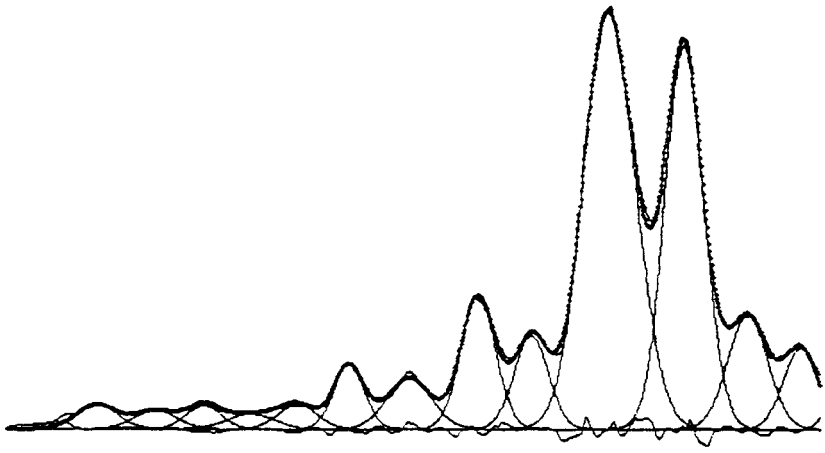


FIGURE 2.

Stripping one peak out of chromatogram in Figure 1: a) selection of peak and limits for fitting by setting cursor lines on the curve, b) fitted gaussian function and c) curve after subtraction.



Peak	Time(min)	Area	%FitArea	%RawArea
1	9.09	1.02	1.65	1.66
2	10.67	0.80	1.28	1.29
3	11.96	0.85	1.36	1.37
4	13.21	0.71	1.15	1.15
5	14.47	0.98	1.58	1.59
6	15.84	2.00	3.22	3.24
7	17.44	2.34	3.76	3.79
8	19.27	4.73	7.60	7.65
9	20.68	3.34	5.37	5.41
10	22.79	22.56	36.27	36.50
11	24.79	15.53	24.96	25.12
12	26.46	4.76	7.65	7.70
13	27.89	2.57	4.14	4.17

Total area under fitted curve = 62.20029
Total area under raw curve = 61.80904
Fitted/Raw area = 1.0063

FIGURE 3.

The same chromatogram as in Figure 1 as analysed to gaussian peaks and residual curve.
%FitArea = the percentage of peak area of the total fitted curve area.
%RawArea = the percentage of peak area of the total raw curve area.
The ratio of fitted to raw total area is presented as a measure of the final fitting quality.

Output

Finally the program displays graphically the initial raw curve, its reconstruction to several peaks, the fitted total sum curve of gaussians and the residual curve (Fig. 3). Numerical values of peak areas and their relative ratios are shown on the screen and will be printed when needed. The quality of fitting is described here as a ratio of fitted to raw curve area. It is printed as well (Fig. 3). In this case the fitted curve follows almost exactly the curve of the raw data demonstrating the good quality of fitting.

Storage

The curves can be stored on magnetic disks or diskettes and read out of files for analysis. This has proved to be convenient in reanalysis of curves and training the integration technique.

Hardware and Software

Coordinate reader: Graphical Tablet MM1201 of Summagraphics.
Computer: Almost any IBM PC compatible personal computer, for example a Decstation 316 with Intel 80386SX processor. A mathematical coprocessor is recommended for a reasonable speed.
Software: Program was written by using PowerBasic version 2.0 having advanced commands to handle the graphical display. PowerBasic produces compiled executive code which makes it fast for this kind of computations. Program is not computer dependant.

Analysing Time

The analysis of a curve like one shown in Figure 3 takes 3-4 minutes when performed by an experienced operator. The most time consuming step is the manual input of chromatogram.

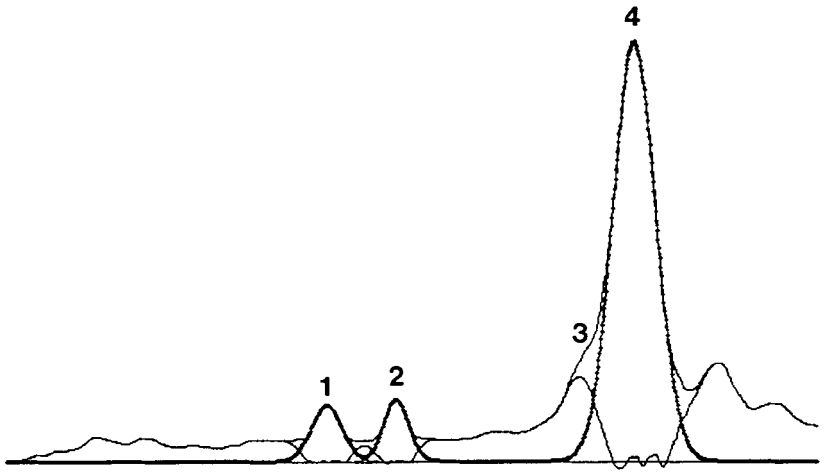


FIGURE 4.

Analysis of hemoglobin fast fraction chromatogram in HbA_{1c} determination: elimination of a shoulderlike peak caused by labile HbA_{1c}. The peaks are: 1, HbA_{1b}; 2, HbF; 3, labile HbA_{1c}; 4, stabile HbA_{1c}.

Testing the Program

All tests were performed with real chromatograms of fast hemoglobins produced by cation exchange liquid chromatography (FPLC system with Mono S Hr 5/5 column, Pharmacia, Uppsala, Sweden, and Shimadzu CR-5A integrator, Shimadzu Corporation, Kyoto, Japan).

The repeatability of the integration has two possible sources of error: the manual input of the curve and the gaussian fitting process. This was tested by entering the chromatogram in Figure 4 25 times repeatedly and computing the relative areas and retention times of three interesting peaks (HbA_{1b}, HbF and HbA_{1c}). Coefficients of variance were 2.4, 1.8 and 1.4 per cent of areas and 0.3, 0.3 and 0.4 per cent of retention times, respectively. These figures thus include the variance due to the whole process; the input of curve and the fitting mathematics. The chromatogram chosen to test indicates the ability of the

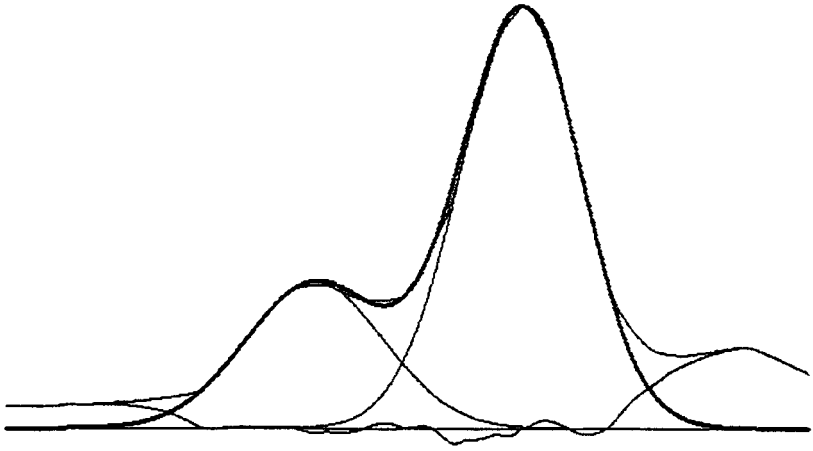


FIGURE 5.

Integration of two overlapped peaks. Chromatogram: labile (left) and stabile (right) HbA_{1c} after incubation of normal blood with glucose.

program to overcome peaks with shoulders. The curve was obtained from HbA_{1c} measurement where the shoulder caused by labile HbA_{1c} is a problem.

It is also obvious that there will be differences in results due to the order in which the overlapped peaks are peeled off. This feature was tested by analysing two adjacent partially overlapped peaks in both orders. In the example the area of the bigger peak was approximately two times the smaller one (Fig. 5). The effect to the results was 2.4 per cent for the smaller and 1.0 per cent for the bigger peak. The strength of the effect markedly depends on the shape of the spectrum. The operator has to be aware of this problem.

DISCUSSION

We certainly do understand that it is primary to optimise the chromatography itself to give clearly separated peaks. All kinds of mathematical methods to fix poor curves can only be secondary. In reality, complex overlapped spectra do exist, anyway. We met severe difficulties in the analysis of hemoglobin

spectra with an ordinary commercial integrator, this lead us to build the present computing method. We do believe that we have been able to analyse them in more detail than it was possible before. We also believe that our method gives a better approximation for peak areas than the older straight cutting method did. Even the shoulderlike peaks, which are impossible to analyse with most straight cutting methods, can be evaluated.

The repeatability of data input was fairly good, although a manual input method was used. The operator must be careful when peeling curves, because the settings of fitting borders and the order of analysis may affect the results. The best order to peel curves is to start from a discrete, clearly visualised peak and then continue to smaller ones. The interactive computation gives a good visual control to the analysis step by step. It is also easy to start calculations from the beginning without the need to enter the whole primary data again.

The baseline setting is quite a controversial subject in integrating chromatograms. The idea of the present method is that the baseline is a horizontal line at the zero level. The peeling method reveals, when we are dealing with hemoglobin, that the repetitive subtraction of low broad peaks will stabilise the baseline. It is also logical to believe, from a chemical point of view, that the whole area represents hemoglobin and must be taken into account when computing the relative portions of peaks. We think that the common method to draw the baseline "from valley to valley" will underestimate peak areas. This will distort mostly the smallest peaks which will be left partly, or totally, beneath the baseline.

In this work we have used gaussian fitting because the form of the peaks in hemoglobin spectra evidently resemble a normal distribution. The correctness of this analysis fails if the form of peaks differs significantly from normal distribution. In practice gaussians seem to fit fairly well to our experimental curves. There are reasons to believe that this kind of peak shape may be a more general phenomenon in chromatography. If the curves were skewed, like Poisson distributions, the program should be modified. Technically, only the form of matrix A should be changed (see Appendix).

It is most probably worthy to install an AD-converter card in the PC and read signals straight into the computer memory to avoid the most laborious phase of the work, the curve input.

Several methods have been presented in literature, where computer programs have been used to detect peaks and integrate peak areas, but they usually do not use analytical formula and numerical integration. The possibility to use this kind of mathematical fittings in chromatography is obvious. Nevertheless, the only article where that type of application has been described thus far is that of Ebel et al. who have used a Gaussian fitting by a non-linear iterative method (5).

APPENDIX

Fitting Gaussians to Measured Curves

Chromatograms are presented as a set of normal distributions

$$y(t) = \sum_{k=1}^{m} \frac{c_k}{\sigma_k \sqrt{2\pi}} e^{-\frac{(t-\mu_k)^2}{2\sigma^2 k}} \qquad (1)$$

where y = value of absorbance in chromatogram
t = time
μ = mean value
σ = standard deviation
c = scaling constant
m = number of gaussian components in the spectrum

Unknown parameters μ, σ and c for every gaussian k can be solved by taking natural logarithm of both sides of the equation (linearization), developing the square term and using following notations:

$$\log\left(\frac{c}{\sigma\sqrt{2\pi}}\right) - \frac{\mu^2}{2\sigma^2} = x_1$$

$$\frac{\mu}{\sigma^2} = x_2 \qquad (2)$$

$$-\frac{1}{2\sigma^2} = x_3$$

$$\log(\, y(t)) = z(t)$$

This can be written for every point i on the curve

$$z_i = x_1 + x_2 t_i + x_3 t_i^{\,2} + d_i \qquad (3)$$

where

z_i = logarithms of measured values

$x_1 + x_2 t_i + x_3 t_i^{\,2}$ = estimated values

d_i = residuals between measured and computed values

By using matrix notations

column vector $\mathbf{z}=\mathbf{z}(i)$, i=1,2,...,n

row vector $\mathbf{x}=\mathbf{x}(j)$, j=1,2,3

column vector $\mathbf{d}=\mathbf{d}(i)$, i=1,2,...,n (4)

n = number of data points in fitting

$$\mathrm{A} = \begin{pmatrix} 1 & t_1 & t_1^{\,2} \\ 1 & t_2 & t_2^{\,2} \\ . & . & . \\ . & . & . \\ 1 & t_n & t_n^{\,2} \end{pmatrix}$$

For simplicity the time step between points on the curve is set to be one unit. Integers can now be used during calculations. From a mathematical point of view the units of the abscissa are irrelevant. The only place where the true time units are needed is the list of retention times at the end of the whole analysis.

The set of equations can now be written

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{d} \qquad (5)$$

This over-determined set of equations can be solved by multiplication from left side with transpose of matrix **A**.

In the least square method the residual will be minimised so that

$$\mathbf{A^T d = 0} \tag{6}$$

Thus

$$\mathbf{A^T(z-Ax) = 0} \tag{7}$$

and

$$\mathbf{A^T z - A^T A x = 0} \tag{8}$$

Finally

$$\mathbf{x = (A^T A)^{-1} A^T z} \tag{9}$$

Superscript **T** means transpose and superscript **-1** inversion of matrix. For computational reasons it is notable that matrix to be inverted is of size of 3 x 3 only. Gauss-Jordan algorithm was used in the inversion.

It proved to be obligatory to use dual precision numbers in the crucial parts of program to avoid severe cumulative rounding errors.

The values of σ, μ and c are now easy to solve by entering components of vector **x** back in equations (2).

REFERENCES

1. Cardot, P.J.P., Trolliard, P. and Tembely, S. A Fully Automated Chromatographic Peak Detection and Treatment Software for Multi-user Multi-task Computers. J. Pharm. Biomed. Anal. 8, 755, 1990

2. Cattan, A.R. and Smith J., Computer Control of Fast Protein Liquid Chromatograpy (FPLC). Comput. Biol. Med. 20, 95, 1990

3. Ebel, S.,Herboth, M. and Wätzig, H. Parameterfreie Integrations-Software in der Chromatographie. Software-Entwicklung in der Chemie 2. Springer-Verlag, Berlin, Heidelberg,1988, 267-290

4. Giles, I.G. and Gore, M.G. Biochemical Data Processing with Microcomputers, Part 2. A Basic Program for the Detection, Integration and Area Assignment of Chromatographic Elution Profiles. Anal. Chim. Acta, 151, 123, 1983

5. Jinno, K. and Kuwajima, M. Microcomputer-Assisted Liquid Chromatographic Separation System: Application to Toxic Compounds Identification in Poisoned Human Fluids. J. Chromatogr. Sci. 27, 57, 1989

6. Olivo, J-P., Cardot, P., Ignatiadis, I. and Vidal-Madjar, C. Microcomputer System for Data Acquisition and Software Handling in Chromatography. J. Chromatogr. 395, 383, 1987.

7. Ryoo, J-P., Greer, R. and Carlson, D. Data Acquisition and Processing System for Drug Release Characterization by Thin Layer Chromatography. Biomed. Sci. Instrum. 25, 251, 1989.

8. Santangelo, J.D. A General Purpose Computer Analysis System for Chromatographic Data. Comput. Appl. Biosci. 4, 275, 1988

9. Singhal, R.P. and Smoll, D.B. Software for Microprocessor Controlled HPLC: Peak Area Integration. J. Liq. Chromatogr. 9, 2719, 1986

10. PowerBasic V2.0, 1990, Borland International Inc, USA

11. Users Manual, IMSL Library, FORTRAN Subroutines for Mathematics and Statistics, Vol. 1, Edition 9.2, Revised November 1984, IMSL Inc, USA

12. Scientific Subroutines, Programmer's Reference Manual, Updated 1982, Digital Equipment Corporation, Marlboro, USA

Received: January 30, 1993
Accepted: February 16, 1993