*Journal of Chromatography*, 587 (1991) 129–136
Elsevier Science Publishers B.V., Amsterdam

CHROM. 23 563

# Deconvolution method for accurate determination of overlapping peak areas in chromatograms

Thomas J. Nelson

*Neural Systems Section, National Institute of Neurological Diseases and Stroke, National Institutes of Health, Bethesda, MD 20892 (USA)*

(First received February 13th, 1991; revised manuscript received June 11th, 1991)

ABSTRACT

A method is described for deconvoluting chromatograms which contain overlapping peaks. Parameters can be selected to ensure that attenuation of peak areas is uniform over any desired range of peak widths. A simple extension of the method greatly reduces the negative overshoot frequently encountered with deconvolutions. The deconvoluted chromatograms are suitable for integration by conventional methods.

## INTRODUCTION

Although considerable effort has been expended over the years to improve the peak-resolving power of chromatographic techniques, one is frequently still faced with data-containing peaks which are incompletely separated from each other, or which contain large baseline drifts or other features which render accurate peak-area determination difficult. Conventional peak integration algorithms utilize running averages to compute baselines and simple tangent-skimming and perpendicular drop methods for separating peaks. Although robust, these algorithms are notoriously inaccurate [1,2], and require numerous parameters which must be determined and adjusted empirically for each chromatogram, a laborious and inexact process which in effect introduces numerous assumptions about peak width, peak shape, etc. In addition, as they generally rely on detecting points at which the first derivative ($dy/dx$) crosses zero to separate peaks, they are usually incapable of detecting peaks which are not separated by a horizontal region or valley.

An alternative approach is to deconvolute the chromatogram mathematically to remove all peaks above a certain threshold peak width (*i.e.*, below a threshold frequency). This also automatically flattens the baseline. In practice, however, attempts at deconvolution often produce unsatisfactory results, negative peaks or discrepancies in areas for peaks of different width. Other methods, such as moments analysis [3], orthogonal polynomial analysis of chromatogram segments [4,5] and an inverse diffusion model [6], have also been proposed for analyzing overlapping peaks. The relative advantages of several types of deconvolution methods have been compared [7]. In this paper, a straightforward method for deconvoluting chromatograms prior to integration is presented and its applicability to actual chromatographic analyses is discussed.

## EXPERIMENTAL AND RESULTS

Two most commonly used mathematical techniques for deconvoluting sets of numbers are nonlinear least-squares (NLR) analysis [8] and the Fourier transform method. In least-squares analysis, the chromatographic data are fitted to a sum of several Gaussians, solving for $3n$ parameters (position, width and height) for each peak. Although successful for single peaks [9], NLR analyses typically fail to converge when more than three or four

peaks are involved. In the Fourier transform method, the data are converted to a set of complex frequencies, then divided by the transform of a transfer function, which is a function which represents the data smearing process [10] (since Fourier transforms yield complex numbers, complex division must be used). The new set of complex frequencies is then inverse transformed to obtain the deconvoluted data. This method is also considerably faster than NLR.

The principle of deconvolution is that convoluting two functions ($b$ and $c$) is equivalent to multiplying their Fourier transforms ($B$ and $C$), *i.e.*, if

$$a(x) = b(x) \otimes c(x)$$

then

$$A(\omega) = B(\omega)C(\omega)$$

The variance of the convolution product of the two functions is equal to the sum of the variances of the individual functions:

$$\sigma_a^2 = \sigma_b^2 + \sigma_c^2$$

Deconvolution is simply the reverse of this process, dividing the transform of the data by that of the transfer function, thereby subtracting its contribution to the variance.

As the observed peaks in a chromatogram are products created by the peak-smearing process which we desire to reverse, the transfer function should in some manner mimic the peak shape. Although peak shape has been modelled as a simple Gaussian [11], a combination of Gauss and Cauchy functions [12,13], a Gaussian in the presence of noise [14], a polynomial series [5], a combination of statistical moments [15] or an exponentially modified Gaussian [16–18], there are reasons for using a simple Gaussian curve. Diffusion processes and response time-limited detector response functions, which are the main contributors to peak broadening, are both Gaussian in form. Also, conveniently, the Fourier power spectrum of a Gaussian is another Gaussian, as well as a real function. This reduces the number of transforms needed in deconvolution from three to two, as the desired Gaussian can be easily generated when needed. Finally, Gaussians do not produce fast Fourier transform (FFT) oscillation artifacts, in contrast with a sharp threshold cut-off function or triangle function.

Unfortunately, performing a deconvolution as described above frequently yields unsatisfactory results, as the amplitudes of some frequencies in the transfer function may be extremely small, resulting in enormous amplitudes of high-frequency components of the deconvoluted data, possibly even exceeding the dynamic range of the computer. An alternative method, shown in Table I, involves creating a high-pass filter function based on a Gaussian, retaining its favorable characteristics such as an absence of oscillation artifacts, $x$-axis symmetry and steep fall-off, but having a wide plateau for high frequencies and attenuating low ones. The non-zero real and imaginary Fourier frequencies are then multiplied by this function, that is, by

$$f = [1 - \exp(-b_f x^2)]\{1 - \exp[-b_f(x - n)^2]\} \qquad x = 1, 2, \ldots, n \qquad (1)$$

where $n$ is the number of data points and $x$ is the index for the points for (in this case) the frequency domain (Fig. 1). The $\omega = 0$ frequency is set to zero. Strictly, this is a type of convolution rather than deconvolution, as it is a multiplication rather than division, but the effect is that of deconvoluting wide and narrow peaks as the term is frequently used in chromatography. Fig. 1 shows the function used, which is essentially a filter function. The adjustable, dimensionless exponential constant for the filter, $b_f$, is related to the attenuation for a peak component of

TABLE I

PROCEDURE FOR DECONVOLUTING CHROMATOGRAMS

| Step | Action |
|---|---|
| 1 | Create Gaussian filter function of length $n$: $f = [1 - \exp(-b_f x^2)]\{1 - \exp[-b_f(x - n)^2]\}$ where $b_f$ is an adjustable parameter and $n$ is the number of data points in the chromatogram |
| 2 | Fourier transform the data |
| 3 | Calculate the new real and imaginary frequencies by multiplying the old frequencies $\omega^{(x)}$ by the filter function $f$ and adding a fraction $c$ of the original data frequencies $\omega_0^x$: $\omega^{(x)\prime} = \omega^{(x)} f^{(x)} + c\omega_0^{(x)}, \quad x = 0, 1, 2, \ldots, n$ |
| 4 | Inverse transform and apply a smoothing filter if necessary |

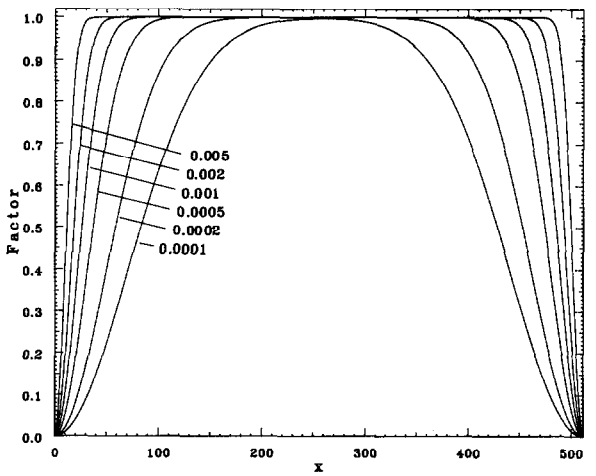


Fig. 1. Filter functions computed from Gaussians with an exponential coefficient ($b_f$) of 0.005, 0.002, 0.001, 0.0005, 0.0002 and 0.0001 for a chromatogram of 512 data points. The lowest frequencies (near both ends of the curve) are attenuated to a greater extent as the Gaussian becomes wider, whereas the highest frequencies (center) remain unchanged. The Nyquist (maximum) frequency is at $x = 256$. A simple step function cannot be used, as it would create artifactual sinusoids in the final result. The function also must be symmetrical in order to also handle negative frequencies, which are customarily placed in reverse order above the positive frequencies by an FFT.

frequency $\omega$ ($\omega$ = 1/peak width in $x$ units) by the equation

$$b_f = -w^2 \ln(1 - a)$$

where $a$ is the attenuation and $w$ is the peak width at half-height expressed as a fraction of the total length of the chromatogram. This equation provides a useful approximation of attenuation. For example, if the maximum acceptable attenuation for the broadest peak is 0.95, and the broadest peak is 4% of the total length of the chromatogram, then $b_f$ should be 0.004793 or greater. Peaks of width >8.3% of the total chromatogram will be attenuated by a factor of 2 or more. This value gives a good margin of safety, because for a typical high-pressure liquid chromatogram most of the peaks are <1% of the length of the chromatogram. Even values of $b_f$ which result in significant attenuation of some broader peaks may still be useful, as peaks eluting at similar retention times usually have similar height/width ratios.

The relationship between attenuation and peak width can be estimated more accurately from considering the power spectrum of an entire individual peak. The power spectrum is generated by taking the modulus of the real and imaginary components of the transform. It can be shown that the $\omega = 0$ component (*i.e.*, the amplitude) of a power spectrum is equal to the area under the original peak ($A_p$) in the time domain, and conversely, the area of the power spectrum curve (in the frequency domain) is equal to the height of the original peak ($h_p$) times $n$, the number of data points. The power spectrum curve of a Gaussian peak is another Gaussian centered at $x = 0$, with an amplitude of $A_p$, or

$$y = A_p \exp(-b_2 x^2)$$

The exponential coefficients of a Gaussian curve and its corresponding power spectrum conveniently happen to be related by a simple inverse multiplied by $\pi^2/n^2$. Thus, the exponential parameter $b_2$, the exponential coefficient of the power spectrum, is calculated from $b_p$, the exponential coefficient of the original peak, by the equation

$$b_2 = \pi^2 / b_p n^2 \tag{2}$$

where $b_p$ is related to the width of the original peak, $w_p$, by

$$b_p = 4 \ln 2 / w_p^2 \tag{3}$$

If the power spectrum is multiplied by the filter function (eqn. 1), it becomes

$$y = [1\text{-}\exp(-b_f x^2)] A_p [\exp(-b_2 x^2)]$$

or

$$y = A_p \{\exp(-b_2 x^2) - \exp[-(b_f + b_2) x^2]\} \tag{4}$$

The maximum value of this new curve, which occurs at a non-zero position, is equal to the area of the new deconvoluted peak. The maximum, found by setting $dy/dx$ to zero, is at

$$x = \left[ \frac{-1}{b_f} \cdot \ln\left( \frac{b_2}{b_f + b_2} \right) \right]^{\frac{1}{2}}$$

After substituting back into eqn. 4 to obtain the $y$ value at this point, one obtains

$$y_{max} = A_{new} = A_p \left( 1 - \frac{b_2}{b_f + b_2} \right) \left( \frac{b_2}{b_f + b_2} \right)^{b_2/b_f} \tag{5}$$

where the equality $\exp(a \ln b) = b^a$ was used to remove the exponential. This equation enables one

to calculate the area ($A_{new}$) of the new, attenuated Gaussian representing the deconvoluted peak.

It is also straightforward to calculate the height and width of the new peak, by integrating eqn. 4. However, the value obtained is the height measured from $y = 0$ and not from the negative minima on either side of the peak. Hence the calculated $h_{new}$ would be an underestimate of the actual new height. The negative regions of the deconvoluted peak arise because the new power spectrum is slightly asymmetric.

From the above equations, it is possible to select a value for $b_f$ to keep the area of a peak of any specified width from decreasing, within any desired tolerance.

For example, suppose that a chromatogram of length 512, with a typical peak width $w_p = 5$ and height $h_p = 7$, was to be deconvoluted using a $b_f$ of 0.01. The area of the peak (estimated by numerical integration or FFT) would be 37.26. Using eqn. 3, $b_p$ would be 0.1109. Then, using eqn. 2, one calculates $b_2$ to be 0.0003395. The new area would be 32.088 from eqn. 5. As this value is only 14% less than the starting value, the selection of 0.01 would be a suitable choice.

To test whether the deconvolution method can be used to obtain a calibration curve, samples containing hemoglobin, cytochrome *c* and various amounts of bovine serum albumin were subjected to size-exclusion high-pressure liquid chromatography (HPLC). The chromatograms were compressed to 512 data points and the areas of the peaks, which were partially overlapping, were measured before and after deconvolution. For the lowest albumin concentrations, area measurement by both the perpendicular drop and tangent-skin methods yielded less accurate results than measurement from the deconvoluted chromatograms. The areas after deconvolution were reduced by *ca.* 10% at all concentrations (Fig. 2), although at the lowest concentrations this effect was difficult to observe because of the greater uncertainty in estimating areas from the original chromatograms. In contrast, the peak in the deconvoluted chromatogram was clearly visible, even at the lowest concentrations (Fig. 2, inset).

Fig. 3a shows a chromatogram produced by an HPLC column under conditions of sample overloading, selected to show features of erratic baseline drift, unresolved peaks and a flat region bounded on

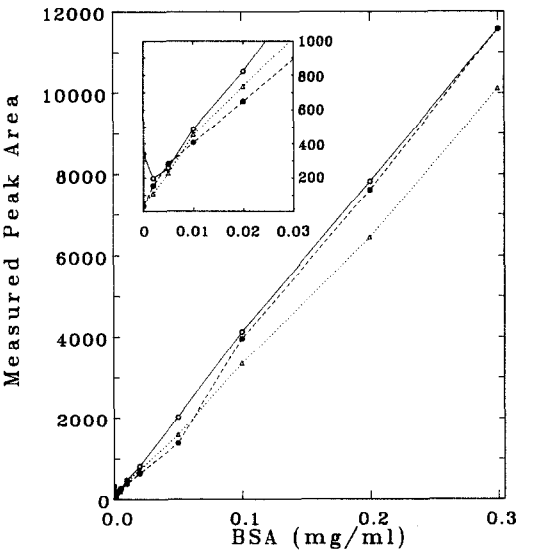


Fig. 2. Calibration curves derived from a set of chromatograms, measuring peak areas with the vertical drop method (○), tangential-skim method (●) or after deconvolution using a simple minimum-to-minimum method (△). The original chromatograms consisted of 0.1 mg/ml hemoglobin, 0.1 mg/ml cytochrome *c* and various concentrations of bovine serum albumin. The samples were chromatographed on a 250 × 4.6 mm I.D. size-exclusion HPLC column (Macrosphere GPC 60, 7 μm) eluted with 0.1 *M* potassium acetate (pH 7.4). Inset: enlargement of low-concentration region.

one side by a sharp edge, all of which make integration difficult. Fig. 3b–d show the chromatogram deconvoluted with the method in Table I, using $b_f$ values of 0.00005, 0.00015, 0.0005 and 0.0015. The peaks are resolved to a greater extent than the original, although in Fig. 3b and c the noise is also increased. Note the absence of oscillation artifacts in the flat region at the right. Fig. 3 also demonstrates the importance of selecting a value appropriate for the largest peak in the sample. In Fig. 3, the curve is too narrow, hence only the high-frequency noise is enhanced.

Also, there is a large negative overshoot near some of the larger peaks. This overshoot is one of the major drawbacks of deconvolution. Several sophisticated methods have been devised for removing these negative regions, which generally introduce a constraint function such as zero-clipping followed by several reiterations of the deconvolution [19]. However, a close examination of the transforms of the original deconvoluted chromatogram and one in

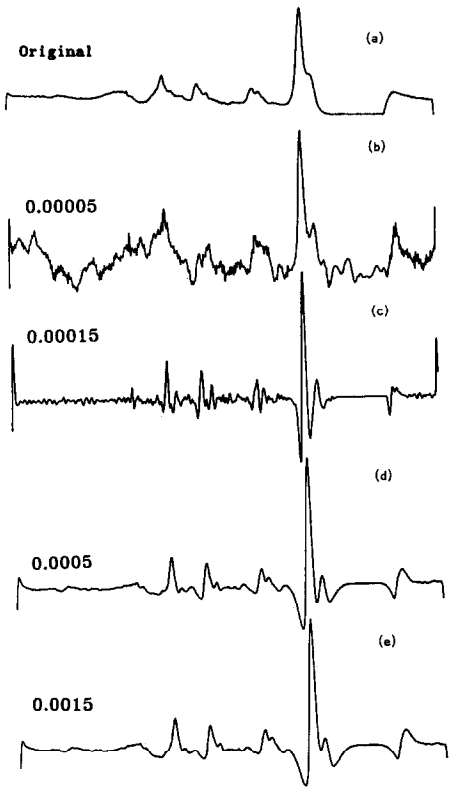


Fig. 3. Chromatograms deconvoluted by the method in Table I using different values for $b_f$. Thirty optic ganglia from squid (*Loligo paelei*) were homogeneized and chromatographed on DEAE-Sephacel. One fraction from this column was concentrated and re-chromatographed by HPLC on a column of AX-300, using a gradient of 0–0.6 *M* potassium acetate (pH 7.4) (0–20 min) and 0.6 M potassium acetate (20–60 min), and the absorbance was measured at 280 nm. The data set was compressed to 512 data points before processing.

which the negative regions have been artificially removed by subtraction would reveal that the desired changes are essentially an increase in the low-frequency regions which were attenuated by the deconvolution. Thus, a simple alternative method is to add back a portion of the frequencies of the original data. As shown in Fig. 4, even retaining a small fraction of original frequencies is sufficient to eliminate the negative regions, while the resolution is impaired to a lesser extent than if a wider curve were used (compare Figs. 3d and 4d). For integration purposes, in fact, it is actually not necessary to eliminate the negative regions, provided that the integration is performed between minima and not from the zero-crossings.

It is also possible to combine the retention of original frequencies by modifying the filter equation, as follows:

$$f = [1 - k\exp(-b_f x^2)]\{1 - k\exp[-b_f(x - n)^2]\}$$

The factor $k$ ranges from 0 to 1 and determines the fraction of the low original frequencies to be discarded. If $k$ is less than 1, the ends of the filter function (Fig. 1) approach 1-$k$ instead of 0, and the $\omega = 0$ frequency should also be set to 1-$k$.

The most important criterion for integration purposes is the degree to which relative peak areas are preserved for peaks of different width. At the same time, extremely wide peaks (resulting from baseline shifts) should be attenuated. Fig. 5 shows the effect of different filter widths on the preserva-

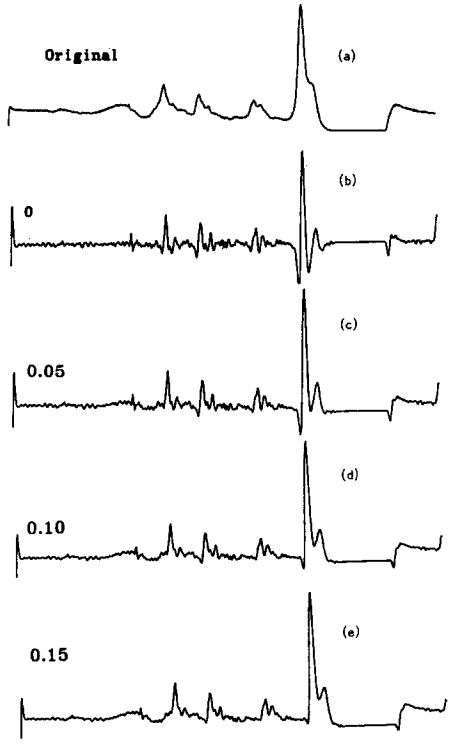


Fig. 4. Effect of adding increasing fractions of the original Fourier set during deconvolution, using a $b_f$ value of 0.0002 in the chromatogram from Fig. 3. Adding 5% of the original frequencies eliminates the negative overshoot almost completely, whereas the resolution is only slightly impaired. As more of the original frequencies are added back, the baseline drift begins to reappear.

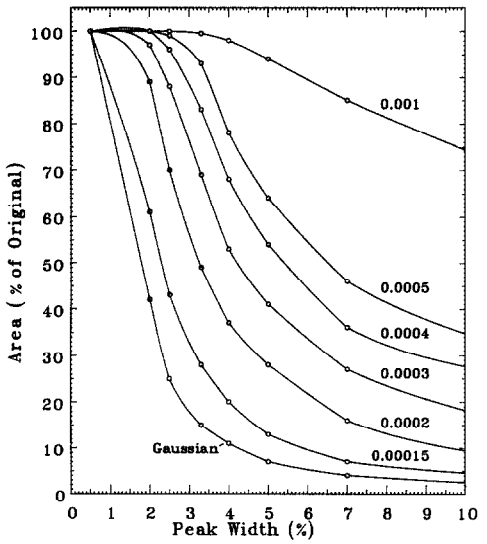


Fig. 5. Effect of width of the filter function used for deconvolution on attenuation of peaks of increasing width. A simulated chromatogram of 100 data points, containing peaks of various widths and area = 1, was deconvoluted using the method in Table I, and the areas in the deconvoluted chromatograms were computed. A broader function yields a wider plateau of low, constant attenuation. Fourier deconvolution using the standard method resulted in the curve marked "Gaussian" for all peak widths within a broad range. In contrast to the present method, the traditional Fourier deconvolution method does not produce a region over which attenuation is constant for different peak widths.

tion of peak areas with increasing peak width. Since for a typical HPLC trace the majority of peaks are less than 1% of the total length of the chromatogram, all but the narrowest peak produce satisfactory results, providing a fairly wide plateau. In contrast, results of a standard FFT deconvolution (using three Fourier transforms and complex division instead of the filtering method described here), are shown on the curve marked "Gaussian". The standard method attenuates peaks of increasing width to an increasing extent; hence the shape of this curve is relatively constant regardless of the width of the Gaussian transfer function used within almost the entire region of useful transfer function widths. Additionally, a peak whose width approached zero would also be enlarged to an unlimited extent, whereas in the present method its area would be unchanged.

The determine whether the recovered peak areas are dependent on the resolution between the components or their height ratios, a sequence of simulated chromatograms containing peaks of differen heights and different degrees of separation wa analyzed. Fig. 6 shows that, at the point at which th two peaks merge into a single peak, there is a sligh decrease (up to 10%) in the total recovered area. Th size of this decrease depends inversely on $b_f$, an results when the greater width of the unresolve peak shifts the filter function (Fig. 1) into a region o greater attenuation. Hence this effect can be mini mized by a careful choice of $b_f$. The effect is als maximal when the two peaks are of identical size

It is also possible to apply the deconvolutio method iteratively, using partial retention of origi nal frequencies and zero-clipping to reduce negativ peaks. This results in chromatograms such as th one in Fig. 7. Although the peaks are all clearl resolved, their areas depend on the original peal width to a greater extent than the result from a singl iteration, *i.e.*, the plateau in Fig. 1 is narrowe Hence, although it can help determine whethe multiple overlapping peaks exist, and return thei

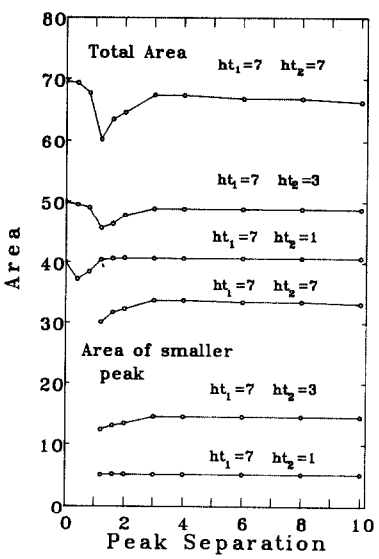


Fig. 6. Effect of the degree of peak separation and relative peal height on the recovered areas. Simulated chromatograms witl peaks of constant peak heights with maxima separated by variou distances were deconvoluted and the individual and total area were measured. For some values of the deconvolution paramete $b_f$, a slight decrease in recovered area can occur when the tw peaks are approximately one peak width apart. This effect i much less pronounced when the two peaks are of different sizes The abscissa is peak separation in units of peak width.

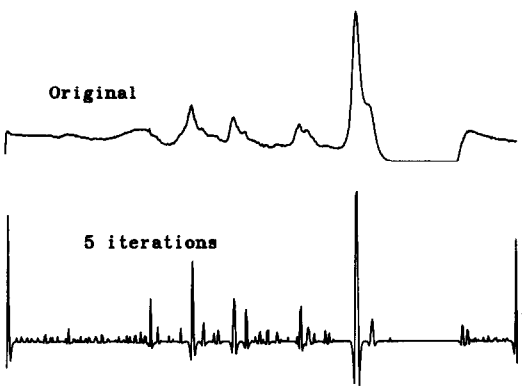


Fig. 7. Deconvolution of a chromatogram by the method in Table I applied for five iterations using a filter function based on $b_f = 0.0002$. The second to fifth iterations were carried out using a carry-over of 0.5 of the previous frequencies, and clipped at $y = 0$. All the peaks contain either one or two data points, making integration computationally trivial and allowing accurate determinations of retention times.

retention time, area measurements from the reiterative method should be viewed with caution.

In some instances, deconvolution adds excessive high-frequency noise, which can usually be removed by data smoothing before or after the deconvolution.

The effect of noise on the deconvolution is shown in Fig. 8. Simulated chromatograms of 512 data points, containing numerous overlapping peaks were deconvoluted with different values of $b_f$. As $b_f$ is decreased, the $S/(S + N)$ ratio ($S$ = signal; $N$ = noise) in the deconvoluted chromatogram becomes progressively lower than the original, because a greater proportion of the signal is being discarded. Hence, the decrease in the signal-to-noise ratio caused by using too low a value of $b_f$ exactly parallels the decrease in area recovery caused by the same factor. In the case of zero noise, the $S/(S + N)$ ratio is unaffected by deconvolution.

More complicated functions than those described in Table I could equally well be applied to this method, as long they contain a flat region over a fixed range of higher frequencies and a smooth cut-off, and produce a small number of artifacts. It is envisaged that some of these other functions could produce an even sharper cut-off of peak widths whose areas are consistently retained.

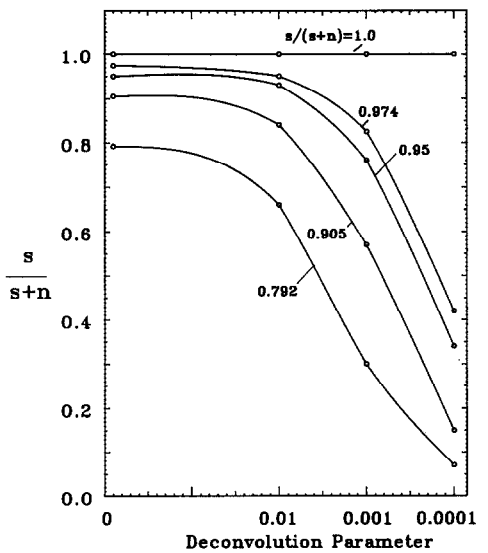


Fig. 8. Effect of chromatogram noise on the signal-to-noise ratio in the deconvoluted chromatogram. Simulated chromatograms containing 512 data points, with $S/(S + N)$ ratios of 1.0, 0.974, 0.95, 0.905 and 0.792, were deconvoluted using $b_f$ values of 0.01, 0.001 and 0.001. Similated white noise was produced by adding random numbers to the chromatogram. Smaller values of $b_f$ decrease the signal-to-noise ratio in direct proportion to their attenuating effect on the peak area (Fig. 5). No noise is added by the deconvolution, as indicated by the top curve.

ACKNOWLEDGEMENT

The helpful criticisms of Dr. Avrama Blackwell of the Environmental Research Institute of Michigan are appreciated.

REFERENCES

1 A. W. Westerberg, *Anal. Chem.*, 41 (1969) 1770–1777.
2 A. N. Papas and T. P. Tougas, *Anal. Chem.*, 62 (1990) 234–239.
3 E. Grushka, M. N. Myers and J. C. Giddings, *Anal. Chem.*, 42 (1970) 21–26.
4 H. J. G. Debets, A. W. Wijnsma, D. A. Doornbos and H. C. Smit, *Anal. Chim. Acta*, 171 (1985) 33–43.
5 P. J. H. Scheeren, Z. Klous and H. C. Smit, *Anal. Chim. Acta*, 171 (1985) 45–60.
6 R. Crandall, M. McClellan, S. Arch, J. Doenias and R. Piper, *Anal. Biochem.*, 167 (1987) 15–22.
7 L. M. Gugliotta, J. R. Vega and G. R. Meira, *J. Liq. Chromatogr.*, 13 (1990) 1671–1708.
8 D. W. Marquardt, *J. Soc. Ind. Appl. Math.*, 11 (1963) 431–441.
9 S. N. Chesler and S. P. Cram, *Anal. Chem.*, 45 (1973) 1354–1359.

10 W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986, pp. 407–414.
11 E. Grushka, M. N. Myers and J. C. Giddings, *Anal. Chem.*, 42 (1970) 21–26.
12 R. D. B. Fraser and E. Suzuki, *Anal. Chem.*, 38 (1966) 1770–1773.
13 R. D. B. Fraser and E. Suzuki, *Anal. Chem.*, 41 (1969) 37–39.
14 J. M. Laeven and H. C. Smit, *Anal. Chim. Acta*, 176 (1985) 77–104.
15 E. Grushka, M. N. Myers, P. D. Schettler and J. C. Giddings, *Anal. Chem.*, 41 (1969) 889–892.
16 J. P. Foley and J. G. Dorsey, *J. Chromatogr. Sci.*, 22 (1984) 40–46.
17 R. Pauls and L. B. Rogers, 49 (1977) 625–628.
18 J. P. Foley and J. G. Dorsey, *Anal. Chem.*, 55 (1983) 730–737.
19 P. A. Jansson, *Deconvolution, with Applications in Spectroscopy*, Academic Press, Orlando, FL, 1984.