NASA Contractor Report 3229

Deconvolution of Gas Chromatographic Data

Samuel Howard and Grayson H. Rayborn

GRANT NSG-1380
FEBRUARY 1980

NASA

NASA Contractor Report 3229

Deconvolution of Gas Chromatographic Data

Samuel Howard and Grayson H. Rayborn
University of Southern Mississippi
Hattiesburg, Mississippi

Prepared for
Langley Research Center
under Grant NSG-1380

National Aeronautics
and Space Administration

**Scientific and Technical
Information Office**

1980

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

This research involves the use of deconvolution methods on gas chromatographic data to obtain an accurate determination of the relative amounts of each material present by separating the merged peaks.

The underlying assumption implied throughout is that the area under the peaks is an accurate indicator of the amount of material present.[1] Due to uncontrollable errors at every stage of operation, this is never exactly true. However, it is a consideration that must always be borne in mind in designing, building, and operating the gas chromatograph, its detector, and other related equipment. As peak areas are less dependent on operating conditions than other methods, they are preferred by most researchers for accurate determinations.[2]

The promising new concept of function continuation in the frequency domain was introduced in striving to bring the art of deconvolution to its theoretical limit of accuracy. This proved only partially successful.

The methods used here will have general applicability to any deconvolution where the desired result consists of a number of well separated distributions.

Gas Chromatography

Gas chromatography is an effective and widely used method for the separation of gases and volatile liquids or solids in the gaseous state.

A small sample of the material being examined is injected into a stream of an inert gas such as nitrogen, hydrogen, carbon dioxide, argon, or helium which carries it into a column containing a suitable medium capable of retarding the flow, by varying degrees, of the individual components of the sample as they flow through the column. Differences in the time various components remain in adsorption or partition on the material in the column is again the factor which makes separation possible. The separated components then emerge from the column at discrete intervals (characteristic of each component) and pass through some form of detector. As a general rule, gas analyses are carried out on adsorption columns (gas/solid chromatography), while liquids and volatile solids are separated on partition columns (gas/liquid chromatography).[3]

The data used in the present study were obtained from a gas/liquid chromatograph. The carrier gas was helium and the detector was a Varian 1400 flame ionization detector. The sample compounds consisted of five xylenes.

The flame ionization detector is a differentiating type, making the area under the peak a suitable measure of the amount of material present.[4]

The flame ionization detector operates on the principle that the electrical conductivity of a gas is directly proportional to the concentration of charged particles within the gas. Effluent gas from the column is mixed with hydrogen and burned in air. Ions and electrons formed in the flame

enter the electrode gap, decreasing the gap resistance, thus permitting a current to flow in the external circuit.[5]

The flame ionization detector is extraordinarily insensitive to air and water, making it especially suitable for the analysis of air pollutants or aqueous samples such as beverages, biological materials, and other liquids. For other materials the areas obtained must be multiplied by the proper correction constant to obtain the true proportions.[6]

The flame ionization detector has the widest linear range of any detector in common use.[7]

The analog output of the gas chromatograph was digitized by bringing its output into the HP2100 minicomputer. Data were acquired at a 10 hz rate.[8]

Electronic digital integration was accomplished by accumulating the data points when an increase in signal was detected and terminating when the signal again increased, which, incidentally, began a new count.[9]

Temperature programming was used on the substances making up this data. The temperature was increased at a linear rate different for each run.[10]

The Data

Data are given for five runs. The same sample composition, consisting of five test substances (see Figures 1 and 2), was injected each time. The graph of Run 3 was not included as nothing new of significance was illustrated. Each run was taken at a different temperature. See Table 1 for specifications.

The interval between each data point was quite small compared to the standard deviation of the narrowest peak, thus making the data suitable for application of the fast Fourier transform.

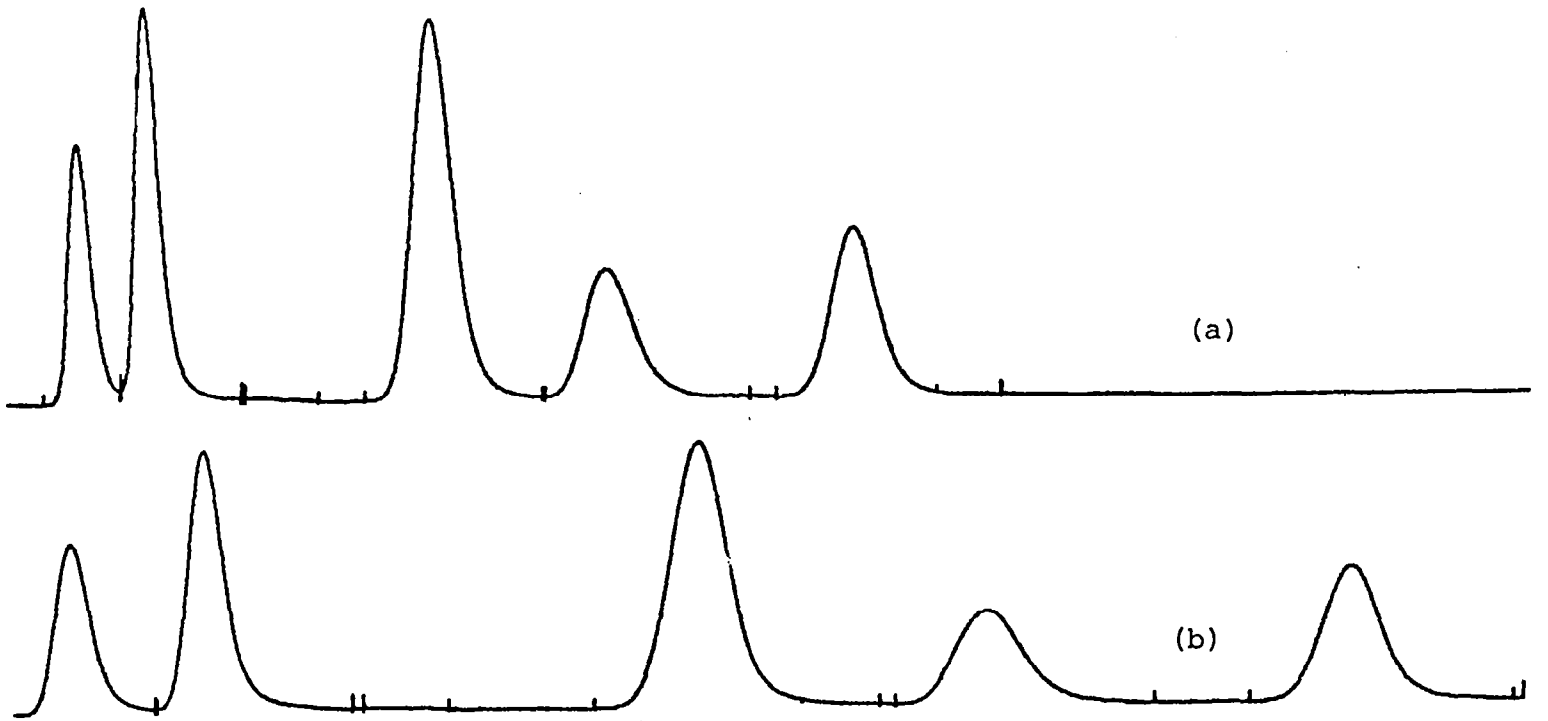


Fig. 1. Data of the first two runs: (a) is data of Run 2, (b) is data of Run 1.

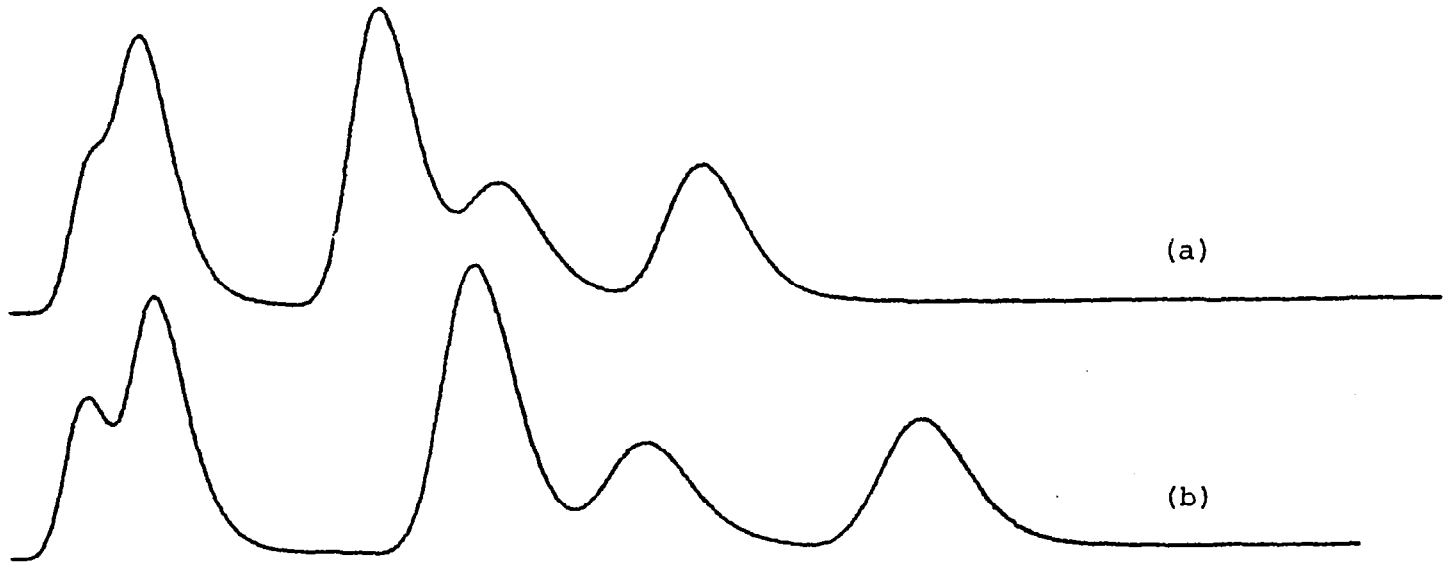


Fig. 2. Data of the last two runs: (a) is data of Run 5, (b) is data of Run 4.

TABLE 1

GAS CHROMATOGRAPH SPECIFICATIONS

Temperature Programming					
	Run 1	Run 2	Run 3	Run 4	Run 5
Beginning Temperature	50°	75°	100°	135°	150°
Rate of Change	4°/min.	4°/min.	4°/min.	6°/min.	6°/min.

G.C. Parameters	Integrator
10^{-9}	100 mv input = 1 v output
He 20 cc/min.	Slope sens. 0.01 up; 1.0 down
H_2 30 cc/min.	Zero delay
Air 300 cc/min.	Noise suppression 3
Injection 0.75µℓ 22A	Area threshold 1 Shoulder control on (front) rear 0.1

An inspection of the data points and its graph reveals that it is relatively smooth and free from noise.

The tic marks, quite noticeable on the graph, were imposed on the data whenever the integrator encountered any increase in the signal. Removing the tic marks was quite easy. Graphs were drawn which included a few points on both sides of the tic marks. A smooth curve was drawn through these points. Intermediate points were then taken from the graph and replaced those of the tic marks on the data. The interval was small and the data was digitized so that this method was sufficiently accurate for the purpose.

The baseline drift is negligibly small and the baseline will be taken at zero.

CHAPTER II

FAST FOURIER TRANSFORM DECONVOLUTION

The fast Fourier transform is a computational algorithm for calculating the discrete form of the complex Fourier series. (See Appendix B for a discussion of the discrete Fourier transform.) The addition theorem, shift theorem, convolution theorem, and other theorems familiar to users of the continuous Fourier transform with infinite limits may also be developed for the complex Fourier series.

The complex Fourier series, defined over the interval (0,2p) for any arbitrary function, f(t), is given by:

$$f(t) = \sum_{n=-\infty}^{\infty} C_n \exp(ni\pi t/p) \tag{2.1}$$

where the complex coefficients, c_n, are determined from:

$$C_n = \frac{1}{2p} \int_0^{2p} f(t) \exp(-ni\pi t/p)dt \tag{2.2}$$

The addition theorem may be stated as follows:

If C_{n1} and C_{n2} are the nth coefficients, respectively, of $f_1(t)$ and $f_2(t)$, then the nth coefficient, C_n, of $f = f_1 + f_2$ is $C_{n1} + C_{n2}$.

Proof:

$$C_n = \frac{1}{2p} \int_0^{2p} f(t) \exp(-ni\pi t/p)dt$$

$$C_n = \frac{1}{2p} \int_0^{2p} [f_1(t) + f_2(t)] \exp(-ni\pi t/p)dt$$

$$C_n = \frac{1}{2p} \int_0^{2p} f_1(t) \exp(-ni\pi t/p)dt + \frac{1}{2p} \int_0^{2p} f_2(t) \exp(-ni\pi t/p)dt$$

$$C_n = C_{n1} + C_{n2} \tag{2.3}$$

The shift theorem will be derived next.

If the function is shifted by an amount a, the coefficients will be changed as follows: Let these coefficients be distinguished from the previous coefficients by primes.

$$C_n' = \frac{1}{2p} \int_0^{2p} f(t-a) \exp(-ni\pi t/p)dt \tag{2.4}$$

$$C_n' = \frac{1}{2p} \int_0^{2p} f(t-a) \exp(-ni\pi a/p) \exp(ni\pi a/p) \exp(-ni\pi t/p)dt$$

$$C_n' = \frac{1}{2p} \exp(-ni\pi a/p) \int_0^{2p} f(t-a) \exp(-ni\,(t-a)/p)dt$$

Change variables: Let $x = t-a$ $\quad dx = dt$

$$C_n' = \frac{1}{2p} \exp(-ni\pi a/p) \int_a^{2p+a} f(x) \exp(-ni\pi x/p)dx \tag{2.5}$$

Consider the integral in Equation (2.5).

$$\int_a^{2p+a} f(x)\ \exp(-ni\pi x/p)dx = \int_a^{2p} f(x)\ \exp(-ni\pi x/p)dx$$

$$+ \int_{2p}^{2p+a} f(x)\ \exp(-ni\pi x/p)dx \qquad (2.6)$$

The complex exponential repeats itself after every interval of 2p, and its value from 2p to 2p+a is the same as its value from 0 to a. If f(x) is a periodic function of period 2p, then its value from 2p to 2p+a will also be the same as from 0 to a (the shift theorem will only be applied to functions with this property). We may then write Equation (2.6) as:

$$\int_a^{2p} f(x)\ \exp(-ni\pi x/p)dx + \int_0^{a} f(x)\ \exp(-ni\pi x/p)dx$$

$$= \int_0^{2p} f(x)\ \exp(-ni\pi x/p)dx \qquad (2.7)$$

Substituting this result in the integral in Equation (2.5):

$$C_n' = \frac{1}{2p}\ \exp(-ni\pi a/p) \int_0^{2p} f(x)\ \exp(-ni\pi x/p)dx$$

$$C_n' = \exp(-ni\pi a/p) \left[\frac{1}{2p} \int_0^{2p} f(x)\ \exp(-ni\pi x/p)dx\right]$$

$$C_n' = \exp(-ni\pi a/p)C_n \qquad (2.8)$$

Thus the nth coefficient for the shifted function can be obtained simply by multiplying the coefficient of the unshifted function by the phase factor $\exp(-ni\pi a/p)$.

With this result the convolution theorem for complex Fourier series may be derived.

The data given are assumed to be a convolution of resolved peaks with some "machine function" that causes a spreading and "smearing" of the peaks together. The complex coefficients, then, are the transform of this convolution:

$$C_n^h = \frac{1}{2p} \int_0^{2p} \left(\int_0^{2p} f(x)g(t-x)dx \right) \exp(-ni\pi t/p)dt \qquad (2.9)$$

where $g(t)$ is the machine function, and $f(t)$ is the assumed form of the data before convolution. Let C_n^h denote the coefficients of the data, C_n^g the coefficients (transform) of the machine function, and C_n^f the coefficients (transform) of the deconvolved data.

Note that the continuous integral was used rather than a discrete sum as of the discrete series. This is because the actual convolution was assumed to be a continuous process. Actually, as long as we have sufficient data points, the results of both methods should be essentially the same. The integral permits convenient mathematical manipulation.

Assuming sufficient continuity so that the order of integration may be interchanged:

$$C_n^h = \frac{1}{2p} \int_0^{2p} f(x) \left(\int_0^{2p} g(t-x) \exp(-ni\pi t/p)dt \right) dx$$

Applying the shift theorem:

$$C_n^h = \int_0^{2p} f(x)\ \exp(-ni\pi x/p) \left(\frac{1}{2p} \int_0^{2p} g(t)\ \exp(-ni\pi t/p)dt \right) dx$$

$$C_n^h = \int_0^{2p} f(x)\ \exp(-ni\pi x/p) C_n^g\ dx$$

$$C_n^h = C_n^g 2p\ \frac{1}{2p} \int_0^{2p} f(x)\ \exp(-ni\pi x/p)dx$$

$$C_n^h = 2pC_n^g C_n^f \tag{2.10}$$

The product of the coefficients of the original function and the machine function multiplied by the constant factor, 2p, gives the coefficients of the convolved function. When all that is given is the data and some machine function, and the separated peaks are desired, Equation (2.10) is usually written in the form:

$$C_n^f = \frac{1}{2p}\ \frac{C_n^h}{C_n^g} \tag{2.11}$$

The coefficients of the separated peaks may be obtained by dividing the coefficients of the data by the coefficients of the machine function and multiplying by the constant factor 1/2p. Transforming to the time domain will give the resolved peaks. This process is known as "deconvolution". It should be apparent that there is no unique set of coefficients that will satisfy this equation. The coefficients of the machine functions used in this research die out rather quickly at the higher frequencies and are essentially zero over most of the spectrum. Assuming C_n^h is zero in these regions, we have the

indeterminate case and almost any spectral function given for these regions where C_n^g is zero added to the unique determination C_n^h/C_n^g where C_n^g is not zero will satisfy Equation (2.11). One important particular solution gives $C_n^f = 0$ when $C_n^g = 0$. This is known as the "principal solution" as given by Bracewell and Roberts.[12]

Note here that a single isolated peak for the case of gas chromatography cannot be taken for the machine function as it is not the response of an impulse; it is the response of one of the original resolved peaks "smeared" by the machine function to give the data. It is not possible nor desirable to assume the original input were "spikes." The sidelobes quickly become formidable as the peaks become sharper. (See Appendix A for a discussion of sidelobes.) In accordance with the hypothesis that the area under the peak is proportional to the amount of material present, the deconvolution needs to be carried only to the point of barely resolving the peaks, hence the assumption that the original distribution of material before "smearing" was in the form of barely resolved peaks.

Of course, one realizes that the form of the convolution integral is obtained by considering the original input peaks to be broken up into infinitesimal "impulses," each one of these giving rise to infinitesimal "machine functions" positioned in a continuous manner on the t-axis and that summing all these together gives the data.[13]

A first guess at a good machine function, then, might be one of the same shape as an isolated peak, but much narrower; more specifically, if an isolated peak can be expressed in some functional form, g(t), then a machine function of the form g(at) would be tried, where a must be greater than one. This was tried with only moderate success. The best results were obtained by taking

this function and giving it more "skewness." The assertion that this function is a good machine function has no firm foundation in theory as will be discussed subsequently. First, the linearity and shift-invariance of the apparatus will be discussed.

The flame ionization detector is linear over a wide range. Stated mathematically, one would say that for certain amounts of different gases run through the chromatograph singly and forming the distributions: $f_1(t)$, $f_2(t)$, $f_3(t)$, etc., then differing amounts of these same substances run through the machine together would form the composite distribution:

$$af_1(t) + bf_2(t) + cf_3(t) + \ldots$$

regardless of whether or not they were overlapping. The constants a,b,c, etc., are determined by the proportions of the substances in the sample.

Even a cursory examination, however, reveals that the distributions are not shift-invariant. Being shift-invariant means that for an amount of gas coming out of the tube earlier or later than a given gas should have the same shape as that gas. To state this more explicitly, if the distributions are shift-invariant, then for an arbitrary distribution of functional form f(t), any other distribution could be expressed in the form $c_1 f(t-c_2)$, where c_1 and c_2 are constants. An examination of the data shows that the peaks seem to be all of the same form but that the ones coming out earlier are narrower than those coming out later. Measurements show that the widths at half-maximum become larger in a continuous manner as one goes from left to right in the data. This lack of shift-invariance will affect the accuracy of the results only slightly. This will be discussed in more detail subsequently.

Taking the area under each peak as a measure of the amount of each material present, how this area would be changed, if any, under deconvolution

needs to be investigated. To look at this problem in its simplest form, free of all nonessentials, consider two isolated peaks of any arbitrary shape and size. The areas of each would be given by, respectively:

$$\int_0^{2p} f_1(t)dt \quad \text{and} \quad \int_0^{2p} f_2(t)dt$$

The ratio of their areas would be:

$$\frac{\int_0^{2p} f_1(t)dt}{\int_0^{2p} f_2(t)dt}$$

Convolving each of these functions with some arbitrary machine function, g(t), would yield:

$$h_1(t) = \int_0^{2p} f_1(t)g(x-t)dt$$

for $f_1(t)$ and an expression of similar form for $f_2(t)$. The ratio of the areas after convolution would be:

$$\frac{\int_0^{2p} h_1(x)dx}{\int_0^{2p} h_2(x)dx} = \frac{\int_0^{2p} \left[\int_0^{2p} f_1(t)g(x-t)dt \right] dx}{\int_0^{2p} \left[\int_0^{2p} f_2(t)g(x-t)dt \right] dx} \qquad (2.12)$$

As they are both of similar form, only one expression will be considered.

$$\int_0^{2p}\left(\int_0^{2p} f_1(t)g(x-t)dt\right)dx = \int_0^{2p} f_1(t)\left(\int_0^{2p} g(x-t)dx\right)dt$$

The expression on the right was obtained by interchanging the order of integration. The integral over g(x-t) will be a constant as it gives the area under g which will remain the same regardless of how much it is shifted along the axis by t. Let this integral be denoted by K. A similar result will be obtained for $h_2(t)$ and Equation (2.12) may be written:

$$\frac{\int_0^{2p} h_1(t)dt}{\int_0^{2p} h_2(t)dt} = \frac{\int_0^{2p} f_1(t)Kdt}{\int_0^{2p} f_2(t)Kdt} = \frac{K\int_0^{2p} f_1(t)dt}{K\int_0^{2p} f_2(t)dt}$$

$$\frac{\int_0^{2p} h_1(t)dt}{\int_0^{2p} h_2(t)dt} = \frac{\int_0^{2p} f_1(t)dt}{\int_0^{2p} f_2(t)dt} \qquad (2.13)$$

and the same ratio of areas as before convolution is obtained. Thus the same ratio of peak areas will be obtained under deconvolution. The case where the peaks are merged needs to be considered. It should be apparent after a little thought that if two or more peaks of unknown shape are merged there is no possible way to determine how much of the smeared together area to assign to each distribution. This situation is actually no different from the one where the peaks are well separated, but only more vividly illustrates the problem. The deconvolution process requires closer examination.

From the convolution theorem, the coefficients of the deconvolved function are given by:

$$C_n^f = \frac{1}{2p} \frac{C_n^h}{C_n^g} \tag{2.11}$$

Suppose the areas are arbitrarily assigned under two smeared together peaks and the complete function written as the sum of two others:

$$h(t) = h_1(t) + h_2(t)$$

By the addition theorem the transform would be:

$$C_n^h = C_{n_1}^h + C_{n_2}^h$$

where $C_{n_1}^h$ and $C_{n_2}^h$ are the transforms of $h_1(t)$ and $h_2(t)$, respectively, Dividing this by the transform of the machine function, we get the coefficient of the deconvolved function.

$$C_n^f = \frac{1}{2p} \frac{C_n^h}{C_n^g} = \frac{1}{2p} \frac{C_{n_1}^h + C_{n_2}^h}{C_n^g} = \frac{1}{2p} \frac{C_{n_1}^h}{C_n^g} + \frac{1}{2p} \frac{C_{n_2}^h}{C_n^g}$$

Taking the inverse transform and making use of the addition theorem, we get for the deconvolved function:

$$f(t) = f_1(t) + f_2(t)$$

where $f_1(t)$ and $f_2(t)$ are the deconvolutions of $h_1(t)$ and $h_2(t)$, respectively. Suppose the areas under the peaks were assigned such that two other quite different functions are obtained.

$$h(t) = h_3(t) + h_4(t)$$

After deconvolution the result obtained is:

$$f(t) = f_3(t) + f_4(t)$$

The deconvolved result is the sum of two quite different distributions. This is not a contradiction; only one result is obtained. For instance, if f_1 and f_2 were both zero on some region of the t-axis and f_3 was negative a certain amount and f_4 was positive by the same amount, then $f_3 + f_4$ would equal zero in that region as would $f_1 + f_2$, giving the same result in each case. Of course, the areas under each of these distributions would in general be different, so that if one didn't know the shapes of the two original peaks before smearing, then there would be no way of determining the correct areas. However, if the correct shapes of the peaks were known, even if they were smeared together, then one could deconvolve them and know how much area to assign to each peak by examining the deconvolution of each peak separately.

For this particular problem it is known within narrow limits what the peak shapes should be. All are reasonably uniform except for small variations in narrowness. If, then, a machine function can be found that gives an isolated peak a deconvolution with small negative regions, and if the deconvolved peaks all have similar form, then one may accept the results with a high degree of confidence. The requirement of small negative regions for the deconvolved peaks is if the deconvolution gives well separated peaks with no negative regions, then there will be no problem of cancellation of the area of adjacent peaks by these negative regions and hence there will be no problem in assigning the areas under the peaks.

For the sake of completeness, a deconvolution of a linear, shift-invariant set of data will be included.

Any number of distributions making up the data could be expressed in the

following functional form:

$$H(t) = ah(t-t_1) + bh(t-t_2) + ch(t-t_3) + \ldots \tag{2.14}$$

where a, b, c, etc., t_1, t_2, t_3, etc., are constants. Apply the convolution theorem, and employ the addition and shift theorems:

$$C_n^F = \frac{1}{2p}\,\frac{C_n^H}{C_n^g} = \frac{a\exp(-ni\pi t_1/p)\,C_n^h + b\exp(-ni\pi t_2/p)C_n^h + \ldots}{2pC_n^g}$$

$$C_n^F = \left(a\exp(-ni\pi t_1/p) + b\exp(-ni\pi t_2/p) + \ldots\right)\,\frac{1}{2p}\,\frac{C_n^h}{C_n^g} \tag{2.15}$$

where C_n^h is the transform of h(t). Transforming back to the time domain would yield the expression:

$$F(g) = af(t-t_1) + bf(t-t_2) + cf(t-t_3) + \ldots \tag{2.16}$$

where f(t) is the deconvolution of h(t).

So, if a machine function can be found that gives a minimum of negative regions for the deconvolution of an isolated peak, then the correct deconvolution of the data will give peaks all of the same shape, differing only in their heights. Note also that the location of the peaks will remain unchanged under deconvolution.

CHAPTER III

DECONVOLUTION OF GAS CHROMATOGRAPHIC DATA

The general method of performing deconvolution is as follows: One must initially obtain a suitable apparatus function. A transform is then made of this machine function and of the data. The transform of the data is divided by the transform of the machine function. Equation (2.11) is used for this.

$$C_n^f = \frac{1}{2p} \frac{C_n^h}{C_n^g} \tag{2.11}$$

These coefficients are then transformed back to the time domain to give the resolved peaks. The small imaginary part of these figures is due to calculations roundoff and can be neglected.

If the entire transform after division is transformed back to the time domain, the result, quite apparently, is nonsense. An inspection of the transforms reveals why this is the case (see Figure 3). Note from the graphs of the machine function and the data that the coefficients of the lower frequencies are the only appreciable ones (see Figure 4). They begin rather large at the lowest frequencies and fall off in a regular manner. It is apparent in nearly all cases that if this trailing off of the function were extrapolated into the "noise" it would approach zero rather quickly, certainly in all cases before at most a hundred coefficients were taken. The coefficients falling to zero this quickly reveal that no information contained in our data was lost, getting at least two points per wavelength on the highest frequency present (see Appendix B where the sampling theorem is discussed further). By its relation to the continuous Fourier transform, it is apparent that the

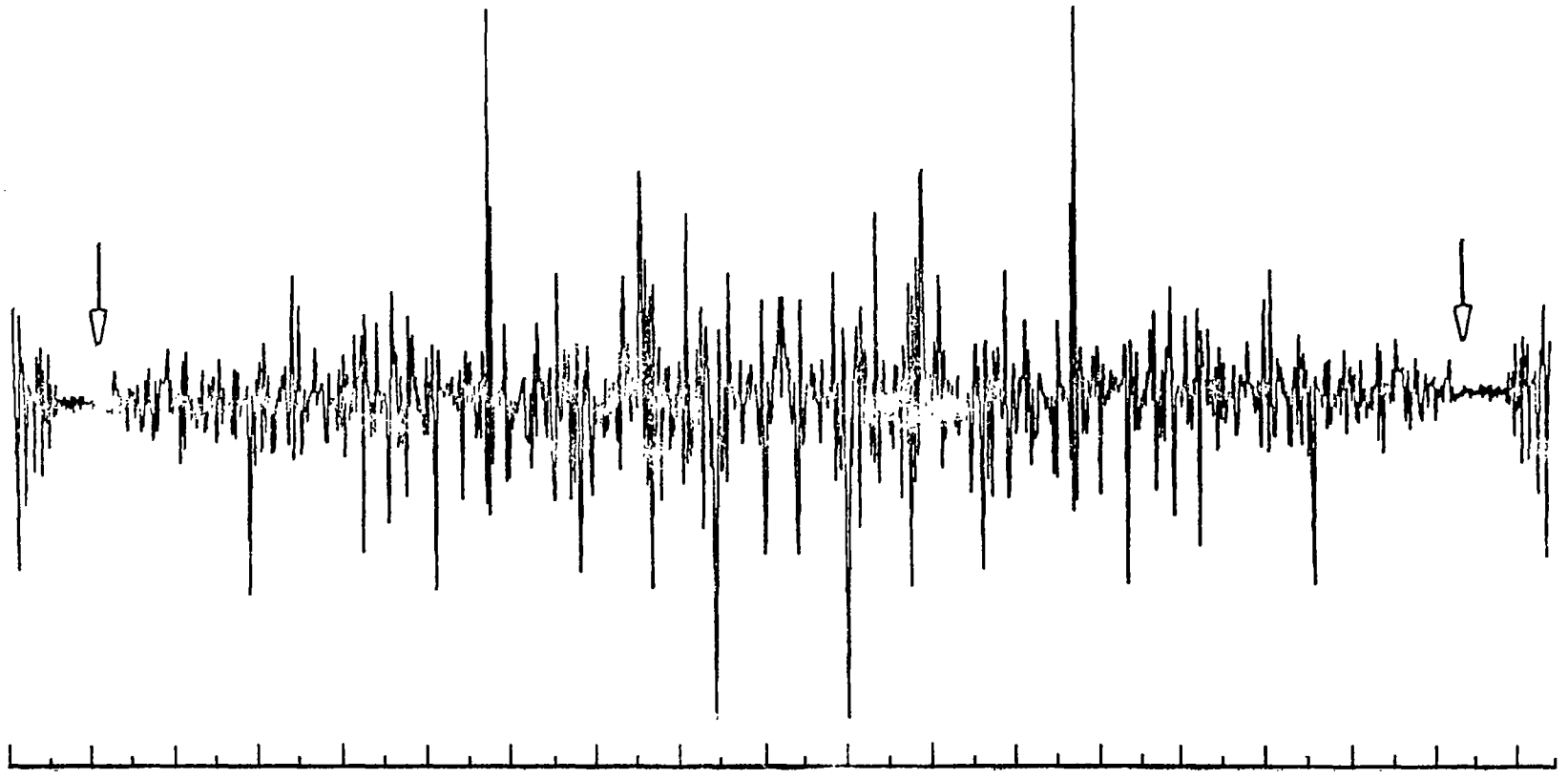


Fig. 3. Graph of the real parts of the complex numbers that result from division of the coefficients of the data by coefficients of the machine function. These coefficients are truncated at some point (see arrows) and transformed back to the time domain to give the deconvolution.

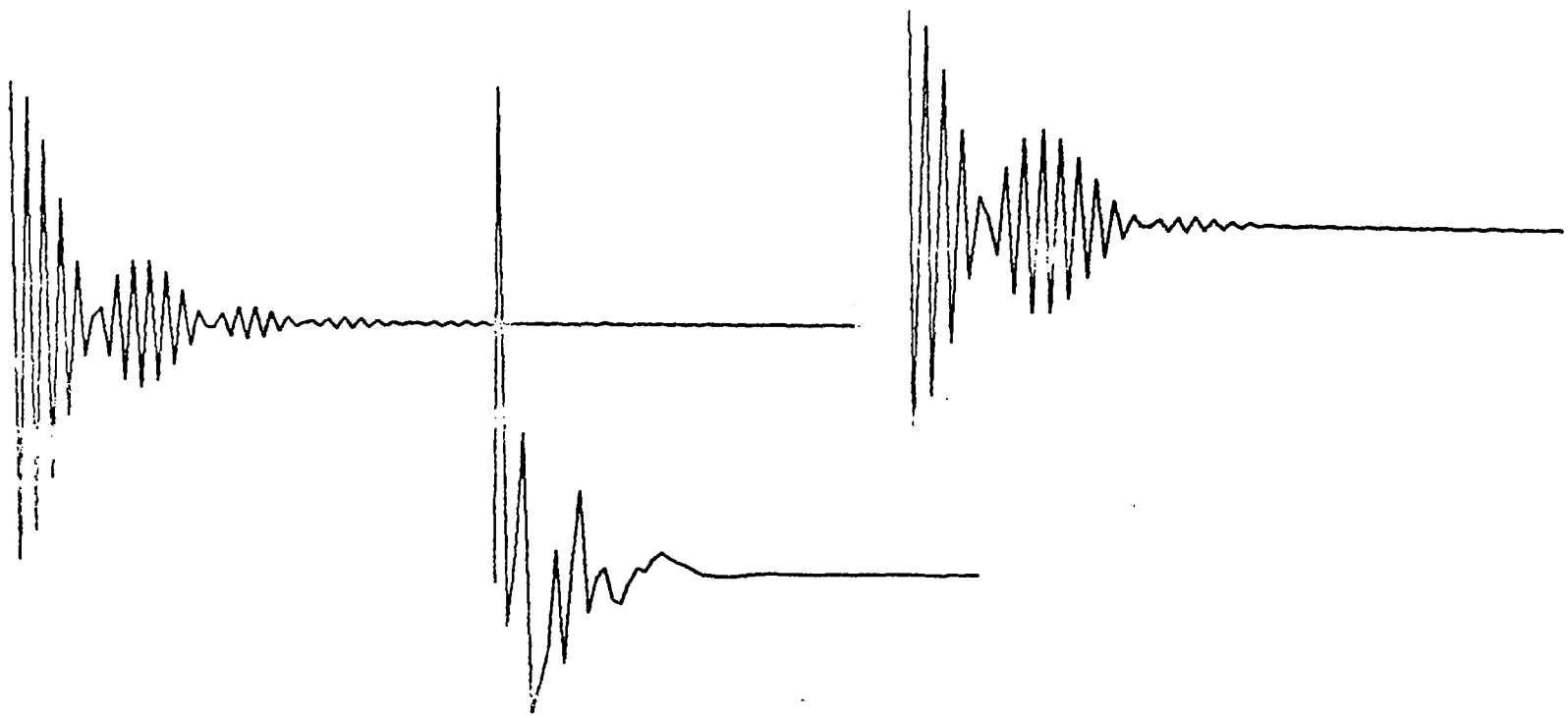


Fig. 4. Graphs of coefficients of the data and two machine functions: (a) is graph of real part of complex coefficients of the machine function, (b) is real part of coefficients of data, and (c) is real part of coefficients of a gaussian function (not centered at zero).

envelope of the transforms of gaussians are gaussian also and would fall to zero rather quickly.[14] However, one notes by examining the transforms that a small residual "noise" extends the whole length of the spectrum. This is due to machine roundoff in the computer, noise and error in the original data, and digitizing of data. This noise seems to be completely random and as much positive as negative, so it is not necessary to raise or lower the spectral function to some "baseline."

Digressing for a moment, the data and the manner in which it was recorded should be considered. Due to the mechanical and electrical properties, all recording instruments have a "smoothing" effect on the recorded data. Speaking in terms of the Fourier transform, one would say that the higher frequencies are filtered out. As discussed in the introduction, the area under the peaks is an accurate but not exact measure of the amount of material present. If the data were not smoothed and the recorder accurately measured the amount of material passing through an arbitrary cross section of the tube at every instant of time, the data would show statistical fluctuation due to the thermodynamic nature of the variables concerned. The smoothed data recorded is some sort of "average" of these statistical fluctuations and the error involved would in most cases be small compared to other inherent errors. Thus, for the purpose of taking the area under the peaks, the smoothed, or "bandlimited" form for the data and the deconvolved peaks will be sufficiently accurate for the purpose. If this is the case, the principal solution for deconvolution will be the correct solution. Practically, the presence of noise complicates the situation. The true spectrum of the machine function and its associated deconvolution buried in the noise will be unknown. This will necessitate truncation of the spectrum at some point before the true value of

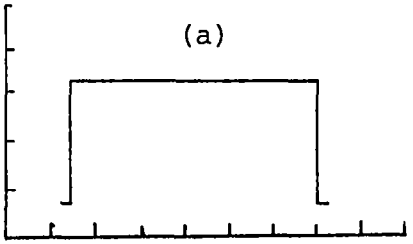
C_n^g becomes zero.

The reason the noise assumes such significance is because of the division of one frequency spectrum by another. One may accept with confidence the results in the region where both coefficients are appreciable. However, in the region of the noise, this division, due to the random nature of the noise, can give rise to enormous numbers. Note from the graph that the largest numbers are in this region. These results are completely meaningless and can quickly overwhelm the correct values if very many coefficients are taken as one extends into this region (see Figure 14). The results obtained when the dividing function is small represents one of the largest problems encountered in practical deconvolution. It is quite apparent that these frequencies must be "cut off" before getting very far into this region. This is referred to as "window closing" in the literature.[15]

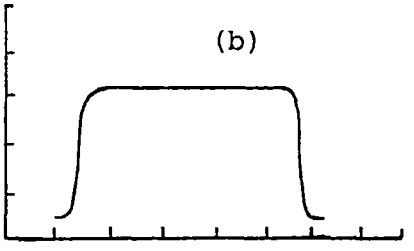
Various methods of dealing with this problem have been suggested. Some merely accept the results as they are, taking it as being accurate enough for their purposes. Others, desiring more accurate results, employ the use of one of the many "data windows" available (see Figure 5).[16, 17] (See Appendix A for a discussion of data windows.)

A Hamming window was applied in the frequency domain with only moderate success (see Figure 6). Most of the problem was due to widening of the transformed peaks. Such a strong deconvolution had to be taken in order to offset this widening that some distortion was introduced. There was no improvement over deconvolution with window not applied. Other workers in the field have used data windows, however, with some measure of success.[18]

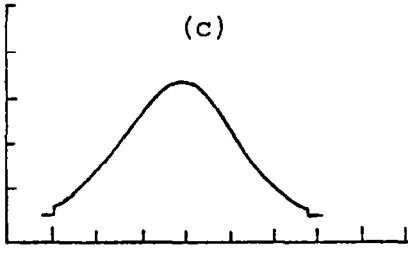
A data window, however, is an artificial device and its determinations only fortuitously resemble the original function because of its reduction of



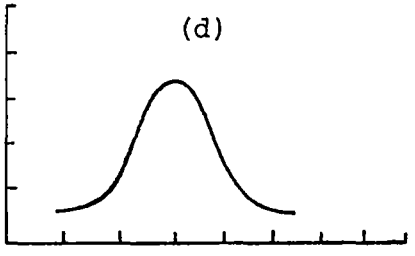
A = 1 for t = 0 to T



A=.5(1-cos2π5t/T)for T=0 to T/10
and t=9T/10 to T
A=1 for t=T/10 to 9T/10



A=.08+.46(1-cos2πt/T) for
t=0 to T



A=1-6(2t/T-1)^2+6|2t/T-1|^{3}for t=
T/4 to 3T/4
A=2(1-|2t/T-1|)^{3}for t=0 to T/4
and t = 3T/4 to T

Fig. 5. Four common data windows. Data is taken over the interval (O,T). (a) Rectangle. (b) Extended cosine bell. (c) Hamming. (d) Parzen.

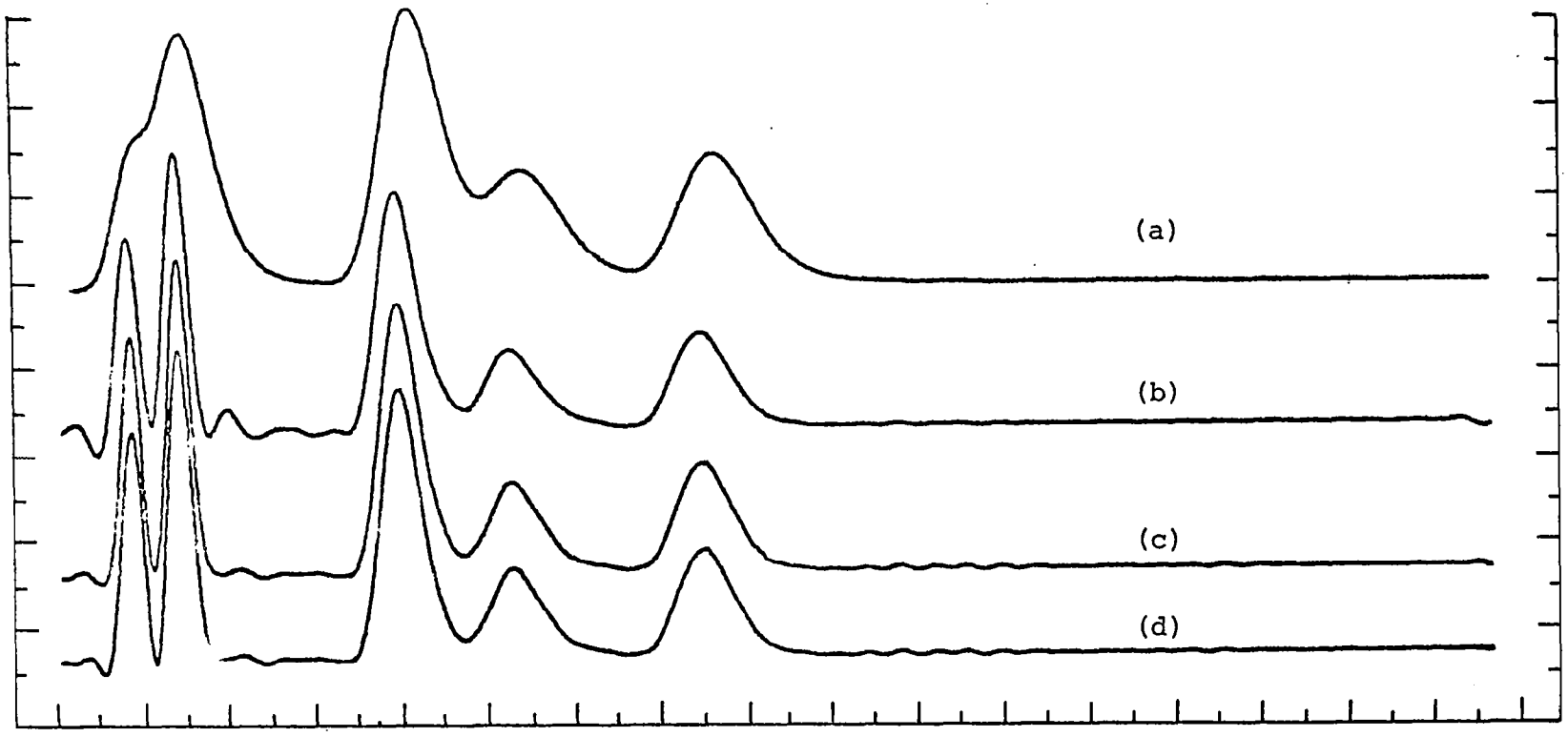


Fig. 6. Deconvolution with Hamming window applied compared to straightforward deconvolution: (a) is data of Run 5, (b) is deconvolution with Hamming window applied, (c) is deconvolution with function continuation, and (d) is straightforward deconvolution.

the sidelobes. The peak shape is also usually changed slightly. To achieve any substantial improvement in deconvolution, more precise information about each particular problem is needed. Specifically, if one knew exactly what the coefficients were supposed to be beyond the point of cutoff, the perfect deconvolution should be obtained with no sidelobes. A measurement of the wavelength of the sidelobes showed that it corresponds closely with the wavelength of the cutoff frequency. This suggested that the sidelobes and other distortion were possibly caused by "cutting off" the frequencies at this point. However, this turned out to be only partially correct. Most of the error was found to stem from a poor choice of the machine function. This will be discussed later. For the following discussion, the assumption will be that the deconvolution was performed with some "ideal" machine function. This was realized in practice, to a large extent, by later deconvolutions.

Even with a good machine function, truncation of the frequencies will cause error; there will be sidelobes. The main thrust of the ensuing argument is that if a function could be found which resembled the deconvolved function closely enough, such that the coefficients matched up with the coefficients of the original deconvolved function, especially in the region close to the point of cutoff, then the coefficients should be close in numerical value to what the coefficients of the original function would have been had they not been cut off. To test this hypothesis, an artificial function was built from gaussian peaks of the same height, location, and width at half-maximum as the deconvolved function (see Figures 7 and 8). The transform was taken and the frequencies were cut off at the same point as the original function to see what effect this would have. The results were very encouraging. The sidelobes appeared at the same positions as the original deconvolution and had almost

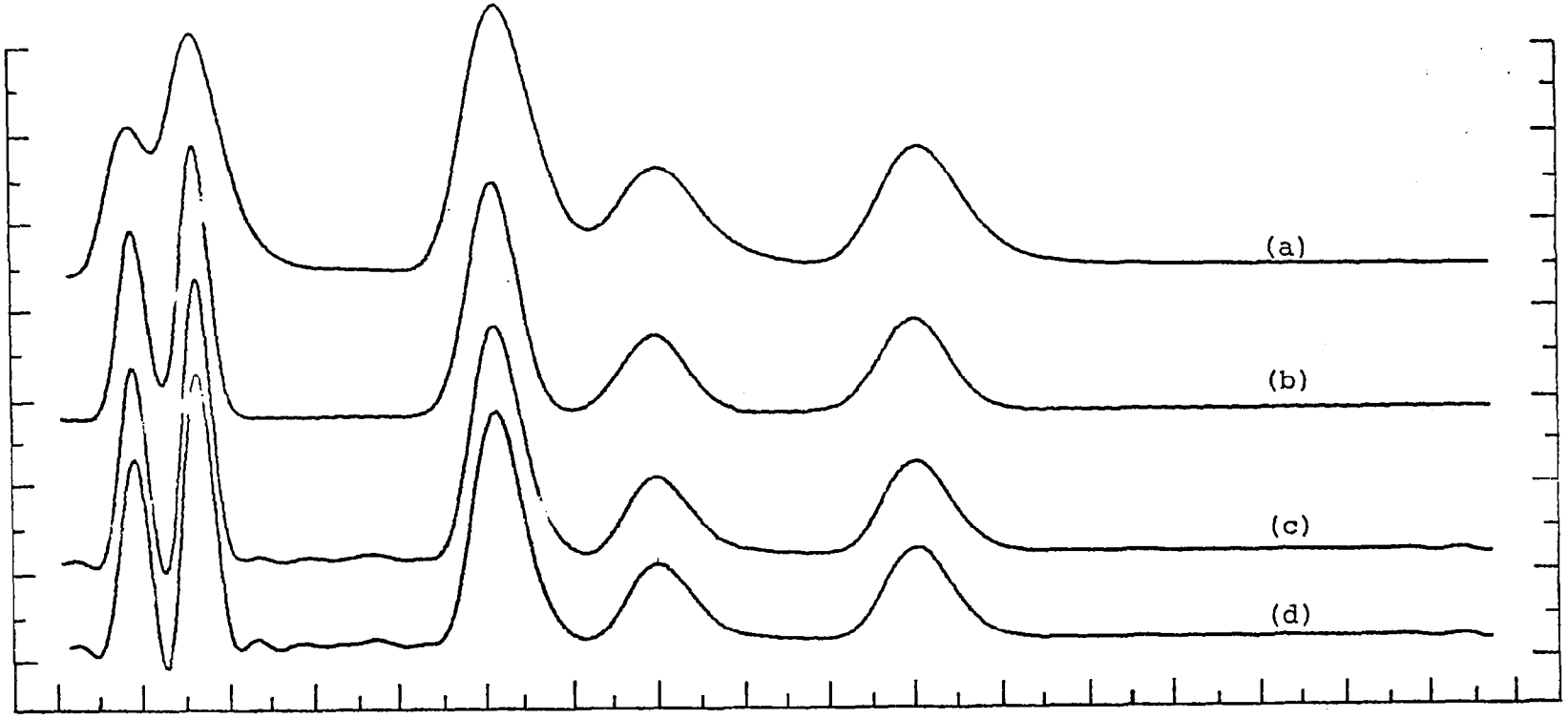


Fig. 7. Deconvolution with function continuation compared to straightforward deconvolution for Run 4: (a) is data of Run 4, (b) is artificial function, (c) is deconvolution with function continuation, and (d) is straightforward deconvolution.

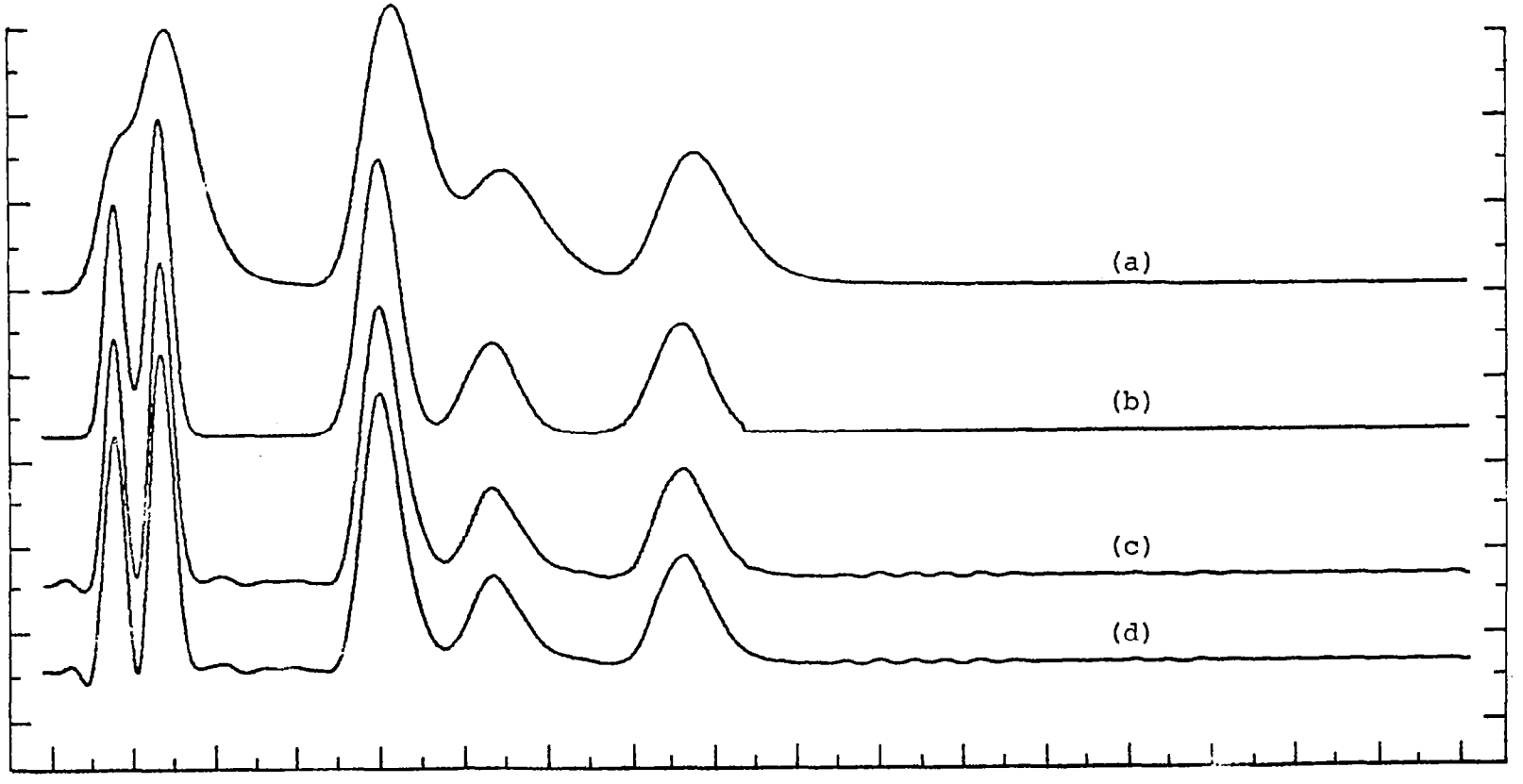


Fig. 8. Deconvolution with function continuation compared to straightforward deconvolution for Run 5: (a) is data of Run 5, (b) is artificial function, (c) is deconvolution with function continuation, and (d) is straightforward deconvolution.

the same amplitude. This indicated that most of the error remaining in the deconvolved data was due to the act of "cutting off" the frequencies. Coefficients of the artificial function were then substituted into the original transform from the point of cutoff on. The transform was then a hybrid with all of its original coefficients up to the point of cutoff, and with coefficients from that point up to a hundred coefficients "grafted" from the artificial function. The rest were set to zero as they originated in noise anyway. Transforming back to the time domain to observe the results: The results were very good for Run 4 (see Figures 7 and 9). The amplitude of the sidelobes were reduced by more than half. The results from Run 5 were not quite as impressive, only slight improvement being gained. However, the coefficients after forty-six terms do not contribute much "ripple" anyway (see Figures 8 and 10). A little more work on the machine function is probably needed here.

More work needs to be done in obtaining a good artificial function also. The matchup between the coefficients of the artificial function and the actual function was crude (see Figure 11). This suggests a more fundamental difference between the two than a comparison of the data reveals. Encouragingly, though, the match was as good in the region close to the cutoff point as at any other region. The actual peaks were different from guassians in that they were narrower close to the base. A promising field of research would be the investigation of the efficacy of various peak shapes for artificial functions in deconvolution. One item that needs to be pointed out here is that the artificial function should resemble the original deconvolution after cutting off its frequency spectrum, not before. This would involve some adjusting as the solution is converged to. At any rate, the match worked well enough

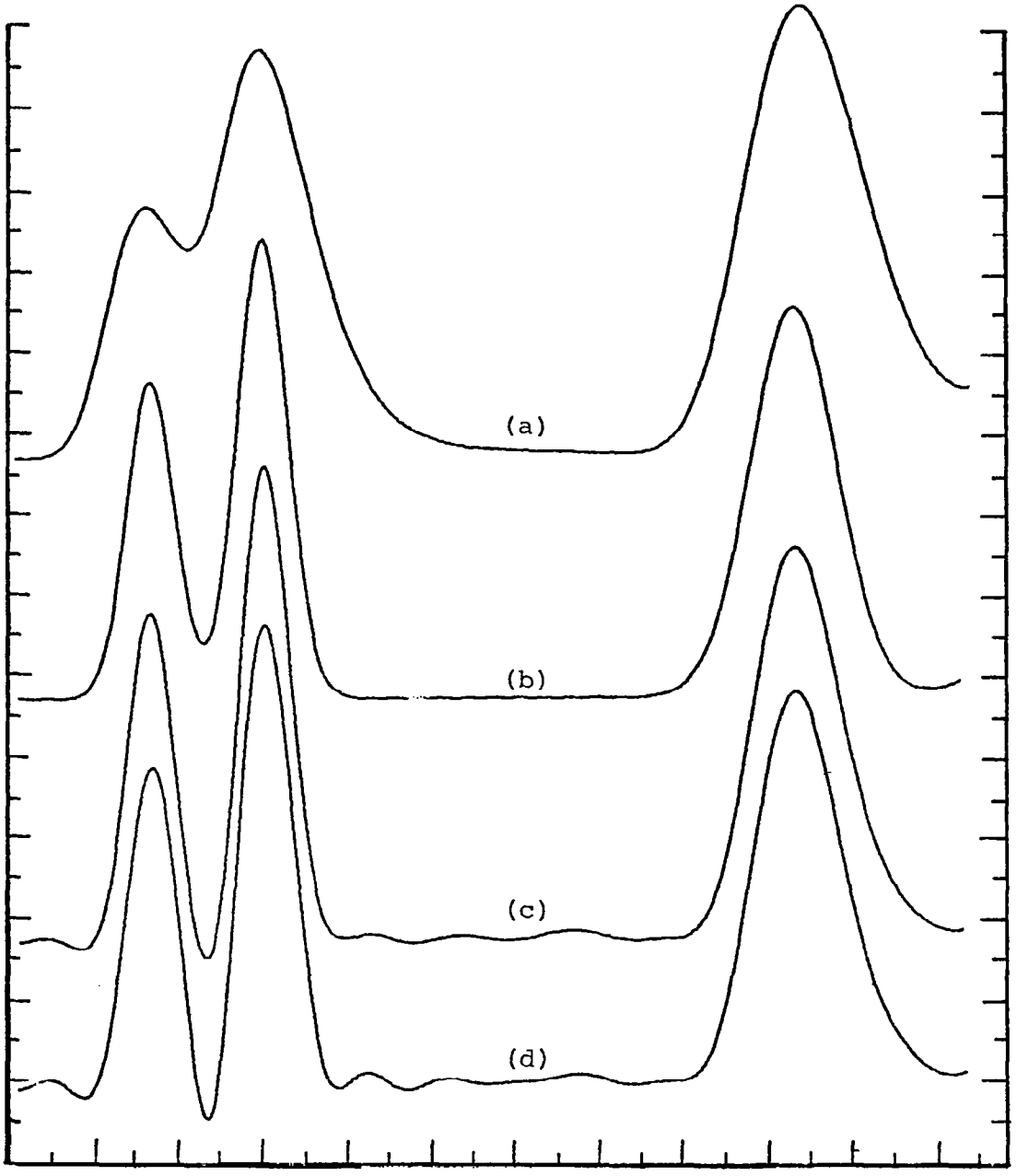


Fig. 9. An enlarged view of the first three peaks of Run 4: (a) is data of Run 4, (b) is artificial function, (c) is deconvolution with function continuation, and (d) is straightforward deconvolution.

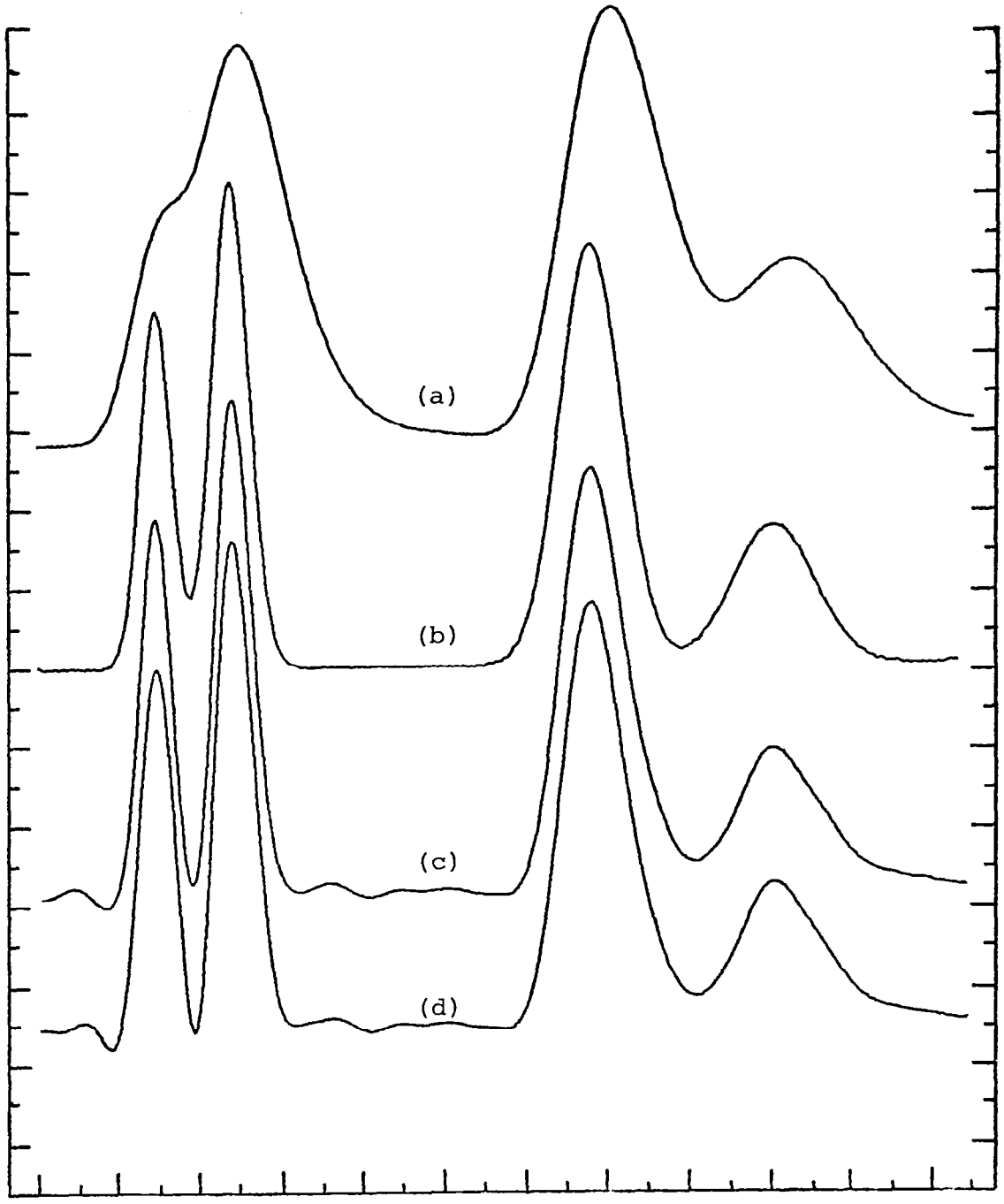


Fig. 10. An enlarged view of the first four peaks of Run 5: (a) is data of Run 5, (b) is artificial function, (c) is deconvolution with function continuation, and (d) is straightforward deconvolution.

Fig. 11. Coefficients of the artificial and deconvolved function. "HOUT" contains the coefficients of the deconvolved function and "AROUT" contains the coefficients of the artificial function.

NAME = HOUT

.201841734344547D 02	.0000000000000000D 00
.326864335342944D 01	.8206243128879270 01
.315513295077103D 00	-.2399258393917180 01
-.447552833583300 01	.5616999255556560 01
-.105697546035332D 02	-.5084771850635090 01
.564344816256069D 01	-.1992944980516280 01
.201863899871019D 01	.7622484671209720 01
.259218966695231D 01	.3263942412206040 01
.290686296584605D 00	.6350827756573050 01
-.649925330930191D 01	-.8796624363708870 00
-.232586448470689D 01	-.6290082962896300 01
.130179814038466D 01	-.3431944692213230 01
.515779866516200 00	.2020369838598430 00
.740113995502188D 00	.1321992074209820 00
.378571278257526D 00	-.3688890854579750 01
-.432442977166412D 01	-.7494727152232510 00
.300647330536524D 01	-.3336216418391320 01
.102139497020750 00	.2579572486559020 01
-.132173580371190D 00	-.3374636803779470 01
.353200194555490D 01	.2984998097219800 01
-.456151989992719D 01	.9033572498546480 00
.210389797684407D 01	-.2857451960800239D 01
-.624423554397115D 00	.3071358228589100 01
-.130158365139623D 01	-.3163724001615780 01
.310805557193679D 01	.1465199280202660 01
-.281117552436823D 01	.9137592033754840 00
.130383188559160D 01	-.1874666589167420 01
-.987955892025272D-01	.1862675684478900 01
-.891290241818138D 00	-.1318708930647490 01
.121175693364077D 01	.2802959245863480 00
-.729244612755146D 00	.3541892043653740 00
.232494203119825D 00	-.5144546739677990 00
.153951070974315D 00	.2664438661816790 00
-.222067637174278D 00	-.6167369506129500-01
-.684165729376068D-01	-.1844550088806360 00
.394749189112762D 00	-.5871840787572070-02
-.371714035454455D 00	.1839372936604740 00
.216704112281573D 00	-.3135047878157490 00
-.367559811124257D-01	.6022863724441700 00
-.351457236324972D 00	-.3978741885132980 00
.422907732704773D 00	.8210310388820560-01
-.597140835744506D 00	.1263901572404850 00
.382721588740852D 00	-.4334228888094040 00
-.505127392886550D-01	.2615147373632350 00
-.141140768646243D 00	-.3046873499613130 00
.374576781188669D 00	.3835158671666810-01
-.346938051017286D 00	-.1151142573908330-01
.401854965874133D 00	-.1473721065944530 00
-.121702852993994D-01	.3460503265118790 00
-.106341387724859D 00	-.1083498030738200 00
.144449362946721D-01	.2356777680602910 00
-.126505668193953D 00	-.5780363735803230-01
-.796081894163097D-01	.1975753680470930 00
.161975248908475D 00	.5191938872720810-01
-.439963384232826D 00	-.3764181424482410-02
.166496918751911D 00	.1621854258767310 00
.136922048026688D 00	-.2686978951596920 00
-.522949074496742D-01	.4302380686692570 00

NAME = ARGJT

.2054882812500000D 02	.0000000000000000D 00
.2573314219497120D 01	.8748693703522130 01
.4097484184047530 00	-.3075172372528770 01
-.4581097734298030 01	.5908051426342130 01
-.1100230078244950 02	-.5769887151398120 01
.6167113734153480 01	-.1683495772877810 01
.1134068676676850 01	.8363544878079900 01
.2012381183381910 01	.3065829424600900 01
-.2582567763351840 00	.6443575381584320 01
-.6811317815615870 01	-.1854402640883300 01
-.1283928033294060 01	-.6512716285495240 01
.1323095833468550 01	-.2853026324930490 01
.6716489384730000 00	.6317938051810640-01
.8890258065568470 00	.7532519053872350 00
-.2023841617208950 00	.3424054023726150 01
-.3838616565033290 01	-.9835887365740070 00
.2508213650717110 01	-.2963022405520990 01
.7834342896114530-02	.1838213320007430 01
.1818918363518640 00	-.2963291762524800 01
.3008400573684730 01	.2755616439245850 01
-.3871358579720330 01	.7596890414419500 00
.1866611668073490 01	-.2253095459135810 01
-.6655269722628670 00	.2633637936852330 01
-.1050965127175480 01	-.2726742504464200 01
.2561246127317720 01	.1194378630086470 01
-.2199798973107890 01	.7039812828255840 00
.1123011237366020 01	-.1403392321067350 01
-.1707162005878940 00	.1612122362644250 01
-.8135498480798370 00	-.1166155903525900 01
.1061908492438900 01	.2526796962668770 00
-.6956107993350610 00	.3473551296590770 00
.2462377623874830 00	-.4646480963568920 00
.8500727883353560-01	.3628554925572860 00
-.2044185638899730 00	-.8047936824856900-01
.5325861661966570-01	-.9979805597034120-01
.1208948619582400 00	.6377138912652310-01
-.1604311070519410 00	.1041846660767330 00
.6768983556676600-01	-.2189657722402880 00
.1024000886888730 00	.2585228667967420 00
-.2735304299941570 00	-.1280847275451920 00
.2798173671819840 00	-.5923087938570880-01
-.1837933932099110 00	.2197435757207560 00
.1103323664790650-01	-.2577056296806000 00
.1277138673829360 00	.2051224988889350 00
-.2124453159914580 00	-.6282615217519050-01
.1671950587952360 00	-.5113150223407780-01
-.9560832765049370-01	.1310897612541890 00
-.5204869302995050-02	-.1256437277058800 00
.5607289687965140-01	.8651869447469680-01
-.8218672280487130-01	-.1802032420855050-01
.5479659049301250-01	-.7688802594275110-02
-.2968304117319020-01	.4088322142868360-01
.6909747143907470-02	-.3175354347040480-01
.1447806004157410-01	.2079023942253660-01
-.9078682793108070-02	.1265504062584690-03
.5742345837191590-02	.1398290123885000-01
-.9418068039711110-02	.8370993359248230-02
-.1381552111154779D-02	-.1526283875492320-02

in practice to considerably improve Run 4.

Also, there is no reason why substitution of coefficients should begin at the point of cutoff. More and more coefficients from the artificial function may be substituted until a more desirable result is obtained (see Figures 12 and 13). However, one must keep in mind in this case that the final peaks will take on more of the character of the artificial function.

Finding a good machine function will probably be the largest problem encountered. A gaussian for the machine function was easiest to implement, but left curious "shoulders" on the right side of some of the peaks, giving the impression that there were additional peaks that could be revealed by taking a stronger deconvolution. This was highly unlikely, however, as there was no other evidence for this. It was suspected that the shoulders were due to not considering the skewness of the peaks.

In the next attempt, several isolated peaks in Runs 1 and 2 were tried as machine functions. Narrowing the peaks was easy to implement in a computer program by selecting every other, or every third data point, or some other interval. Interpolation between the points was introduced later to obtain a finer gradation. The amount of resolution desired in the final deconvolution is obtained by adjusting the narrowness of the machine function.

However, only small improvement over a guassian was gained by using isolated peaks (see Figure 14). It was apparent that a function with more "skewness" was needed. There are several functional forms, all referred to generically as "skewed gaussians" that have been used successfully by other workers in the field.[19] The method used with success in this research was to choose an isolated peak, which was close to the desired shape anyway, and give it an additional "skewness" (see Figures 7 and 8). This was accomplished

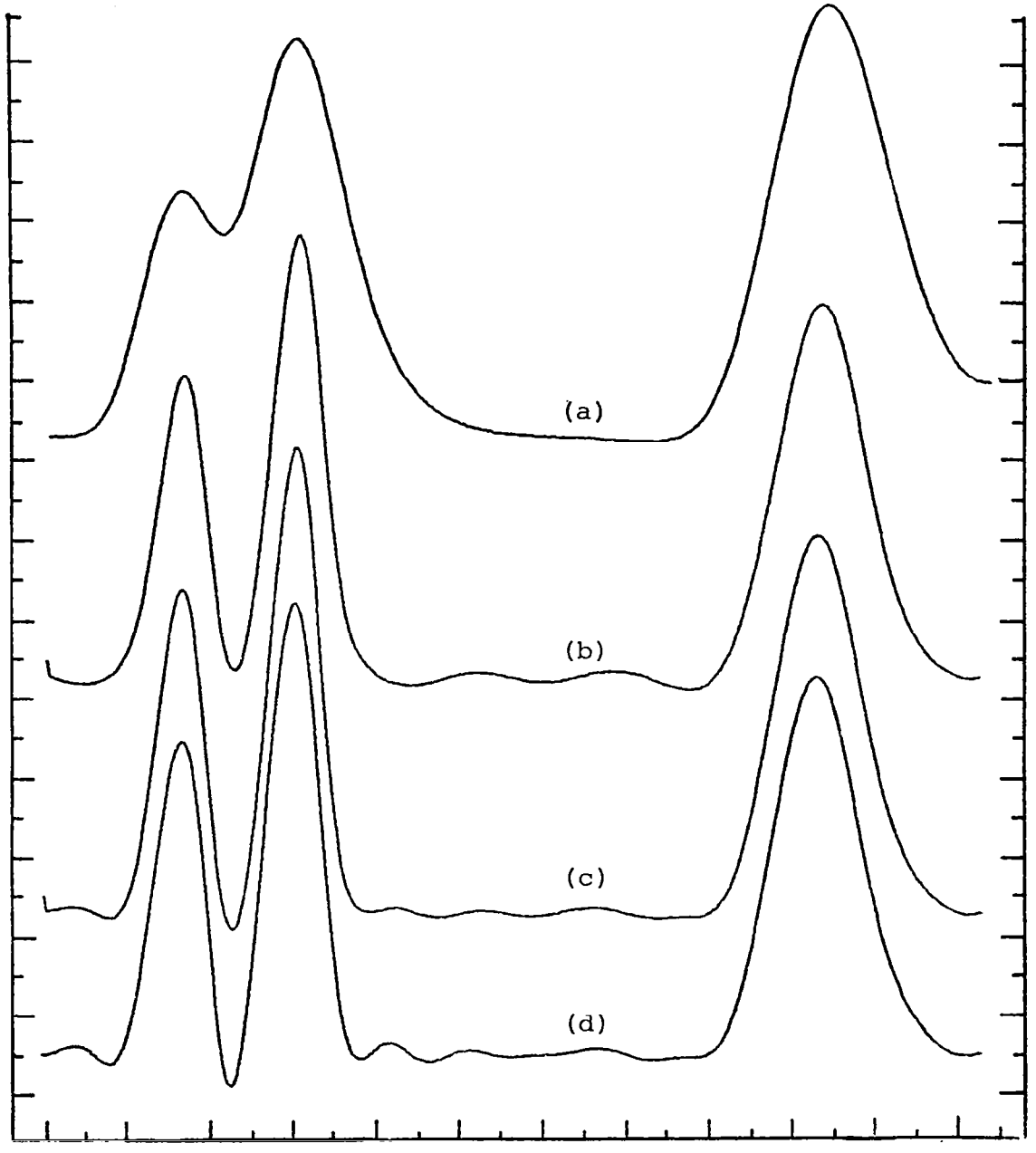


Fig. 12. Comparison of two different cases of function continuation for Run 4: (a) is data of Run 4, (b) is deconvolution with function continuation (20 to 100 coefficients substituted from artificial function), (c) is deconvolution with function continuation (35 to 100 coefficients substituted), and (d) is straightforward deconvolution.

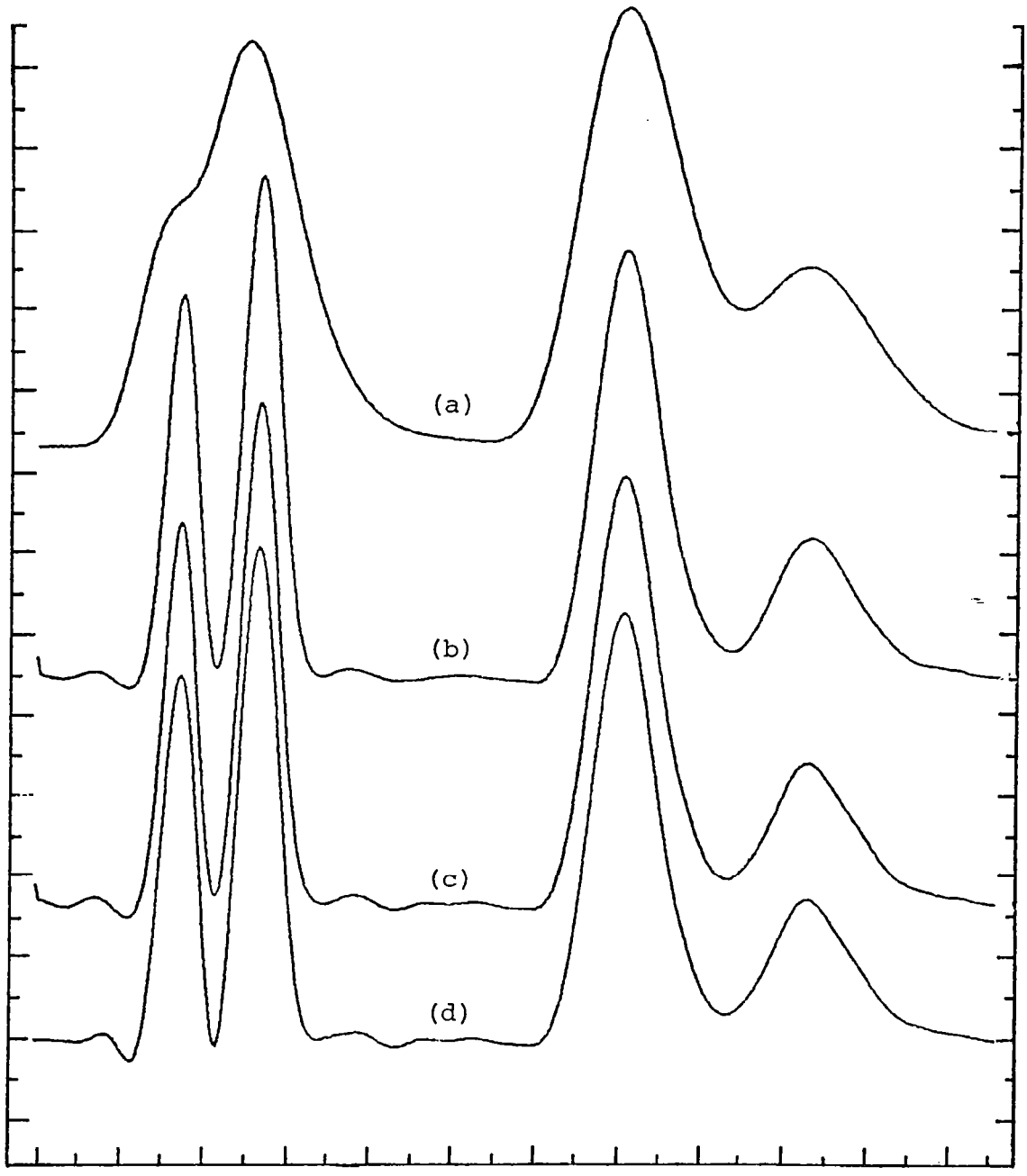


Fig. 13. Comparison of two different cases of function continuation for Run 5: (a) is data of Run 5, (b) is deconvolution with function continuation (37 to 100 coefficients substituted from the artificial function), (c) is deconvolution with function continuation (47 to 100 coefficients substituted), and (d) is straightforward deconvolution.

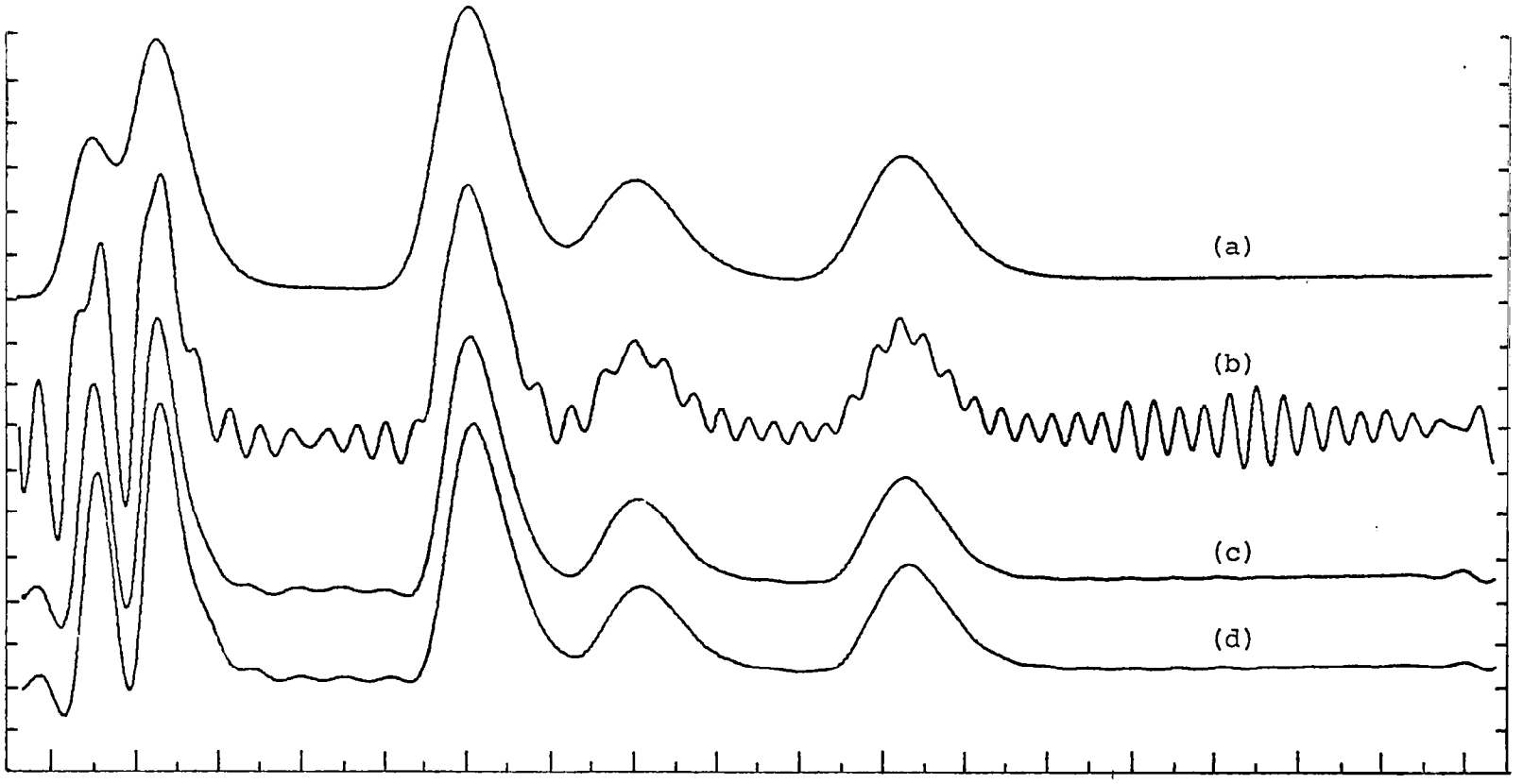


Fig. 14. Deconvolution with a gaussian and an isolated peak as machine function. Also distortion in deconvolution introduced when too many coefficients are taken: (a) is data of Run 4, (b) is deconvolution with coefficients truncated at 60 coefficients (distortion is considerable, (c) is deconvolution with 5th peak in Run 1 as machine function, and (d) is deconvolution with a gaussian as machine function.

as follows: Taking an origin at the center of the peak, the function was compressed behind the origin and expanded in front of it, the compression and expansion being accomplished by only one factor (see Figure 16). A computer program was written to implement this and also to adjust the narrowness of the peak. By using a convenient feature of Xerox extended FORTRAN, the INPUT statement, three parameters were input after the program was run which seemed to be the only ones necessary for good deconvolution. The first input controls the narrowness, the second adjusts the skewness, and the third shifts the function either left or right on the t-axis. Various combinations of these parameters are tried until a good deconvolution is obtained. What constitutes a "good" deconvolution is more fully discussed in Chapter II. If the sidelobes are comparable in size to that obtained from the artificial function when its coefficients are truncated at the same point, then one has good reason to believe that a limit has been reached beyond which one cannot go without using function continuation.

Also, one should keep in mind that a good deconvolution may not be unique. To reiterate, a "good" deconvolution for the purpose of gas chromatography is one in which the area associated with the deconvolution of an isolated peak is coalesced into one well defined "peak" and the negative regions are small so that there will be no problem in assigning the areas under the peaks. There may be a number of machine functions and their associated deconvolutions that satisfy these criteria.

An important consideration is that a good deconvolution seems to depend critically on small irregularities in the machine function (see Figures 15 and 17). Note from Figure 15 that when the machine function that drops to zero in the skewed portion is used, the graph of the deconvolution has larger

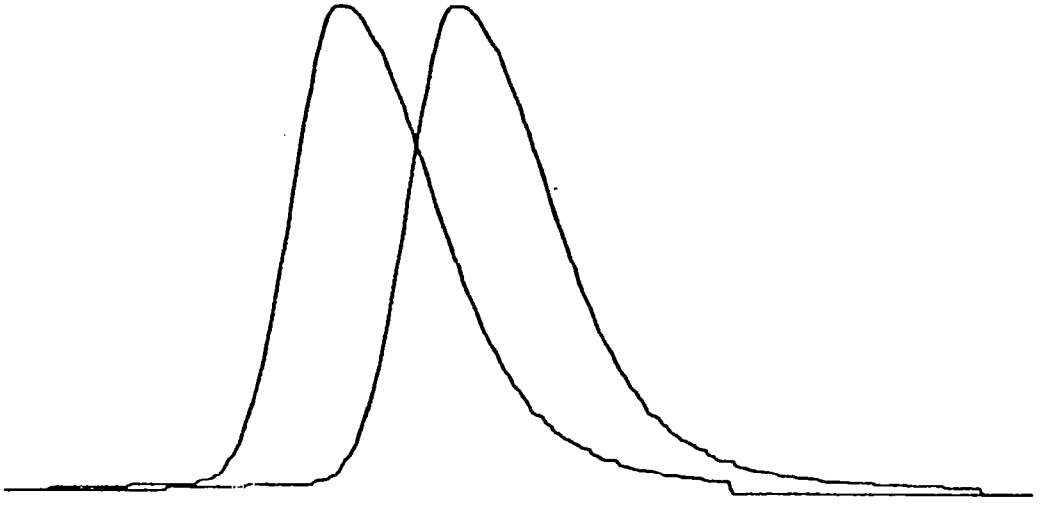


Fig. 15. Comparison of a smooth machine function to one with a sharp discontinuity. On left is machine function as program MACHFUN outputs it. Note the dropoff in the tail. On right is corrected machine function.

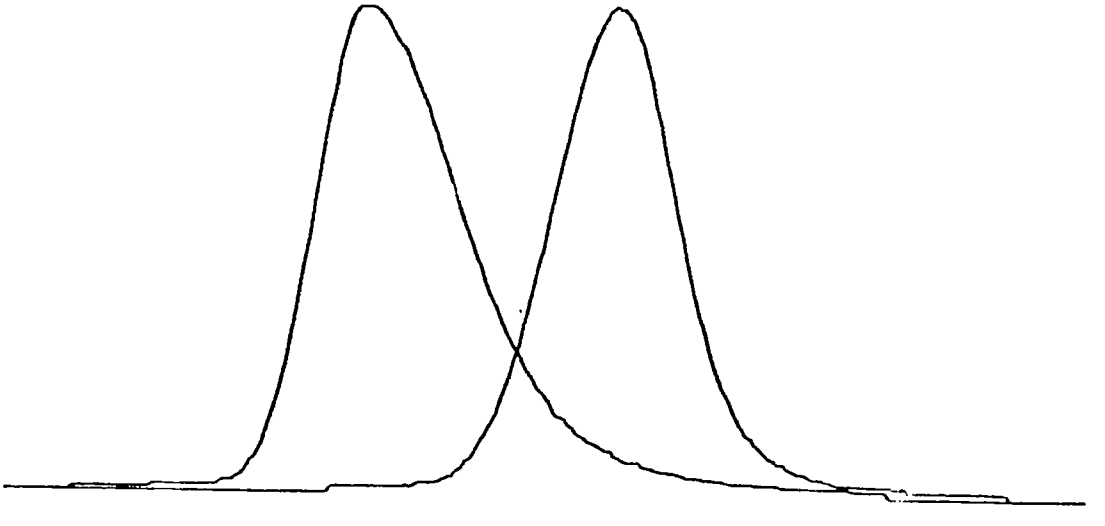


Fig. 16. Comparison of a skewed with an unskewed machine function. On right is 5th peak in Run 1. On left is same peak with "skewing." This was machine function used for Run 4. c = .4 for this peak.

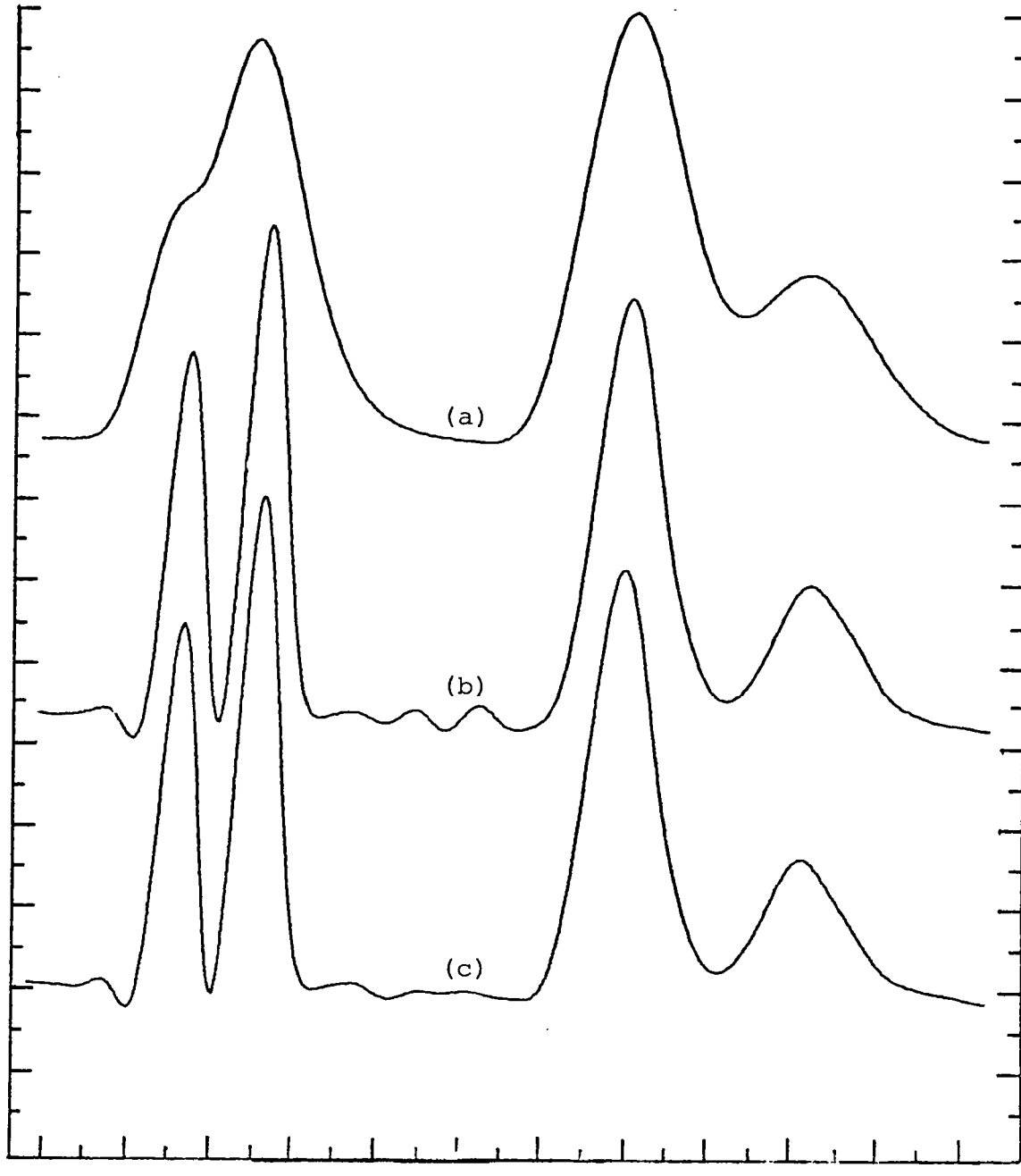


Fig. 17. Effects of irregular machine function on deconvolution: (a) is data of Run 5, (b) is deconvolution with irregular machine function (left peak in Fig. 15), and (c) is deconvolution with corrected machine function (right peak in Fig. 15).

sidelobes and shows more distortion than the case where the machine function is continued in an approximately linear fashion to zero.

To obtain a good deconvolution involves a certain amount of trial and error. An inspection of the outputs at each stage is usually necessary.

Once good deconvolved peaks were obtained, the areas under the peaks were acquired by numerical integration. Because of the sidelobes there were usually small negative regions on both sides of the peaks. The areas under the peaks were obtained by integrating between these negative regions as no other methods were known for accurately treating this particular situation. The suspicion is that if the sidelobes did not exist the peaks would probably be slightly-broader at the base and that the true area would be slightly greater than the values determined.

CHAPTER IV

ERROR ANALYSIS AND CONCLUSION

Error Analysis

The trapezoidal rule was used to determine the areas under the peaks.

$$\text{Area} = \left(\frac{1}{2} X_1 + X_2 + X_3 + \ldots + \frac{1}{2} X_{N-1}\right) \Delta t \qquad (A.1)$$

The data points were so close together that any difference between this and Simpson's rule would be so small that it could safely be neglected. Also, the endpoints were either very small or zero, so that an excellent approximation can be obtained by simply summing the values of the data points. The true areas would be obtained by multiplication by the proper proportionality constant.[20]

$$\text{Area} = \text{prop. const.} \cdot \left(X_1 + X_2 + X_3 + \ldots X_{N-1}\right) \qquad (A.2)$$

The results of any particular run are usually normalized; that is, the percentage composition is calculated by measuring the area of each peak and dividing the individual areas by the total area, e.g.,[21]

$$\% \text{ B} = \frac{\text{Area of B}}{\text{Total Area}} \cdot 100 \qquad (A.3)$$

The normalized calculations for all four runs are given in Table 2. The agreement of the percentage ratio of each peak among runs is relatively good,

even for the deconvolved results.

Areas of compounds are not directly proportional to the percentage composition, i.e., different compounds have different detector responses; therefore, it is necessary to determine correction factors. Once determined, these correction factors can be used to calculate the percentage composition.[22]

In an attempt to determine some idea of the inherent errors in the taking of gas chromatographic data, ratios were taken of the peaks in Run 1 and compared to the same ratios in Run 2 to see if there were significant differences. Only Runs 1 and 2 were compared because the peaks were well separated here. Although most compared quite well, one difference of 5 percent was observed. This means that it is difficult to get a very exact error analysis.

Other methods of determining the amounts of material present from overlapped peaks are triangulation, peak height measurement, and the dropping of a perpendicular at the lowest point of the valley between two peaks.[23] Obviously, none of these methods could apply to the first two peaks of Run 5 as the peaks are almost completely merged. This leaves deconvolution as the only method available for treating problems of this type. In Run 4 the first two peaks are merged to such an extent that none of the above methods are recommended.[24] The perpendicular drop method will be investigated to get some idea of the error involved. From Table 2 the ratio of Peaks 1 and 2 in Run 4 results in a value of 2.47. Deconvolution gives a value of 1.57. The ratio of the same peaks in Run 1 where the peaks are well separated is 1.64. The ratio of the deconvolved peaks differs from the ratio of the same peaks in Run 1 by 4 percent. The ratio from the perpendicular drop method differs from the ratio in Run 1 by 51 percent. This is much larger than one would expect from experimental error. In such cases one would prefer the

TABLE 2

AREAS UNDER THE PEAKS
(Normalized)

Sample Mixture	Percentage Composition					
	Run 1	Run 2	Run 4 (After Decon.)	Run 5 (After Decon.)	Run 4 (Before Decon.)	Run 5 (Before Decon.)
Compound A	13.010	12.768	13.372	13.418	9.814	{33.840
Compound B	21.371	21.428	20.960	20.827	24.219	
Compound C	33.856	33.470	33.082	33.115	32.125	31.950
Compound D	14.223	14.656	14.714	14.531	15.380	15.346
Compound E	17.541	17.678	17.871	18.108	18.462	18.865

deconvolved value if it were available.

Discussion of Results and Conclusion

One of the goals of this project was to develop methods for obtaining a good deconvolution with a minimum of time and effort. This was achieved in large measure by developing computer programs to handle all aspects of the problem and using INPUT statements to most conveniently vary the important parameters.

Striving for generality was another goal. Working in the timesharing mode and using FORTRAN exclusively will not be the methods everyone will use. However, the programs will not be too difficult to adapt to the minicomputer or other data handling machine the researcher may have. The methods used here will not have general applicability. They would be most appropriately applied to data where the deconvolved result would consist of a number of quite distinct "peaks." However, in cases where the machine function is known exactly, the concept of function continuation could probably be applied in some sort of "iterative" process that converges to the correct result. Perhaps the sidelobes could be identified to give a starting point.

The concept of function continuation in the frequency domain provides a twofold advantage. Not only does it serve to improve the deconvolution considerably, but, by examining the truncated artificial function, a basis for determining how close one is to the attainable limit in a straightforward deconvolution is obtained.

More work needs to be done in making good artificial functions as evidenced by the crude matchup of coefficients. That the peak shapes are not gaussian is obvious when the artificial function is overlaid over the deconvolved result.

One needs to find some distributions that can be expressed algebraically that closely resemble the peak shape. Alternatively, the shape of the deconvolved peaks may be changed slightly by changing the shape of the machine function.

The shape of the machine function used in this investigation was obtained by taking the top of a given peak as the origin, compressing the t-axis behind this origin and expanding the t-axis in front of it. The amount of compression and expansion was varied by inputting a value for the variable "c" in the computer program. This was expressed in the program as (x+c*x), which is algebraically the same as (c+1)x; c+1 is the linear factor of expansion and compression. It took the value 1.4 for the deconvolution of Run 4 and 1.36 for Run 5. It is easy to see how this might be generalized to give the t-axis any variable amount of expansion and compression, and hence give the machine function and shape desired, by using a general polynomial in place of this linear factor.

$a+bx+cx^2+dx^3+ \ . \ . \ .$

The constants a, b, c, etc. are varied to give the results desired. A different polynomial may be used for expansion from the one used for compression. Note that the expression above is a special case of this general polynomial with b=c+1 and a=c=d= . . . = 0.

Some of the programs are inefficient as far as computer time is concerned. No attempt has been made to make them more efficient nor to write them in more polished form. In particular, if it was all to be done over, the program for the machine function would be rewritten using a guassian rather than an isolated peak as a beginning function; thus eliminating the need for interpolating between the points.

APPENDIX A

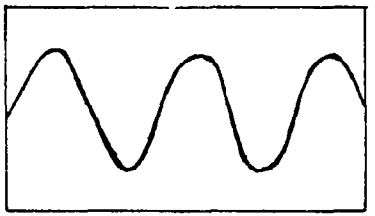
DATA WINDOWS

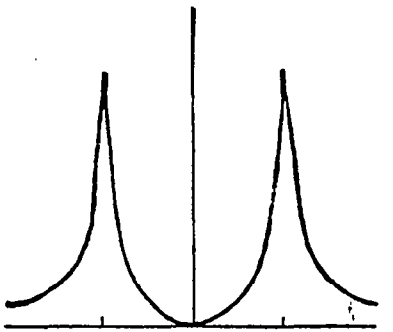
Data windows will be discussed only qualitatively here. For a more rigorous mathematical discussion, one is referred to the references.[11,17,25]

The frequency and time transforms are very similar in form, differing only by sign on the exponential term. The results obtained are nearly always of the same form regardless of in which domain we begin initially.[26] Thus, a data window will have essentially the same effect regardless of the domain in which it is applied. It is better interpreted when applied in the time domain so attention will initially be restricted to this.

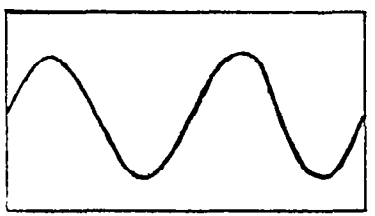
A knowledge of Fourier transforms shows that the sharper points of a function require high frequencies to represent it, an abrupt discontinuity taking frequencies ranging to infinity. The Fourier series is periodic so that if the function and its slope are not the same at its endpoints there will be an abrupt discontinuity and the frequencies will range to the highest ones present as the repeating series tried to join itself over the endpoint values. These frequencies usually group around the major frequency components making up the function and is referred to as "leakage" in the literature.[27] (See Figure 18.)

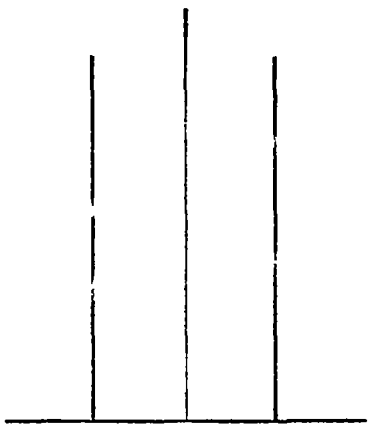
To more faithfully represent the frequency spectrum over the real, infinite time domain would require a data window over a longer interval of time; the wider the data window, the better the results. However, this is not often





(a) Function and its slope not the same at the endpoints. Note the additional frequencies introduced.





(b) Function and its slope the same at the endpoints. Transform is two symmetrical spikes.

Fig. 18. Cosine function and associated transform with data taken over two different intervals to illustrate leakage.

possible to realize in practice. Also if, by chance, the function and its first derivative happened to be the same at both endpoints, these additional frequency components would not be introduced. This happy situation seldom occurs naturally, however. One may, in a sense, bring this about, however, by extending the function at its endpoints in some smooth curve such that the function and its first derivatives are smooth and continuous across this region. Another way of reducing these additional frequency components is by using a "data window."

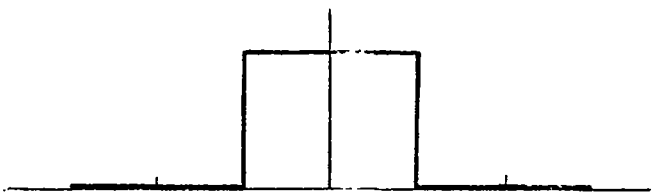
Actually, the operation of taking data over a finite range is referred to as applying a "rectangular" data window as it can be viewed as the multiplication of a rectangular function of unit height times the actual time function over the real, infinite time domain.

Data windows are functions that decrease to zero outside the range over which data is taken and are multiplied by the real, infinite time function. They all usually have a value of unit in the center and fall off in a regular manner to small values or zero at the endpoints. In all cases their purpose is to make the function and its first derivatives smooth and continuous at the endpoints.

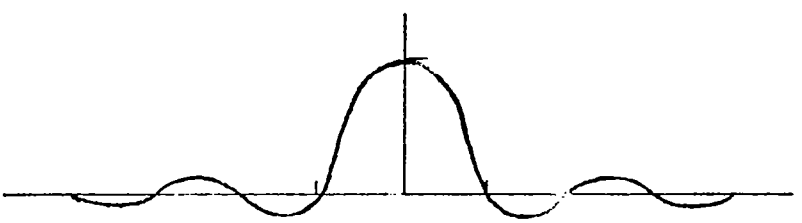
Note that the frequency spectrum introduced by a sharp discontinuity usually expresses itself in the form of "sidelobes" on both sides of the function. (See Figure 19.)

Fortunately, the data in this research is not beset with the problem of abrupt discontinuities. The data and its slope go smoothly to zero at both ends of the data.

In the frequency domain, however, the researcher is confronted with this abrupt behavior as the frequency spectrum is cut off at some point.



(a) Rectangle function, $\Pi(x)$.



(b) Transform of rectangle function, sinc x. Note the sidelobes.

Fig. 19. Rectangle function and its transform, sinc x.

As might be expected, sidelobes are obtained with the transformation to the time domain. A data window has the effect of bringing the frequency function and its slope relatively smoothly to zero or small values, and, hence, to reduce the sidelobes. As the data window tends to narrow the function somewhat, the effect in the transform domain is to broaden the peaks slightly. This broadening, which varies with the type of data window used, is one of the worst disadvantages of these windows. Also the shape of the peaks may be distorted slightly.

The above statements, apparent for a single distribution, may not be so obvious when the data consist of a number of peaks at differing locations. But if one considers that the transform of a sum is the same as the sum of the transforms (addition theorem) and that a shift in location means only multiplication of the transform by a "phase factor" of unit amplitude, then it is easier to see how all the peaks will be affected in the same manner.

APPENDIX B

THE DISCRETE FOURIER TRANSFORM

The complex Fourier series representation of a function defined over the interval (0,2p) is given by:

$$f(t) = \sum_{n=-\infty}^{\infty} C_n \exp(ni\pi t/p) \tag{2.1}$$

where the complex coefficients, C_n, are given by:

$$C_n = \frac{1}{2p} \int_0^{2p} f(t) \exp(-ni\pi t/p)dt \tag{2.2}$$

If only discrete data points are given the above integral can be numerically approximated by a summation. Suppose there are N equally spaced data points, X_k:

$$t = k\Delta t \qquad N\Delta t = 2p \qquad \Delta t \simeq dt$$

$$C_n = \frac{1}{N\Delta t} \sum_{k=0}^{N-1} X_k \exp\left(\frac{-ni\pi k\Delta t}{N\frac{\Delta t}{2}}\right)\Delta t$$

$$C_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \exp(-ni2\pi k/N) \tag{B.1}$$

If the data are band-limited, that is, the frequency spectrum does not extend past a certain point, and if the sample point interval is smaller than a half-wavelength of the highest frequency present, then no information will be lost in taking the summation rather than the continuous integral in

determining the coefficients.[28] If the above is true, then more than N terms need not be included in order to completely restore the original function f(t).

$$f(t) = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} C_n \exp(ni\pi t/p) \qquad (B.2)$$

If discrete points are desired rather than a continuous function, this may be written as:

$$X_k = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} C_n \exp\left(\frac{ni\pi k\Delta t}{N\frac{\Delta t}{2}}\right)$$

$$X_k = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} C_n \exp(ni2\pi k/N) \qquad (B.3)$$

By considering the periodicity of the exponent and by noting that the derivation of C_n from -N/2 to -1 is the same as from N/2 to N-1, it is apparent that this expression may be written:

$$X_k = \sum_{n=0}^{N-1} C_n \exp(ni2\pi k/N) \qquad (B.4)$$

The expression in this form is called the "inverse discrete Fourier transform" and its companion:

$$C_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \exp(-ni2\pi k/N) \qquad (B.5)$$

the "discrete Fourier transform."

The equations expressed in this symmetric form make it possible to write only one computer program to handle both computations. One has only to change the sign of the exponent and to consider the factor 1/N when doing the inverse transform.

The fast Fourier transform, usually written FFT, is only a computational algorithm to shorten the length of time over conventional methods in computing the discrete Fourier transform.[27,29,30]

APPENDIX C

COMPUTER PROGRAMS WRITTEN TO PERFORM FFT DECONVOLUTION

Due to the enormous volume of calculation involved, computer programs were written to handle almost all aspects of the program.

The computer used was the Xerox Sigma 9 which operates under Control Program V (CPV version E00). Most computing was done on the cathode-ray terminals (time-sharing mode). The Tektronix 4010-1 graphics terminal and a Calcomp 565 drum plotter were used to plot all graphs.[a] All programs were written in FORTRAN IV with the exception of two Xerox extended FORTRAN options, the INPUT and OUTPUT statements. They serve as free form read and write, respectively.

No subroutines were used with these programs. In the time-sharing mode it was more convenient to use the SET statements of the terminal executive language, especially in the earlier stages of the research when it was necessary to examine the output of each program. It should not be difficult for the programmer to adapt the programs to any system, however, with slight modifications.

FFT deconvolution was implemented with these programs as follows:

First, a machine function was generated using either MACHFUN or TESTFUN. Program TESTFUN generated data points from a gaussian distribution. The height and width of the gaussian were varied by inputting two variables via the INPUT statement. MACHFUN first read in data points from the 5th peak in Run 1. Then by interpolating between the points the narrowness was varied in

[a] Tommy Dearmond, personal correspondence.

a continuous manner. The skewness was varied with one parameter by compressing the function behind the peak and expanding the function in front of it. The third parameter enabled the function to be shifted along the t-axis. This has the effect of shifting the deconvolved output. This was necessary in order that the program that builds the artificial function would not have "split" peaks to contend with.

Next, a forward transform was taken of the machine function using program FFPEAK[31] to get its coefficients. Also, the forward transform was taken of the data using FFPEAK.

These coefficients were read in program FGH and the coefficients of the data were divided, one by one, by the coefficients of the machine function. When applying a data window, this division is multiplied by a factor which depends on the data window applied. See program FGHAM to see how a Hamming window was applied to Run 5.

The coefficients obtained by division are read into FFPEAK and an inverse transform is taken. The user is allowed to select the number of coefficients wanted to go into the transform via an INPUT statement. The user can thus cut off the frequency spectrum at any point and note which gives the best deconvolution.

The real part of these numbers is read by program CALCOM and plotted. (If the imaginary part of these numbers is very large, one should suspect an error.) This is usually plotted along with the original data, and usually the hybrid function and artificial function, in order to better compare them.

Program SUB is used to construct the hybrid function. Coefficients from the artificial function are substituted in beginning at any desired point. To use this program, input the number of coefficients of the original function

that is wanted and the rest will be substituted from the artificial function. An inverse transform of this is taken using FFPEAK. The result is usually plotted along with the straightforward deconvolution in order to better compare them.

ARTFN is the program that builds the artificial function. It reads the deconvolved function and determines the location, height, and full width at half-maximum of each peak. An artificial function is then constructed with gaussian distributions having these parameters. This function may be plotted directly using CALCOM. The forward transform is then taken using FFPEAK to get the coefficients.

In coneluding, there is one point that needs clarifying. In some of the programs one, loosely, is asked to input the number of coefficients of that number of terms of the lowest frequencies. With the exception of the first coefficient, there are two coefficients that correspond to each wavelength.

```
  NAME = TESTFUN

C
C  THIS PROGRAM GENERATES DATA POINTS OF A GAUSSIAN
C  DISTRIBUTION. THERE ARE TWO INPUTS, THE FIRST
C  DETERMINES THE HEIGHT OF THE PEAK AND THE SECOND
C  DETERMINES THE WIDTH.
C
      DIMENSION Y(256)
      INPUT S,T
      DO 1 I=1,256
      Y(I)=S*DEXP(-(I-129.)**2*T)
    1 CONTINUE
      WRITE(3,2)Y
    2 FORMAT(10F5.0)
      STOP
      END
```

```
      NAME = MACHFUN

C
C  GIVEN ANY INITIAL SET OF DATA POINTS OF A PROSPECTIVE
C  MACHINE FUNCTION, THIS PROGRAM, BY USING THREE INPUT
C  PARAMETERS, GIVES THE FUNCTION A CERTAIN AMOUNT OF
C  SKEWNESS, NARROWS OR BROADENS THE FUNCTION, AND SHIFTS
C  THE FUNCTION LATERALLY, IN THAT ORDER.
C
      DIMENSION RUNIN(1024),MF(256),RU(1024)
      DATA MF/0/
      READ(35,1)RUNIN
    1 FORMAT(10F5.0)
      INPUT C,SC,IBEGIN
C
C  THE FOLLOWING BIT OF CODE COMPRESSES THE X-AXIS BEHIND
C  THE PEAK. POINT 433 IS AT THE TOP OF THE PEAK.
C
      DO 2 I=1,1024-433
      RU(I+433)=RUNIN(I+433)
      X=I
      K=INT(X-C*X)
      RUNIN(I+433)=(X-C*X-K)*(RU(K+834)
     *-RU(K+433))+RU(K+433)
    2 CONTINUE
C
C  THIS BIT OF CODE EXPANDS THE X-AXIS IN FRONT OF THE PEAK.
C
      DO 3 I=432,1,-1
      X=433-I
      K=433-INT(X+C*X)
      IF(K.LT.1)RUNIN(I)=3;GO TO 3
      RUNIN(I)=RUNIN(K)-((X+C*X)-INT(X+C*X))*
     *(RUNIN(K)-RUNIN(K-1))
    3 CONTINUE
C
C  THE FOLLOWING BIT OF CODE NARROWS AND SHIFTS THE FUNCTION.
C  IT CAN OBTAIN DATA POINTS FOR ANY INCREMENT BY INTERPOLATING
C  BETWEEN THE POINTS.
C
      DO 5 I=1,INT(1024/SC)
      RI=I
      K=RI*SC
    5 MF(I+IBEGIN)=RUNIN(K)+(RI*SC-K)*(RUNIN(K+1)-RUNIN(K))
      WRITE(36,4)MF
    4 FORMAT(10F5.0)
      STOP
      END
```

```
      NAME = FFPEAK

C
C     THIS PROGRAM COMPUTES THE FORWARD OR INVERSE FOURIER
C     TRANSFORM FOR ANY SET OF DISCRETE DATA POINTS WHICH IS AN
C     INTEGRAL POWER OF TWO. INPUT -1. FOR A FORWARD TRANSFORM
C     AND 1. FOR AN INVERSE TRANSFORM. IF -1. IS INPUTTED THE
C     PROGRAM WILL READ ONLY REAL DATA POINTS OF A SET OF ANY
C     LENGTH. IF 1. IS INPUTTED THE PROGRAM WILL READ ONLY
C     PROPERLY FORMATTED COMPLEX NUMBERS. GIVEN ANY ARBITRARY
C     SET OF DATA, BE SURE TO DIMENSION THE PROPER ARRAYS
C     TO THE SMALLEST POWER OF TWO THAT WILL CONTAIN THE DATA.
C     THE REST OF THE VALUES IN THE ARRAY WILL BE SET TO ZERO
C     TO GIVE AN OVERALL DATA FIELD OF LENGTH AN INTEGRAL POWER
C     OF TWO. SET NSTAGE TO THIS INTEGRAL POWER OF TWO.
C     ALSO, FOR THE INVERSE TRANSFORM YOU WILL NEED TO INPUT
C     THE NUMBER OF COEFFICIENTS YOU WISH; THE REST WILL BE SET
C     TO ZERO.
C
C     NOTE: ALTHOUGH THE PRESENT ARRAYS ARE DIMENSIONED TO HOLD
C     10_24 DATA POINTS, THE PROGRAM IS SET TO READ ONLY 256.
C
C
C     THIS FIRST BIT OF CODE IS TO FILL THE ARRAYS.
C
      REAL*8 DATA(1024),COEF(2,1024)
      REAL*8 FLTN,PHI2N,SIGN,FLIN2J,TEMP
      DOUBLE COMPLEX W,X(2,1024)
      DATA DATA/0/
      INTEGER R
      NSTAGE=10
      SIGN=-1.
      N=2**NSTAGE
      IF(SIGN.GT.0.)GO TO 33
      READ(11,77,END=66)(DATA(K),K=385,640)
   77 FORMAT(10F5.0)
   66 DO 22 I=1,N
   22 X(1,I)=DCMPLX(DATA(I),0.)
      GO TO 55
   33 INPUT M
      READ(13,99)COEF
   99 FORMAT(D21.15,2X,D21.15)
      DO 9 I=M+1,N+1-M
      COEF(1,I)=0.
    9 COEF(2,I)=0.
      DO 44 I=1,N
   44 X(1,I)=DCMPLX(COEF(1,I),COEF(2,I))
   55 CONTINUE
C
C     ACTUAL CALCULATION STARTS HERE.
C
      N2=N/2
      FLTN=N
      PHI2N=6.283185307179586/FLTN
      DO 3 J=1,NSTAGE
      N2J=N/(2**J)
      NR=N2J
      NI=(2**J)/2
      DO 2 I=1,NI
      IN2J=(I-1)*N2J
```

```
      FLIN2J=IN2J
      TEMP=F_IN2J*PHI2N*SIGN
      W=DCMPLX(DCOS(TEMP),DSIN(TEMP))
      DO 2 R=1,NR
      ISUB=R+IN2J
      ISUB1=R+IN2J*2
      ISUB2=ISUB1+N2J
      ISUB3=ISUB+N2
      X(2,ISUB)=X(1,ISUB1)+W*X(1,ISUB2)
      X(2,ISUB3)=X(1,ISUB1)-W*X(1,ISUB2)
    2 CONTINUE
      DO 3 R=1,N
    3 X(1,R)=X(2,R)
      IF(SIGN.GT.0.)GO TO 5
      DO 4 R=1,N
    4 X(2,R)=X(1,R)/FLTN
    5 WRITE(12,88)(X(2,J),J=1,N)
   88 FORMAT(D21.15,2X,D21.15)
      STOP
      END
```

```
  NAME = F34

C
C  THIS PROGRAM DIVIDES THE COEFFICIENTS OF THE TRANSFORM
C  OF THE GIVEN DATA POINTS BY THE COEFFICIENTS OF THE
C  MACHINE FUNCTION TERM BY TERM TO GIVE THE COEFFICIENTS
C  OF THE DECONVOLVED FUNCTION.
C
C
      DOUBLE COMPLEX F(1024),G(1024),H(1024)
      READ(8,1)G
    1 FORMAT(D21.15,2X,D21.15)
    2 READ(9,3)H
    3 FORMAT(D21.15,2X,D21.15)
    4 DO 5 I=1,1024
    5 F(I)=H(I)/G(I)
      WRITE(10,6)F
    6 FORMAT(D21.15,2X,D21.15)
      STOP
      END
```

```
  NAME = CALCOM

C
C   THIS IS THE PROGRAM THAT INITIATES THE PLOTTING ROUTINES.
C   IT WILL READ AS MANY AS FOUR DIFFERENT DATA FILES AND
C   PLOT THEM ALL ON ONE GRAPH. THERE ARE THREE INPUTS. FOR
C   THE FIRST, INPUT EITHER A '2' OR A '0', A '2' IF YOU
C   WANT THE OUTPUT ON THE TEKTRONIX SCREEN, OR '0' IF YOU WANT
C   OUTPUT ON THE CALCOMP PLOTTER. THE SECOND INPUT IS A SCALING
C   FACTOR. THE PROGRAM IS DESIGNED TO GRAPH ALL NUMERICAL
C   VALUES IN INCHES. ALL NUMERICAL VALUES TIMES THE SCALING
C   FACTOR WILL GIVE AS THE OUTPUT THIS PRODUCT IN INCHES. THE
C   THIRD INPUT IS THE AMOUNT IN DATA-POINT UNITS THAT SOME OF
C   THE GRAPHS ARE SHIFTED. THIS IS NECESSARY BECAUSE A SHIFT
C   IN THE MACHINE FUNCTION CAUSES A SHIFT IN THE OUTPUT. THIS
C   MUST BE ADJUSTED IN ORDER TO COMPARE IT WITH THE ORIGINAL
C   DATA.
C
      DIMENSION X(1024),Y(1024),YY(1024),Z(1024)
      REAL*8 EQVERV
      COMMON /MODEL/ ITEK
      INPUT ITEK,FCTR
      CALL FACTOR(FCTR)
 6    EQVERV=0.
      READ(15,55)Y
 55   FORMAT(321.15)
      READ(17,111)Z
 111  FORMAT(321.15)
      READ(16,66,END=77)YY
 66   FORMAT(10F5.0)
 77   DO 11 I=971,1024
 11   YY(I)=0.
      INPUT 4
      DO 8 I=1,1024
      X(I)=EQVERV+.5
      YY(I)=YY(I)*.003+5.
      Y(I)=Y(I)*.026+1.
      Z(I)=Z(I)*.026+2.5
 100  IF(YY(I).GT.10.)YY(I)=9.
      IF(YY(I).LT.0.)YY(I)=.001
      IF(Y(I).GT.10.)Y(I)=10.
      IF(Y(I).LT.0.)Y(I)=.001
      EQVERV=I/40.
 8    CONTINUE
      CALL PLOTS(1,1,1)
C     CALL AXIS(.5,0.,1HT,1,26.,0.,5.,0.,1.,4HF4.1)
C     CALL AXIS(.5,0.,1HG,1,10.,90.,1.,0.,1.,4HF4.1)
      CALL PLOT(X(1),Y(M),3)
      DO 88 K=2,1024-M
 88   CALL PLOT(X(K),Y(K+M-1),2)
      DO 22 K=1025-M,1024
 22   CALL PLOT(X(K),Y(K+M-1024),2)
      CALL PLOT(X(1),Z(1),3)
      DO 99 K=2,1024-M
 99   CALL PLOT(X(K),Z(K+M-1),2)
      DO 95 K=1025-M,1024
 95   CALL PLOT(X(K),Z(K+M-1024),2)
      CALL PLOT(X(1),YY(1),3)
      DO 44 K=2,1024
 44   CALL PLOT(X(K),YY(K),2)
```

```
C       CALL LINE(X,Y,1024,1,4,.2,0)
C       CALL LINE(X,YY,1024,1,4,.2,0)
        CALL STOPPLOT
        STOP
        END
```

```
      NAME = ARTFN

C
C     THIS PROGRAM SCANS THE DATA POINTS OF THE DECONVOLVED
C     FUNCTION AND DETERMINES THE PEAK HEIGHTS, LOCATION, AND
C     THE FULL WIDTH AT HALF MAXIMUM OF EACH. AN ARTIFICIAL
C     FUNCTION IS THEN BUILT FROM GAUSSIAN DISTRIBUTIONS WITH
C     THE SAME NUMBER OF PEAKS, AND WITH HEIGHTS, WIDTHS, AND
C     LOCATIONS AS DETERMINED ABOVE.
C     THE PEAK LOCATIONS AND WIDTHS ARE OUTPUTTED IMMEDIATELY
C     TO CHECK FOR GROSS ERRORS.
C
          DIMENSION PEAKFN(1024),HT(5),LOC(5),LR(5)
          DIMENSION LL(5),WF(5),WW(5)
          READ(50,4)PEAKFN
      4   FORMAT(D21.15)
C
C     THE FOLLOWING BIT OF CODE DETERMINES THE PEAK HEIGHT,
C     LOCATION, AND FWHM.
C
          K=0
          L=0
          ISW=0
          DO 1 I=1,1024
          IF(PEAKFN(I).LT.20.)GO  TO 1
          IF(PEAKFN(I).GT.PEAKFN(I-1).AND.PEAKFN(I).GT.
         *PEAKFN(I+1))L=L+1;ISW=1;HT(L)=PEAKFN(I);
         *LOC(L)=I;GO TO 2
          IF(ISW.EQ.1)GO TO 2
          GO TO 1
      2   K=K+1
          IF(PEAKFN(I).LT.HT(L)/2.)LR(L)=K-1;ISW=0;K=0
      1   CONTINUE
          DO 7 I=1,5
          N=LOC(I)-LR(I)
          DO 8 J=1,9
          IF(PEAKFN(J+N-3).GT.HT(I)/2.)LL(I)=LR(I)+4-J;
         *WF(I)=.69315/((LR(I)+LL(I))/2.)**2;GO TO 9
      8   CONTINUE
      9   WW(I)=LR(I)+LL(I)
          OUTPUT WW(I)
          OUTPUT LOC(I)
      7   CONTINUE
C
C     THE FOLLOWING BIT OF CODE BUILDS THE ARTIFICIAL FUNCTION
C     FROM GAUSSIANS
C
          DO 6 J=1,1024
      6   PEAKFN(J)=HT(1)*EXP(-(J-LOC(1))**2*WF(1))
         *+HT(2)*EXP(-(J-LOC(2))**2*WF(2))
         *+HT(3)*EXP(-(J-LOC(3))**2*WF(3))
         *+HT(4)*EXP(-(J-LOC(4))**2*WF(4))
         *+HT(5)*EXP(-(J-LOC(5))**2*WF(5))
          WRITE(60,5)PEAKFN
      5   FORMAT(10F5.0)
          STOP
          END
```

```
      NAME = SJ3

C
C     THIS PROGRAM SUBSTITUTES COEFFICIENTS OF THE ARTIFICIAL
C     FUNCTION INTO THE TRANSFORM OF THE DECONVOLVED FUNCTION.
C     INPUT THE NUMBER OF COEFFICIENTS OF THE ORIGINAL
C     FUNCTION THAT YOU WANT AND THE REST WILL BE AUTOMATICALLY
C     SUBSTITUTED FROM THE ARTIFICIAL FUNCTION.
C
          DOUBLE COMPLEX HOUT(1024),HHOUT(1024)
          READ(1,11)HOUT
   11     FORMAT(D21.15,2X,D21.15)
          READ(2,22)HHOUT
   22     FORMAT(D21.15,2X,D21.15)
          INPUT M
          DO 33 I=M+1,1025-M
   33     HOUT(I)=HHOUT(I)
          WRITE(3J,44)HOUT
   44     FORMAT(D21.15,2X,D21.15)
          STOP
          END
```

```
NAME = F3HA4

      DOUBLE COMPLEX F(1024),G(1024),H(1024)
      READ(8,1)F
    1 FORMAT(D21.15,2X,D21.15)
    2 READ(9,3)H
    3 FORMAT(D21.15,2X,D21.15)
    4 DO 5 I=1,510
      IF(I.GT.46)G(I)=0;G(1026-I)=0;GO TO 5
      G(I)=(.54+.46*DCOS(3.1415926589793*
     *((I-1)/45.)))*(H(I)/F(I))
      G(1025-I)=(.54+.46*DCOS(3.1415926589793*
     *(I/45.)))*(H(1025-I)/F(1025-I))
    5 CONTINUE
      WRITE(10,6)G
    6 FORMAT(D21.15,2X,D21.15)
      STOP
      END
```

```
  NAME = AREA

C
C   THIS PROGRAM FINDS THE AREAS UNDER THE PEAKS. THE ENDPOINTS
C   MUST BE DETERMINED BY INSPECTION.
C
      DIMENSION A(5),DECON(5030)
      DATA A/0/
      READ(70,6)DECON
    6 FORMAT(10F5.0)
      DO 1 I=1,470
    1 A(1)=A(1)+DECON(I)
      DO 2 I=472,1120
    2 A(2)=A(2)+DECON(I)
      DO 3 I=1921,2861
    3 A(3)=A(3)+DECON(I)
      DO 4 I=2916,3770
    4 A(4)=A(4)+DECON(I)
      DO 5 I=4088,4960
    5 A(5)=A(5)+DECON(I)
      DO 7 J=1,5
    7 WRITE(71,8)J,A(J)
    8 FORMAT(' ','AREA1(',I1,')=',F8.0)
      OUTPUT ,THESE RESULTS MUST BE MULTIPLIED BY THE
     #PROPER PROPORTIONALITY CONSTANT TO GIVE CORRECT AREA'
      STOP
      END
```

REFERENCES

1. H.M. McNair and E.J. Bonelli, Basic Gas Chromatography, 5th ed. (Berkeley, Ca.: Consolidated Printers for Gow-Mac Instrument Co., 1969), pp. 82-84.

2. Ibid., p. 153.

3. David Abbott and R.S. Andrews, An Introduction to Chromatography (Boston: Houghton Mifflin Co., 1965), p. 36.

4. McNair and Bonelli, pp. 82-85.

5. Ibid., pp. 99-101.

6. Ibid., pp. 101-104.

7. Ibid., p. 104.

8. Memorandum from NASA/Langley Research Center, July 9, 1976.

9. Jack M. Gill and Frank T. Tao, Quantitation and Gas Chromatography, Technical Bulletin 135-67 (Walnut Creek, Ca.: Varian Aerograph, 1967), pp. 4-7.

10. McNair and Bonelli, Chapter 3.

11. Ron Bracewell, The Fourier Transform and Its Applications (New York: McGraw-Hill Book Co., 1965).

12. R.N. Bracewell and J.A. Roberts, Aust. J. Phys. 7, 615 (1954).

13. Bracewell, Chapter 3.

14. Ibid., pp. 204-214.

15. G.M. Jenkins and D.G. Watts, Spectral Analysis and Its Applications (San Francisco, Ca.: Holden-Day, 1969), pp. 280-282.

16. Ibid., pp. 48-50.

17. R.B. Blackman and J.W. Tukey, The Measurement of Power Spectra (New York: Dover Publications, 1959).

18. D.W. Kirmse and A.W. Westerberg, Anal. Chem, 43, 1035 (1971).

19. H.M. Gladney, B.F. Dowden, and J.D. Swalen, Anal. Chem. 41, 883 (1969).

20. McNair and Bonelli, pp. 82-85, 140.

21. Ibid., p. 139.

22. Ibid., pp. 137-140.

23. Ibid., pp. 158-160.

24. Ibid., p. 159.

25. Jenkins and Watts.

26. Bracewell, Chapter 8.

27. G.D. Bergland, IEEE Spectrum 6, 41 (July 1969).

28. W.T. Cochran, et al., IEEE Trans. Audio Electroacoust. 15, 45(June 1967), cited by W.T. Cochran, et al., Proc. IEEE 54, 830 (1966).

29. J.W. Cooley and J.W. Tukey, Math. Comput. 19, 297 (1965).

30. W.T. Chochran, et al., IEEE Trans. Audio Electroacoust. 15, 45(June 1967).

31. Mark L. Uhrich, IEEE Trans. Audio Electroacoust. 17, 170 (June 1969).

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-3229		

4. Title and Subtitle	5. Report Date
DECONVOLUTION OF GAS CHROMATOGRAPHIC DATA	February 1980
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Samuel Howard and Grayson H. Rayborn	
	10. Work Unit No. 506-16-21-02
9. Performing Organization Name and Address	11. Contract or Grant No.
University of Southern Mississippi Hattiesburg, MS 39401	NSG-1380
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Washington, DC 20546	14. Army Project No.

15. Supplementary Notes

Langley Technical Monitor: George M. Wood
Final Report

16. Abstract

This report discusses the use of deconvolution methods on gas chromatographic data to obtain an accurate determination of the relative amounts of each material present by mathematically separating the merged peaks. The underlying assumption implied throughout is that the area under the peaks is an accurate indicator of the amount of material present. Due to uncontrollable errors at every stage of operation, this is never exactly true. However, it is a consideration that must always be borne in mind in designing, building, and operating the gas chromatograph, its detector, and other related equipment.

Data used in the present study were obtained on a gas chromatograph with a flame ionization detector. Chromatograms of five xylenes with differing degrees of separation were generated by varying the column temperature at selected rates. The merged peaks were then successfully separated by deconvolution. The concept of function continuation in the frequency domain was introduced in striving to reach the theoretical limit of accuracy, but proved to be only partially successful.

The methods used here will have general applicability to any deconvolution where the desired result consists of a number of well separated distributions.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Gas Chromatography Gas Analysis Data Enhancement	Unclassified - Unlimited Subject Category 35

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	76	$6.00

*For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1980