TECHNOMETRICS VOL. 13, NO. 3 AUGUST 1971

Self Modeling Curve Resolution

WILLIAM H. LAWTON AND EDWARD A. SYLVESTRE
Eastman Kodak Company, Rochester, New York

This paper presents a method for determining the shapes of two overlapping functions $f_1(x)$ and $f_2(x)$ from an observed set of additive mixtures, $\{\alpha_i f_1(x) + \beta_i f_2(x); i = 1, \cdots, n\}$, of the two functions. This type of problem arises in the fields of spectrophotometry, chromatography, kinetic model building, and many others. The methods described by this paper are based on the use of principal component techniques, and produce two bands of functions, each of which contains one of the unknown, underlying functions. Under certain mild restrictions on the $f_j(x)$, each band reduces to a single curve, and the $f_j(x)$ are completely determined by the analysis.

1. THE PROBLEM

The techniques described in this paper may be applied to any experiment whose outcome is a continuous curve $Y(\lambda)$ which is an additive mixture of two unknown, nonnegative, linearly independent functions. That is, $Y(\lambda)$ is of the form

$$Y(\lambda) = \alpha f_1(\lambda) + \beta f_2(\lambda), \quad L_1 \leq \lambda \leq L_2 \tag{1}$$

where the $f_i(\lambda)$ are unknown, nonnegative, linearly independent functions. We shall also assume that the $f_i(\lambda)$ are normalized to unit area, so that

$$\int_{L_1}^{L_2} f_i(\lambda)\, d\lambda = 1, \tag{2}$$

and the coefficients α and β are nonnegative random variables with $\alpha \not\equiv \delta\beta$.

The condition (2) imposes no restriction on the shapes of the unknown functions $f_i(\lambda)$, since any bounded function on the interval $[L_1, L_2]$ may be normalized to unit area. The condition that α not be a multiple of β asks only that the relative amounts of the two unknown functions vary from sample to sample. If this condition were not satisfied, then (1) becomes

$$Y(\lambda) = \beta f(\lambda), \tag{3}$$

where $f(\lambda) = \delta f_1(\lambda) + f_2(\lambda)$. The observed curves from such a process cannot be distinguished from those of a process which has randomly varying amounts of a single function $f(\lambda)$.

If the above mentioned conditions are satisfied, and there is no region in $[L_1, L_2]$ in which both $f_1(\lambda)$ and $f_2(\lambda)$ are simultaneously identically zero, then it is possible to construct a band of functions which contains $f_1(\lambda)$, and a band which contains $f_2(\lambda)$. These bands are constructed from a sample of observed

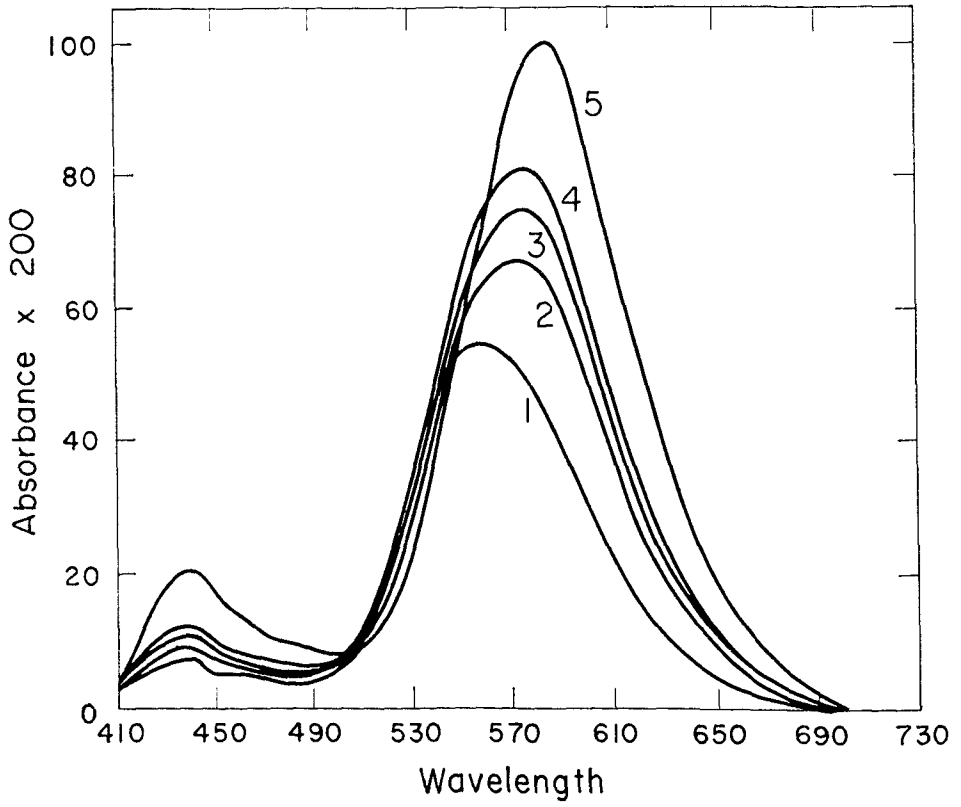
Received Dec. 1968; revised June 1969.

mixtures $\{Y_i(\lambda), i = 1, 2, \cdots, n\}$. The usual approach has required that one assume that the $f_i(\lambda)$ had some known shape (Gaussian, Lorentzian, etc.) and then the underlying functions were determined by nonlinear regression as discussed in [4]. The techniques presented in this paper are referred to as *self modeling* resolution because no assumptions are made concerning the shapes of the unknown $f_i(\lambda)$.

2. Application to Spectrophotometry

Experiments whose outcomes are of the form (1) are frequently found in the field of spectrophotometry. The techniques for self modeling curve resolution will now be developed by means of a spectrophotometric example.

An analytic chemist has received 5 samples of material from an experimental production process, and these samples exhibit noticeable color variations as indicated by the 5 spectrophotometric curves shown in Figure 1. One would like to know the nature and cause of this variation in the spectrophotometric curves. The experimental process involves a chemical reaction which is supposed to yield a 50/50 mixture of two standard dyes. The spectrophotometric curve of the production material is supposed to match the curve of this 50/50 mixture (the aim curve). The 5 observed curves exhibit considerable variation, and they all differ from the desired aim curve. There are, of course, many possible causes for this variation in the observed curves. It might be that process in-



stability is producing varying amounts of the correct standard dyes, or it might be producing varying amounts of dyes other than the standard dyes, or it might even be that the correct 50/50 mixture of the standard dyes is being formed, but there are also varying amounts of some impurity (a third unwanted dye) present. The chemist must determine which, if any, of these explanations is correct.

The chemist is willing to postulate that these five solution samples are made up of varying amounts of two dyes, Dye A and Dye B. He is, however, unwilling to assume that these are the desired standard dyes. Under this assumption of a two dye system, it follows from the laws of absorbtion spectrophotometry (see [8]), that the five observed curves given in Figure 1 must be of the form

$$Y_i(\lambda) = a_i g_1(\lambda) + b_i g_2(\lambda), \quad L_1 \leq \lambda \leq L_2 \tag{4}$$

where $[L_1 , L_2]$ is some suitable range of wavelengths. In this expression $g_1(\lambda)$ is the spectrophotometric curve of Dye A at unit concentration and a_i is the concentration of Dye A in solution i. $g_2(\lambda)$ and b_i are the same for Dye B.

Since Dye A and Dye B are distinct, it follows that their spectrophotometric curves are linearly independent. These two functions are also nonnegative since they represent optical densities at varying wavelengths λ. These functions would, however, not normally have unit area. If one defines

$$c_j = \int_{L_1}^{L_2} g_j(\lambda)\, d\lambda, \qquad j = 1, 2, \tag{5}$$

then it is easily shown from (4) and (5), that the observed curves in Figure 1 must also satisfy

$$Y_i(\lambda) = \alpha_i f_1(\lambda) + \beta_i f_2(\lambda), \quad L_1 \leq \lambda \leq L_2 , \tag{6}$$

where $f_j(\lambda) = g_j(\lambda)/c_j$, and $\alpha_i = a_i c_1$, $\beta_i = b_i c_2$.

Thus, if the chemist's two-dye hypothesis is correct, the 5 observed spectrophotometric curves should satisfy (6) where the $f_j(\lambda)$ and the α's and β's satisfy the conditions of Section 1. If one assumes that the relative concentrations of Dye A and Dye B change from sample to sample, then this means that there is no δ such that $\alpha_i = \delta\beta_i$ for all $i = 1, 2, \cdots , n$. It follows then, from the chemist's assumptions, that the observed curves shown in Figure 1 should satisfy all the conditions imposed in Section 1.

3. Solution of the Problem

The first step of the analysis is to determine whether or not the two-dye mix hypothesis is reasonable. This premise will be checked by means of principal component analysis. This approach was used by Wernimont [7] for a one-dye system.

The ordinate values of each of the observed spectrophotometric curves are recorded at a set of p selected wavelengths λ_k , $k = 1, 2, \cdots , p$. For each observed curve $Y_i(\lambda)$ one now has a vector $\mathbf{Y}_i = (Y_i(\lambda_1), \cdots , Y_i(\lambda_p))$. The 5 vectors derived from the curves of Figure 1 are shown in Table 1. The λ_k

TABLE 1

Curve 1	Curve 2	Curve 3	Curve 4	Curve 5
0.924	2.478	1.239	0.413	2.774
4.406	8.006	6.845	5.075	11.920
5.488	9.009	10.110	6.393	18.392
6.530	11.900	11.586	9.009	20.969
4.977	9.422	10.307	7.475	16.681
4.898	8.419	8.242	6.452	13.907
3.875	6.432	6.845	4.839	10.878
3.600	6.157	6.255	4.485	10.032
3.501	4.780	5.272	3.796	8.439
4.878	5.429	5.724	4.917	7.534
9.992	9.953	9.678	8.950	9.068
16.739	16.601	17.762	15.815	12.845
27.341	27.715	28.521	26.653	20.083
40.146	44.041	42.566	42.015	33.321
52.735	62.570	58.085	55.155	48.752
54.801	72.995	66.858	62.944	70.262
51.260	80.155	72.405	66.701	88.633
46.775	81.512	73.448	66.937	98.350
39.832	74.962	68.353	60.623	97.800
30.272	64.950	57.613	52.676	88.240
22.463	51.496	45.969	39.969	72.799
15.795	34.875	34.442	30.154	55.352
11.350	25.728	25.079	21.303	41.189
7.947	17.900	17.703	15.087	30.135
4.760	11.271	11.684	9.796	20.103
2.813	7.317	7.140	5.842	13.632
2.065	4.485	4.544	3.698	8.065
1.593	2.813	2.655	2.419	5.134
0.964	1.436	1.318	1.259	2.833
0.669	0.472	0.079	0.138	0.551

are selected so that connecting the points $(\lambda_k , Y_i(\lambda_k))$ by straight lines yields a good graphic representation of the original curve $Y_i(\lambda)$.

In vector notation the two-dye mix assumption (6) becomes

$$\mathbf{Y}_i = \alpha_i \mathbf{f}_1 + \beta_i \mathbf{f}_2 \tag{7}$$

where the $\mathbf{f}_j$ are unknown, linearly independent vectors with all elements nonnegative; namely, $\mathbf{f}_j = (f_j(\lambda_1), \cdots , f_j(\lambda_p))$. The α_i and β_i are as defined in (6). The unit area restriction is replaced by

$$\sum_{k=1}^{p} f_i(\lambda_k)\, \Delta_k = 1, \tag{8}$$

where $\Delta_1 = \lambda_2 - \lambda_1$, $\Delta_p = \lambda_p - \lambda_{p-1}$, $\Delta_k = (\lambda_{k+1} - \lambda_{k-1})/2$ for $k = 2, 3, \cdots , p - 1$. In our present example the λ_k are 410, 420, 430, $\cdots$ 700 so that $\Delta_k = 10$ mμ. Equation (8) becomes

$$\sum_{k=1}^{p} f_i(\lambda_k) = 0.1. \tag{9}$$

As noted at the beginning of this section, the first goal is the verification of the structure suggested by (7). That is, one would like to verify that the 5 observed vectors $\mathbf{Y}_i$ are all linear combinations of two basis vectors. Now if (7) holds, then it is easily shown that all the $\mathbf{Y}_i$ are simply linear combinations of the first two eigenvectors of the second moment matrix. That is,

$$\mathbf{Y}_i = \xi_{i1}\mathbf{V}_1 + \xi_{i2}\mathbf{V}_2 \tag{10}$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are the eigenvectors of the second moment matrix

$$M = [Y'Y]/n \tag{11}$$

associated with the two largest eigenvalues of M. These $\mathbf{V}_i$ are the eigenvectors of unit length, see Anderson [1], [2]. In (11) Y denotes the n by p data matrix.

$$Y = \begin{bmatrix} Y_1(\lambda_1) & \cdots & Y_1(\lambda_p) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ Y_n(\lambda_1) & \cdots & Y_n(\lambda_p) \end{bmatrix}. \tag{12}$$

The data matrix for the present example is given in Table 1; the eigenvectors of the resulting M are given in Table 2.

TABLE 2

$\mathbf{V}_1$	$\mathbf{V}_2$
0.009	0.015
0.041	0.048
0.056	0.103
0.068	0.103
0.055	0.080
0.047	0.058
0.037	0.045
0.034	0.041
0.029	0.030
0.031	-0.004
0.050	-0.078
0.082	-0.169
0.134	-0.291
0.210	-0.400
0.290	-0.477
0.349	-0.302
0.389	-0.074
0.402	0.094
0.377	0.218
0.327	0.284
0.261	0.285
0.191	0.243
0.140	0.192
0.100	0.150
0.065	0.109
0.042	0.085
0.026	0.044
0.016	0.025
0.009	0.014
0.002	-0.002

For each $\mathbf{Y}_i$ the best (least squares) fit using the model in (10) is

$$\hat{\mathbf{Y}}_i = \hat{\xi}_{i1}\mathbf{V}_1 + \hat{\xi}_{i2}\mathbf{V}_2 , \tag{13}$$

where $(\hat{\xi}_{i1} , \hat{\xi}_{i2})$ are the values which minimize

$$Q(\xi_{i1} , \xi_{i2}) = \sum_{k=1}^{p} [Y_i(\lambda_k) - \xi_{i1}V_1(\lambda_k) - \xi_{i2}V_2(\lambda_k)]^2 . \tag{14}$$

Due to the orthogonality of the eigenvectors, it is easily shown that

$$\xi_{ij} = \mathbf{V}_j \cdot \mathbf{Y}_i . \tag{15}$$

Figure 4 shows the $\mathbf{Y}_i$, $i = 1, 2, \cdots , 5$ plotted as points with the least squares estimates $\hat{\mathbf{Y}}_i$ plotted as continuous curves. The errors of fit

$$s_i^2 = \sum_{k=1}^{p} [Y_i(\lambda_k) - \hat{Y}_i(\lambda_k)]^2/(p - 2) \tag{16}$$

are on the order of experimental error. Since the reconstruction of the individual curves shown in Figure 4 gives a satisfactory representation of the original curves, the chemist concludes that the 5 observed spectrophotometric curves do represent varying additive mixtures of two unknown dyes. Thus, the structure in (7) is considered verified.

Having decided that the two-dye mix assumption is reasonable, then chemist must now determine the shapes of the unknown spectrophotometric curves. That is, he must determine the vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ of equation (7).

In the absence of experimental error it follows that

$$\begin{aligned} \mathbf{f}_1 &= \eta_{11}\mathbf{V}_1 + \eta_{12}\mathbf{V}_2 \\ \mathbf{f}_2 &= \eta_{21}\mathbf{V}_1 + \eta_{22}\mathbf{V}_2 . \end{aligned} \tag{17}$$

That is, the true, underlying spectrophotometric vectors $\mathbf{f}_i$ are simply linear combinations of the eigenvectors $\mathbf{V}_1$ and $\mathbf{V}_2$.

If the observed curves are subject to experimental error, and one observes $\mathbf{Y}_i^* = \mathbf{Y}_i + \mathbf{e}_i$, where $\mathbf{e}_i$ is a vector of independent errors, then it is clear that the eigenvectors $\mathbf{V}_i^*$ estimated from the data matrix Y^* are also subject to random errors. Equation (17) then fails to hold since the lefthand side contains the deterministic vectors $\mathbf{f}_i$, and the righthand side contains the stochastic vectors $\mathbf{V}_i^*$. However, it can be shown that (17) continues to hold as a limit in probability.

If experimental error is present, make m replicate measurements on each of the $n = 5$ chemical solutions. This action yields an observed set of $5m$ curves, and a $5m$ by p data matrix Y^*. Under these conditions, it can be shown, see Anderson [2], that the estimated eigenvectors $\mathbf{V}_i^*$ are consistent estimators of the true vectors $\mathbf{V}_1$ and $\mathbf{V}_2$ as $m \to \infty$. Thus, (17) does continue to hold as a limit in probability. The effects of experimental error in general, and the effect of m in particular, will be discussed further in Section 5 which contains the results of some simulations.

All of the results which follow in this section hold exactly in the absence of experimental error, and hold as limits in probability in the presence of experimental error.

It follows from (10) and (17) that all of the observed vectors $\mathbf{Y}_i$, and the true underlying spectrophotometric vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ can be expressed as linear combinations of the orthogonal eigenvectors $\mathbf{V}_1$ and $\mathbf{V}_2$. This leads to a particularly useful representation of the random process. Since each of the observed vectors $\mathbf{Y}_i$ is completely described by the amounts (ξ_1, ξ_2) of the eigenvectors, it is possible to represent each observed vector as a point in the (ξ_1, ξ_2)-plane. Using (15), one can obtain the point representations of our 5 observed vectors in Table 1.

$$\begin{array}{ccc} \mathbf{Y}_1 & \mathbf{Y}_2 & \mathbf{Y}_3 \\ (125.7, -33.4) & (201.7, -6.6) & (186.5, -8.2) \\ \mathbf{Y}_4 & \mathbf{Y}_5 & \\ (169.3, -13.2) & (235.4, 39.5). & \end{array} \tag{18}$$

These points are plotted in Figure 2. It might be worth noting that the restriction that there be no δ such that $\alpha_i \neq \delta\beta_i$ for all i, mentioned in (7), is equivalent to asking that all the curves must not fall on a ray in the (ξ_1, ξ_2)-plane.

Since the true, underlying spectrophotometric vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ are themselves linear combinations of the $\mathbf{V}_i$, it follows that they too can be represented as points in the (ξ_1, ξ_2)-plane. By (17) the points (η_{11}, η_{12}) and (η_{21}, η_{22}) define $\mathbf{f}_1$ and $\mathbf{f}_2$ respectively. The only question is where these two points are located in the (ξ_1, ξ_2)-plane. Determination of their locations is equivalent to the determination of the spectrophotometric vectors $\mathbf{f}_1$ and $\mathbf{f}_2$.

One can show that the first eigenvector $\mathbf{V}_1$ of the matrix M will have all its elements nonnegative. However, because of the orthogonality condition imposed on eigenvectors, the second eigenvector $\mathbf{V}_2$ must have some negative elements (see Table 2). Thus, it is clear that $\mathbf{V}_2$ cannot be one of the underlying vectors $\mathbf{f}_i$ since these vectors are known to have nonnegative elements. This known property of the $\mathbf{f}_i$ then implies that the point (0, 1) in the (ξ_1, ξ_2)-plane could not be one of the desired points in the plane representing an unknown $\mathbf{f}_i$. However, the exclusion of the point (0, 1) raises the question of whether or not one might not be able to rule out other points in the (ξ_1, ξ_2)-plane.

Since the unknown spectrophotometric vectors $\mathbf{f}_i$ have nonnegative elements, it follows that the points (η_{11}, η_{12}) and (η_{21}, η_{22}) must lie in the region of the (ξ_1, ξ_2)-plane which satisfies

$$\xi_1 v_{1k} + \xi_2 v_{2k} \geq 0 \quad \text{for all} \quad k, \tag{19}$$

where v_{jk} is the k-th element of the j-th eigenvector $\mathbf{V}_j$. Not all (ξ_1, ξ_2) satisfy this condition in (19) since it was just noted that (0, 1) does not satisfy (19).

Let $K^+ = \{k: v_{2k} \geq 0\}$ and $K^- = \{k: v_{2k} < 0\}$, then the set of points in the (ξ_1, ξ_2)-plane which satisfy (19) is given by the set of points

$$\begin{aligned} \xi_2 &\geq \zeta\xi_1 \\ \xi_2 &\leq \tau\xi_1 \\ \xi_1 &\geq 0, \end{aligned} \tag{20}$$

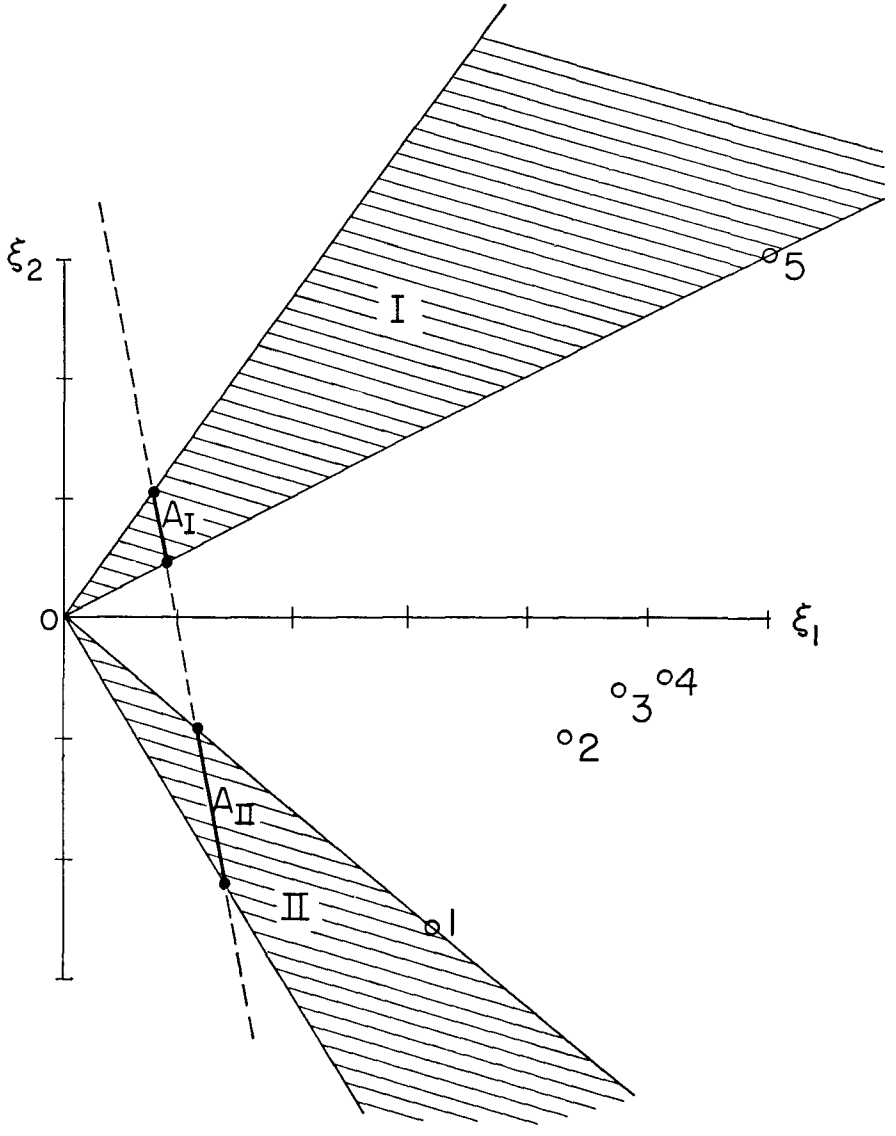


FIGURE 2

where

$$\zeta = -\min_{k \in K^+} \left| \frac{v_{1k}}{v_{2k}} \right|$$

$$\tau = \min_{k \in K^-} \left| \frac{v_{1k}}{v_{2k}} \right|.$$

The region defined by (20) is the region of coefficients (ξ_1 , ξ_2) yielding linear combinations of the eigenvectors $\mathbf{V}_1$ and $\mathbf{V}_2$ with all elements nonnegative. Since the underlying $\mathbf{f}_j$ are linear combinations of these eigenvectors with all elements nonnegative, it follows that the points (η_{11} , η_{12}) and (η_{21} , η_{22}) must lie in the region of the (ξ_1 , ξ_2)-plane described by (20). This region of non-

negative linear combinations is the pie-shaped region in Figure 2 defined by the upper boundary of the shaded Region (I) and the lower boundary of the shaded Region (II).

Notice that the points representing the observed $\mathbf{Y}_i$ fall in this region defined by (20). This is as it should be since all these observed vectors had all elements nonnegative. Thus, by making use of one of the known properties of the unknown $\mathbf{f}_j$, it has been possible to restrict the region of the (ξ_1 , ξ_2)-plane which contains the unknown (η_{11} , η_{12}), (η_{21} , η_{22}). One now seeks to further restrict the region containing these desired points.

The model in (7) not only specified that the elements of the $\mathbf{f}_j$ were nonnegative, but also specified that the amounts of these vectors, the (α_i , β_i), be nonnegative. That is, every observed vector must be of the form

$$\mathbf{Y}_i = \alpha_i \mathbf{f}_1 + \beta_i \mathbf{f}_2 \tag{21}$$

with $\alpha_i \geq 0$ and $\beta_i \geq 0$.

This restriction in (21) must now be translated into a restriction in the (ξ_1 , ξ_2)-plane. Using (10), (17), and (21), one finds that (21) is equivalent to asking that

$$(\xi_{i1} , \xi_{i2}) = \alpha_i(\eta_{11} , \eta_{12}) + \beta_i(\eta_{21} , \eta_{22}) \tag{22}$$

for all $i = 1, 2, \cdots , n$ with (α_i , β_i) as defined in (21). That is, the vectors (η_{11} , η_{12}), (η_{21} , η_{22}) representing the underlying vectors $\mathbf{f}_j$ must be such that the point (ξ_{i1} , ξ_{i2}), representing any observed vector $\mathbf{Y}_i$, can be obtained from the η-vectors by means of nonnegative linear combinations. This restriction in (22) means that one of the points, say (η_{11} , η_{12}) must lie in Region (I) described by (20), but with

$$\xi_2 \geq \max_{1 \leq i \leq n} [\xi_{i2}/\xi_{i1}]\xi_1 . \tag{23}$$

The region boundary in (23) represents a ray through the point representing $\mathbf{Y}_5$. The other unknown point (η_{21} , η_{22}) must lie in Region (II) defined by (20), but with

$$\xi_2 \leq \min_{1 \leq i \leq n} [\xi_{i2}/\xi_{i1}]\xi_1 . \tag{24}$$

The region boundary in (24) is a ray through the point representing $\mathbf{Y}_1$. Applying these results to the 5 curve example then yields the two disjoint regions shown in Figure 2. The shaded Region (I) contains the point (η_{11} , η_{12}), while the shaded Region (II) contains the point (η_{21} , η_{22}).

Thus, by making use of two of the known properties of the unknown $\mathbf{f}_j$ one has obtained two disjoint regions, Region (I) and Region (II), in the (ξ_1 , ξ_2)-plane each of which contains one of the unknown η points. There is yet another restriction on the location of the η's; namely, the unit area restriction of (8).

It was assumed that the unknown spectrophotometric vectors $\mathbf{f}_j$ were normalized so that they satisfied (8). If this is translated into a restriction in the (ξ_1 , ξ_2)-plane, then one finds that the unknown (η_{11} , η_{12}) and (η_{21} , η_{22}) must lie on the line

$$c_1\xi_1 + c_2\xi_2 = 1, \tag{25}$$

where

$$c_j = \sum_{k=1}^{p} v_{jk} \Delta_k . \tag{26}$$

Equation (25) follows from the fact that all linear combinations $\xi_1 \mathbf{V}_1 + \xi_2 \mathbf{V}_2$ satisfying (8) satisfy (25).

Combining (25) with our previous restrictions, it follows that (η_{11}, η_{12}) must lie on the segment cut from the line in (25) by the Region (I). Call this segment A_I. Similarly, (η_{21}, η_{22}) must lie on the segment A_{II} cut from the line in (25) by the Region (II). The line in (25) and the segments A_I and A_{II} are shown in Figure 2 for our present example.

At this point one can say that $(\eta_{11}, \eta_{12}) \in A_I$ and $(\eta_{21}, \eta_{22}) \in A_{II}$. The location of these unknown points has been greatly restricted. It is desirable, however, to convert this statement about the line segments A_I and A_{II} back into statements about spectrophotometric curves. This transition is easily accomplished when one notes that the vector band $F_I = \{\xi_1 \mathbf{V}_1 + \xi_2 \mathbf{V}_2 : (\xi_1, \xi_2) \in A_I\}$ must contain the underlying vector $\mathbf{f}_1$, while the band $F_{II} = \{\xi_1 \mathbf{V}_1 + \xi_2 \mathbf{V}_2 : (\xi_1, \xi_2) \in A_{II}\}$ must contain $\mathbf{f}_2$. F_I and F_{II} will be called the *solution bands* for $\mathbf{f}_1$ and $\mathbf{f}_2$ respectively.

It can easily be shown that these solution bands have the form

$$\begin{aligned} F_I &= \{a\mathbf{f}_I^* + (1 - a)\mathbf{f}_I^{**}; 0 \leq a \leq 1\} \\ F_{II} &= \{a\mathbf{f}_{II}^* + (1 - a)\mathbf{f}_{II}^{**}; 0 \leq a \leq 1\} \end{aligned} \tag{27}$$

where $\mathbf{f}_I^*$ is the linear combination of the $\mathbf{V}_j$ associated with the point in the (ξ_1, ξ_2)-plane determined by the intersection of the line in (25) with the line $\xi_2 = \zeta\xi_1$ of (20). In similar fashion, $\mathbf{f}_I^{**}$ is the linear combination associated with the point of intersection of the line in (25) with the line in (23). The intersection of the line in (25) with the upper boundary of Region (II) yields $\mathbf{f}_{II}^{**}$, while the intersection with the lower boundary of Region (II) yields $\mathbf{f}_{II}^{*}$.

One might note at this point that, in the present example, $\mathbf{f}_I^{**}$ is just $\mathbf{Y}_5$ normalized to unit area, and $\mathbf{f}_{II}^{**}$ is $\mathbf{Y}_1$ normalized to unit area.

Denote the elements of the vector $\mathbf{f}_I^*$ by $f_I^*(\lambda_k)$ so that $\mathbf{f}_I^* = (f_I^*(\lambda_1), f_I^*(\lambda_2), \cdots, f_I^*(\lambda_p))$, and one can think of the elements of the vector as the ordinate values of a function $f_I^*(\lambda)$ at the wavelengths λ_k, $k = 1, 2, \cdots, p$. Denote the elements of $\mathbf{f}_I^{**}$, $\mathbf{f}_{II}^{*}$, $\mathbf{f}_{II}^{**}$ in the same manner. It then follows that for all $1 \leq k \leq p$

$$\begin{aligned} f_1(\lambda_k) &= af_I^*(\lambda_k) + (1 - a)f_I^{**}(\lambda_k) \\ f_2(\lambda_k) &= bf_{II}^*(\lambda_k) + (1 - b)f_{II}^{**}(\lambda_k), \end{aligned} \tag{28}$$

for some $0 \leq a \leq 1$, $0 \leq b \leq 1$. That is, $f_1(\lambda)$ is some weighted average of $f_I^*(\lambda)$ and $f_I^{**}(\lambda)$, while $f_2(\lambda)$ is some weighted average of $f_{II}^*(\lambda)$ and $f_{II}^{**}(\lambda)$. Since the average of two numbers always lies between the two averaged values, one has

$$\min [f_I^*(\lambda_k), f_I^{**}(\lambda_k)] \leq f_1(\lambda_k) \leq \max [f_I^*(\lambda_k), f_I^{**}(\lambda_k)] \tag{29}$$

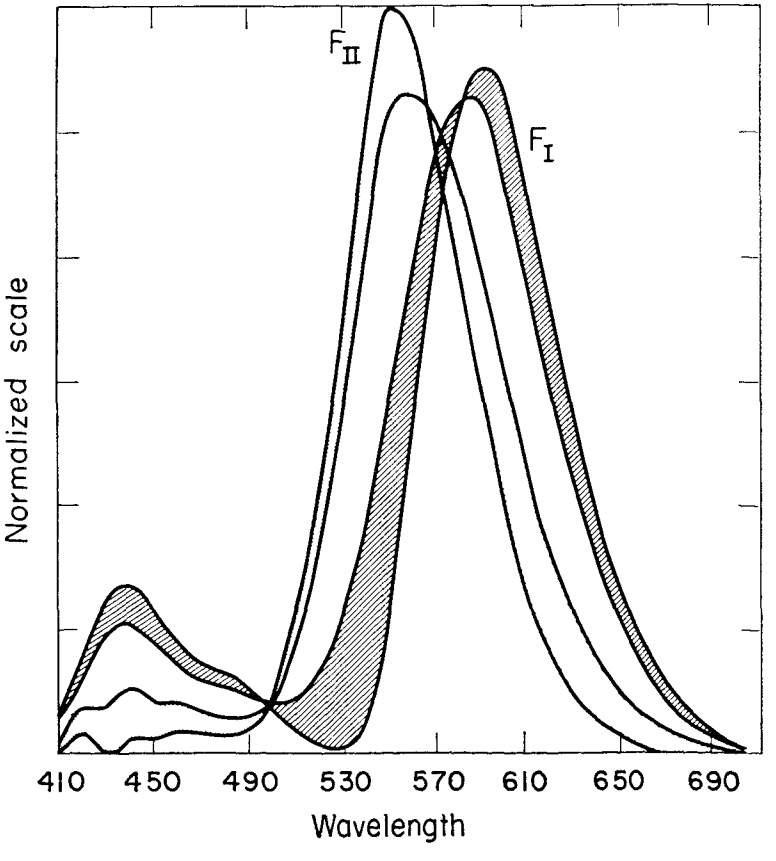
and

$$\min [f_{II}^*(\lambda_k), f_{II}^{**}(\lambda_k)] \leq f_2(\lambda_k) \leq \max [f_{II}^*(\lambda_k), f_{II}^{**}(\lambda_k)]. \tag{30}$$

Equations (29) and (30) yield a pair of function bands which contain the unknown spectrophotometric curves. If the analytic chemist uses (29) and (30) to construct the bands determined from the data of Table 1, then he obtains the bands shown in Figure 3.

At this point the chemist has determined that the 5 observed spectrophotometric curves of Figure 1 represent varying mixtures of two dyes, Dye A and Dye B. He has further determined that, whatever the spectrophotometric curve for Dye A, it must lie in the band F_I plotted in Figure 3 and given by (29). Similarly, the curve for Dye B must lie in the band F_{II} determined by (30) and plotted in Figure 3. The exact spectrophotometric curves are still unknown, but the possible curves have been restricted to rather narrow bands which determine many of the features of these $f_i(\lambda)$.

The chemist notes that the band F_{II} indicates that Dye B has a peak absorbance near 550 mμ and is slightly skewed to the right (higher wavelengths). Since the spectrophotometric curve for one of the Standard Dyes of the production process falls in this band, he concludes that Dye B is this Standard Dye. On the other hand, band F_I delineates a more complex spectrophotometric curve. There appear to be two absorbance peaks, one at 600 mμ and one at about 440 mμ. Both peaks are skewed to higher wavelengths. Since the second Standard Dye in the production process has a peak at 650 mμ, and its spec-



trophotometric curves does not fall in F_I it follows that Dye A is not the second Standard Dye. Knowing the general form of Dye A's spectrophotometric curve and its peak absorbances will aid the chemist in identifying the dye's structure.

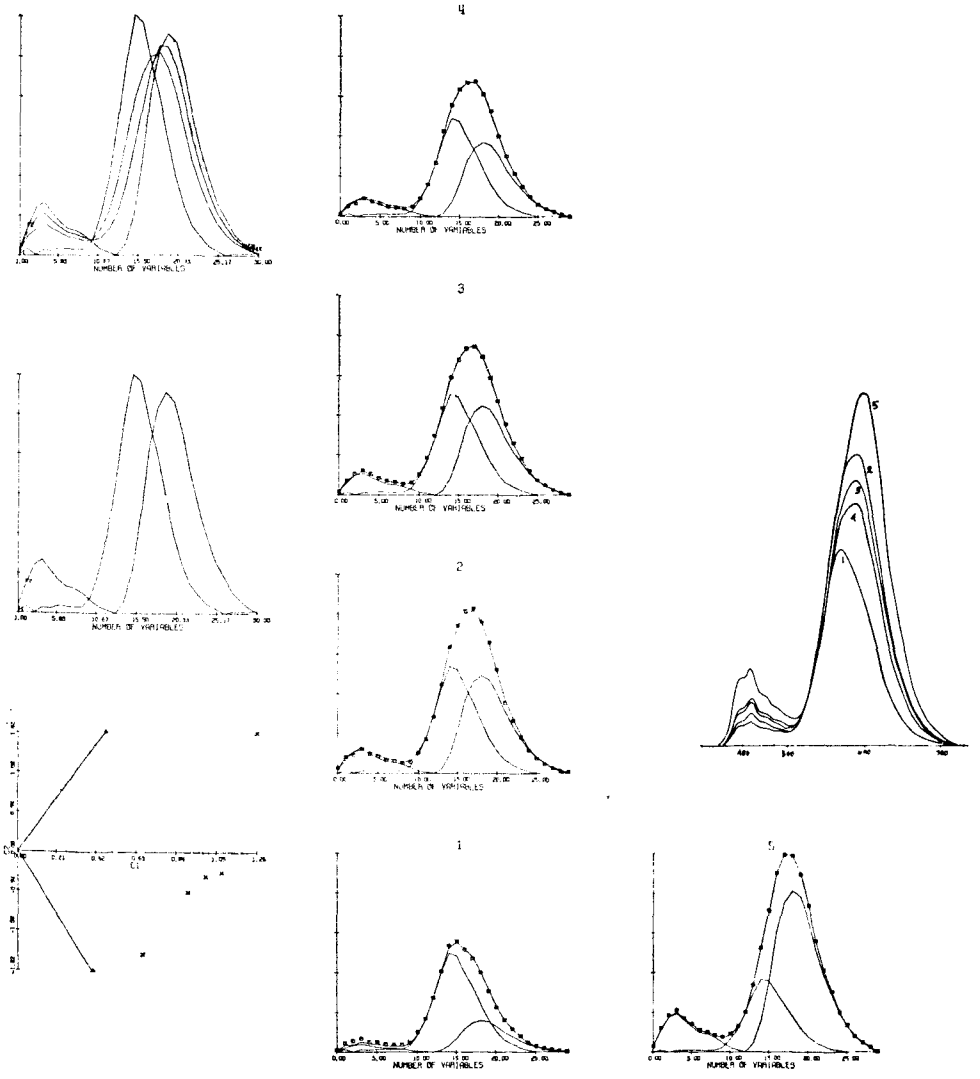
In this instance the bands were reasonably narrow, and allowed the chemist to determine a great deal about the five solutions and the process from which they came. However, $f_I^{**}(\lambda)$ and $f_{II}^{**}(\lambda)$ will always be one of the observed $Y_i(\lambda)$ normalized to unit area. The narrowness of the bands then depends on obtaining observed sample mixtures which are close to "pure" $f_1(\lambda)$ and $f_2(\lambda)$.

If the chemist is willing to make one more assumption, then he need not be satisfied with bands of solutions. If he assumes that $f_1(\lambda)$ is zero for some λ_r^* at which $f_2(\lambda)$ is nonzero, and $f_2(\lambda)$ zero at some λ_m^* where $f_1(\lambda)$ is nonzero, then the exact underlying functions $f_1(\lambda)$ and $f_2(\lambda)$ may be determined. It can be shown that if there is some λ_r^* and λ_m^* such that

$$\begin{aligned} f_1(\lambda_r^*) &= 0 \quad \text{and} \quad f_2(\lambda_r^*) > 0, \\ f_1(\lambda_m^*) &> 0 \quad \text{and} \quad f_2(\lambda_m^*) = 0, \end{aligned} \tag{31}$$

TABLE 3

F_I^*	F_{II}^*
0.0000	0.0004
0.0005	0.0015
0.0011	0.0025
0.0005	0.0028
0.0004	0.0023
0.0005	0.0018
0.0004	0.0014
0.0004	0.0013
0.0004	0.0010
0.0009	0.0007
0.0024	0.0003
0.0046	0.0001
0.0077	0.0000
0.0112	0.0006
0.0145	0.0017
0.0137	0.0051
0.0117	0.0087
0.0098	0.0109
0.0074	0.0117
0.0051	0.0112
0.0033	0.0096
0.0020	0.0074
0.0012	0.0056
0.0007	0.0042
0.0003	0.0028
0.0000	0.0020
0.0001	0.0011
0.0001	0.0007
0.0001	0.0004
0.0001	0.0000



then

$$f_1(\lambda_k) = f_I^*(\lambda_k) \quad \text{and} \quad f_2(\lambda_k) = f_{II}^*(\lambda_k) \tag{32}$$

for all $k = 1, 2, \cdots, p$.

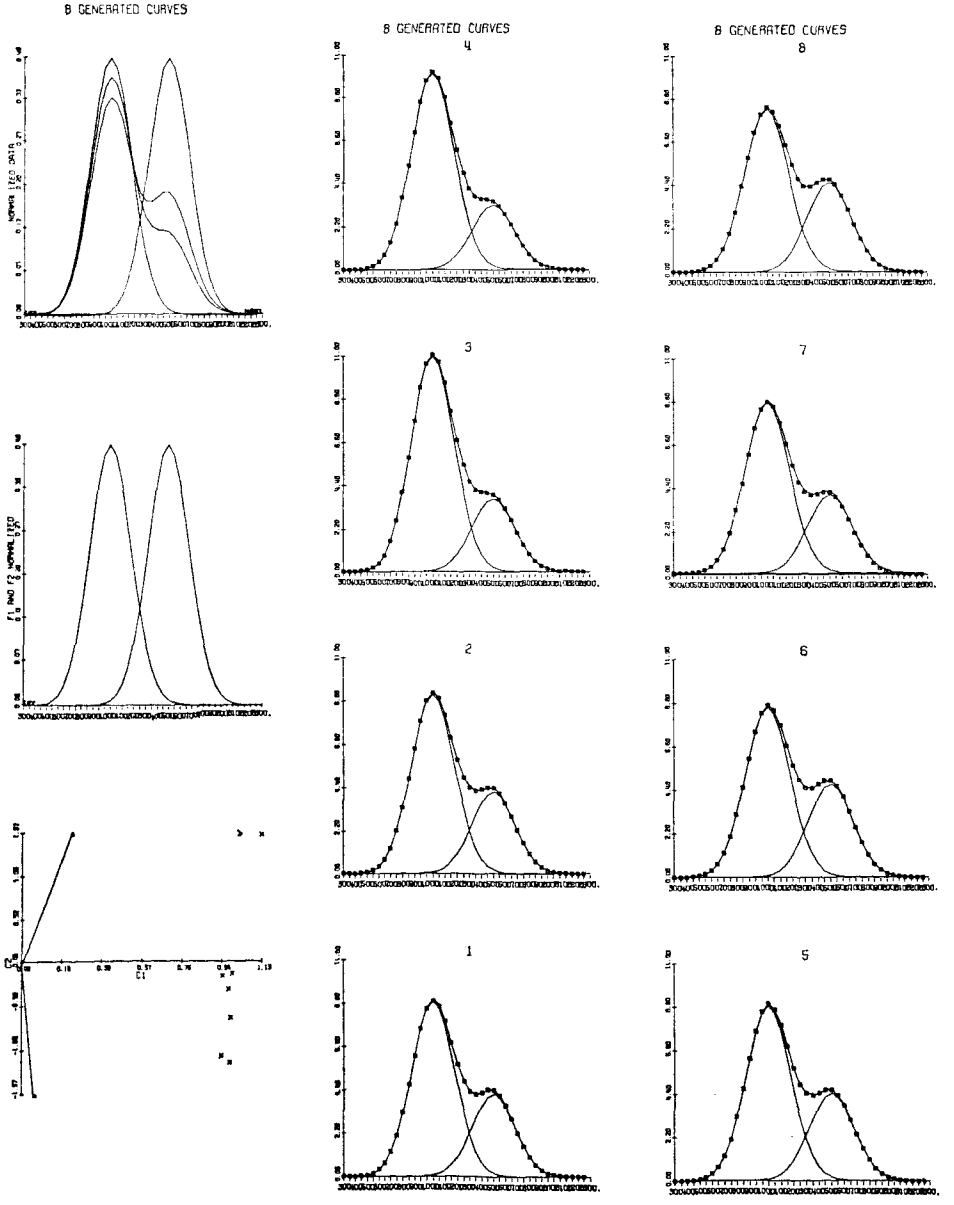
Our chemist feels that the assumption in (31) is reasonable, and so the spectrophotometric curves of Dye A and Dye B must be given by the two curves listed in Table 3 and shown in the middle lefthand plot in Figure 4. Each of the 5 observed curves is shown in Figure 4 resolved into these two components.

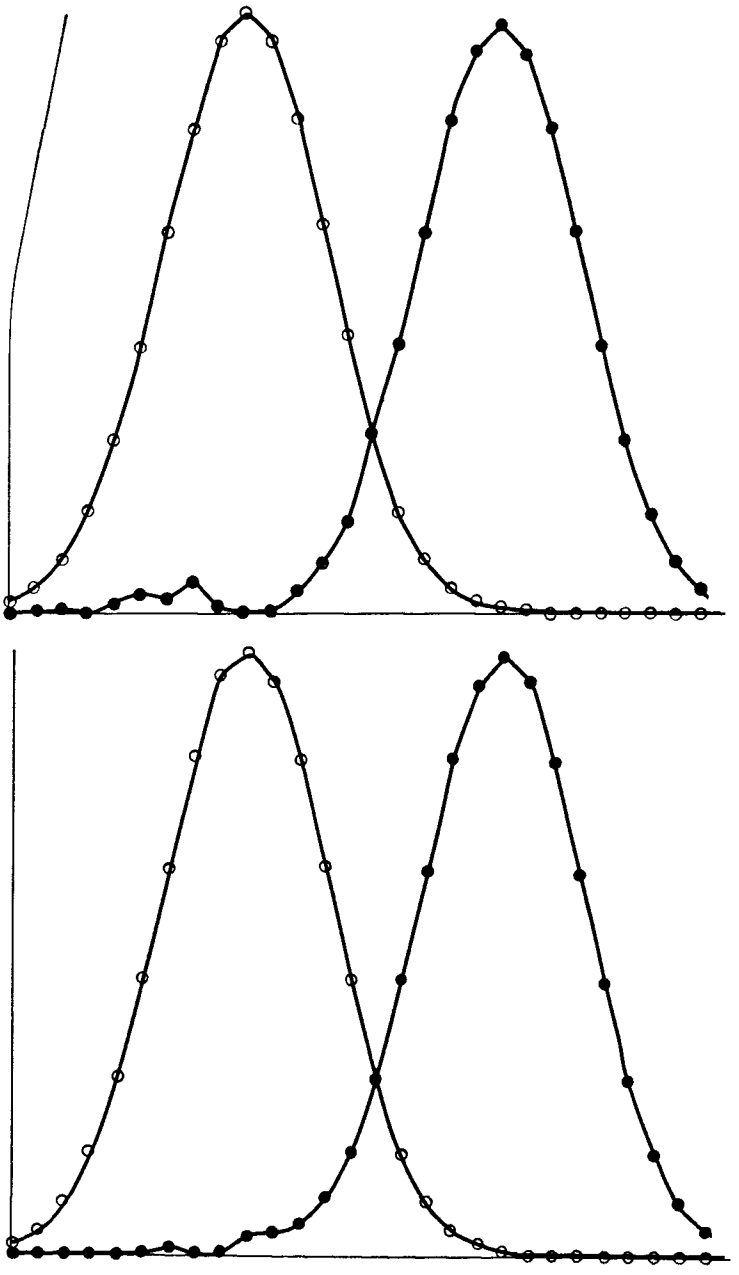
It is a relatively simple matter to build a computer program to carry out this self modeling curve resolution automatically. The input to such a program is the data matrix of (12), the output, mostly graphic, is shown in Figure 4 for the present five curve example. The lower lefthand plot is the (ξ_1, ξ_2)-plot,

above this is the plot of the "pure" components under the assumption in (31), and above that are the two solution bands F_I and F_{II} . The plots to the right show the raw data plotted as small squares with the least squares fit of the model $\alpha_i f_I^*(\lambda) + \beta_i f_{II}^*(\lambda)$ shown as the solid line. The pure components $\alpha_i f_I^*(\lambda)$ and $\beta_i f_{II}^*(\lambda)$ are also shown in these same plots. A plot of all five of the original spectrophotometric curves is shown at the far right.

4. Additional Remarks

Up to this point actual chemical data have been used. The next analysis,





however, is based on computer generated data. The data have no error other than that introduced by roundoff error. 8 random mixtures of two Gaussian components were generated. One component is centered at 450 mμ and the other at 550 mμ. The curves and their analysis are shown in Figure 5. Since none of the observed curves was anywhere near a "pure dye", the function bands F_I and F_{II} are extremely wide as shown in the upper lefthand plot in Figure 5. Since the assumption in (31) is satisfied, the exact underlying Gaussian components are recovered as shown in the middle lefthand plot in Figure 5.

Thus, even though one observed nothing close to a pure "dye" it was still possible to recover the pure underlying components.

It has been noted that the results of Section 4 hold only as limits in probability, if the observed spectrophotometric vectors are subject to random error. In the presence of error the solution bands F_I and F_{II} become a kind of confidence bands whose probability of containing the unknown $f_i(\lambda)$ approaches 1 as m, the number of replicate readings of $Y_i(\lambda)$, approaches infinity. Precise statements about these "confidence bands" and their distributions are not easily obtained. The distribution problems in principal component analysis, see [2], are difficult, and the present work is further complicated by rotation of these eigenvectors.

Experimental error was added to the 8 random mixtures of the two Gaussian components used in the previous example. The raw data for the analysis is then $\mathbf{Y}_i^* = \mathbf{Y}_i + \mathbf{e}_i$. The elements of the error vector, the e_{ik}, are independent, Normally distributed with mean 0 and standard deviation $0.01y_{ik}$. Negative elements of $\mathbf{Y}_i^*$ were set to 0. Figure 5 showed the analysis of the 8 curves without error. Figure 6 shows the "pure components" obtained with 1% error. The only difference in the two analyses is the appearance of a small amount of noise in the tail of the component centered at 550 mμ. If one makes $m = 5$ replicate readings of the 8 curves, the noise in the tail is reduced.

It is the relative error, not the absolute error, which is critical in this analytic technique. Large errors are tolerable in regions where both components are large. However, even small errors in a region where both components are small can produce considerable noise in the component estimates and bands. *It is wise to avoid including extreme tail areas in the use of this technique, since it is these areas in which the two components are both small and their relative error likely to be large.*

The report [9] contains additional examples of applications of this technique, and also contains proofs of the stated mathematical results. It may be obtained by writing the authors.

Acknowledgement

Our thanks to Mary S. Maggio for aiding us in the simulation work, and to Dr. Robert Tuite for bringing this class of problems to our attention.

References

[1] Anderson, T. W. (1958). *Introduction to Multivariate Analysis.* John Wiley and Sons, New York.

[2] Anderson, T. W. (1963). Asymptotic Theory for Principal Component Analysis. *Ann. Math. Statist. 34,* 122-148.

[3] Kendall, M. G. (1957). *A Course in Multivariate Analysis.* Chas. Griffin and Co.

[4] Pitha and Jones, (1967). Evaluation of Mathematical Functions to Fit IR Band Envelopes. *Canadian Journal Chem., 45,*.

[5] Rao, C. R. (1964). The Use and Interpretation of Principal Component Analysis in Applied Research. Tech. Report No. 9, Dept. of Stat., Stanford University.

[6] SIMONDS, J. L. (1963). Application of Characteristic Vector Analysis to Photographic and Optical Response Data. *JOSA 53*, 968–974.

[7] WERNIMONT, G. T. (1947). Evaluating Laboratory Performance of Spectrophotometers. *Jour. Anal. Chem., 39*, 554–562.

[8] WEST, W. (1956). Techniques for Organic Chemistry; Applications of Spectroscopy. Interscience 9.

[9] LAWTON, W. H. and SYLVESTRE, E. A. (1969). Self Modeling Curve Resolution. *MSDD/ Kodak Technical Report.*