Self Modeling Nonlinear Regression
Author(s): W. H. Lawton, E. A. Sylvestre and M. S. Maggio
Source: *Technometrics*, Vol. 14, No. 3 (Aug., 1972), pp. 513-532
Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association and American Society for Quality
Stable URL: http://www.jstor.org/stable/1267281
Accessed: 27-06-2016 06:50 UTC

Self Modeling Nonlinear Regression

W. H. LAWTON, E. A. SYLVESTRE, AND M. S. MAGGIO

Eastman Kodak Company
Kodak Park Division
Rochester, New York

The paper is concerned with parametric models for populations of curves; i.e. models of the form $y_i(x) = f(\theta_i; x) +$ error, $i = 1, 2, \cdots, n$. The shape invariant model $f(\theta_i; x) = \theta_{0i} + \theta_{1i}g([x - \theta_{2i}]/\theta_{3i})$ is introduced. If the function $g(x)$ is known, then the θ_i may be estimated by nonlinear regression. If $g(x)$ is unknown, then the authors propose an iterative technique for simultaneous determination of the best $g(x)$ and θ_i. Generalizations of the shape invariant model to curve resolution are also discussed. Several applications of the method are also presented.

KEY WORDS

Nonlinear Regression
Mathematical Modeling
Curve Resolution
Least Square Splines

1. INTRODUCTION

A large class of problems in the physical sciences give rise to experiments whose observed response is a continuous curve. This is the situation when, for example, one studies hearing response, or EKG's in the human population, or when one measures spectrophotometric curves from sampled product, or observes reaction curves in designed chemical experiments.

The analysis of such experiments poses certain problems because the bulk of the statistical procedures is concerned with the analysis of populations in which the physical observable is a single numerical value, or a vector of such values, rather than an entire curve. For this reason one has usually sought a parametric model for the total population of curves. By using such a model, one can associate a vector of parameter values with each possible observable curve in the population. The study of the population of curves is then reduced to a study of the population of derived parameters as in Box [2], Church [3], or Mandel [8].

It is not surprising, therefore, to find considerable interest in the construction of parametric models

$$y(x) = f(\theta_1, \theta_2, \cdots, \theta_r; x) + \text{error} \tag{1}$$

and in the estimation of the parameters $\theta_1, \cdots, \theta_r$. Once one has selected an

Received Nov. 1970; revised May 1971.

appropriate parametric function $f(\boldsymbol{\theta}; x)$, any one of a number of algorithms will yield least squares estimates of the parameter vector $\boldsymbol{\theta}$. These algorithms involve linear or nonlinear regression and are discussed at length in [4], [5], [6], and [9].

While the estimation of $\boldsymbol{\theta}$ has received considerable attention, the selection of a satisfactory parametric function has received relatively little attention. This paper deals with an empirical selection of this parametric function.

2. The Shape Invariant Model

The basis of the analysis will be a population of observed curves $y_i(x)$, $i = 1, 2, \cdots, n$. Three such populations are in the computer plots of Figures 1, 2, and 3. Figure 1 shows optical density as a function of wavenumber (spectrophotometric curves) for a dye in solutions of varying polarity. Figure 2 shows the volume of air expelled from the lungs as a function of time (Spirometer curves) for a number of different individuals. Figure 3 shows optical density of a photographic film as a function of log exposure to light (Sensitometric curves) for several developer formulations. The processes which produced these curve populations are vastly different, yet we shall see that there is a common mathematical structure.

Each population has its own characteristic curve shape; that is, the curves within a population all seem to have the same basic shape. It appears that all the curves in a given population might coincide if they were properly scaled and shifted. Such a structure would be represented by the nonlinear mathematical model

$$f(\boldsymbol{\theta}_i ; x) = \theta_{0i} + \theta_{1i} g([x - \theta_{2i}]/\theta_{3i}) \tag{2}$$

where $g(x)$ represents the characteristic shape of the response curves, and $\boldsymbol{\theta}_i$ is the parameter vector $(\theta_{0i}, \cdots, \theta_{3i})$. The model in (2) will be termed the *shape invariant model.*

The parameters θ_{1i} and θ_{3i} represent the choice of scale for the i-th curve on the y and x axes respectively. θ_{0i} and θ_{2i} are the corresponding shift, or location parameters. These parameters then have a direct graphical interpretation. For many choices of $g(x)$ these parameters also have familiar physical interpretations. Consider the meaning of the parameters when $g(x) = \sin(x)$ or $g(x) = \exp(-x^2)$.

It has been our experience that this shape invariant model in (2) is applicable to a broad class of physical processes. Using this shape invariant approach, *the problem of model selection is reduced to that of selecting the appropriate shape function $g(x)$.*

Faced with this problem, one would probably seek a simple mathematical expression for $g(x)$. For example, in the Spirometer analysis (Figure 2), one might select the function

$$g(x) = [1 - e^{-x}]u(x) \tag{3}$$

where $u(x) = 1$ for $x \geq 0$ and 0 otherwise. This function rises exponentially from 0 to an upper asymptote of 1, and has the same general features as the

observed curves of Figure 2. Since all of the observed curves have a minimum of 0, we shall take $\theta_0 = 0$ and omit this parameter from the model. The shape invariant model for the Spirometer curves is then given by

$$y_i(x) = \theta_{1i}[1 - e^{-(x-\theta_{2i})/\theta_{3i}}]u([x - \theta_{2i}]/\theta_{3i}) + \text{error} \qquad i = 1, 2, \cdots, n. \quad (4)$$

Assume that the curve $y_i(x)$ has been read at the p_i points $x_{i1}, x_{i2}, \cdots, x_{ip_i}$. The least squares estimate of $\boldsymbol{\theta}_i = (\theta_{1i}, \theta_{2i}, \theta_{3i})$ may then be obtained by performing a nonlinear regression on the i-th observed curve. Let $\hat{\boldsymbol{\theta}}_i$ denote these least squares estimates, then the error sum of squares for the i-th curve

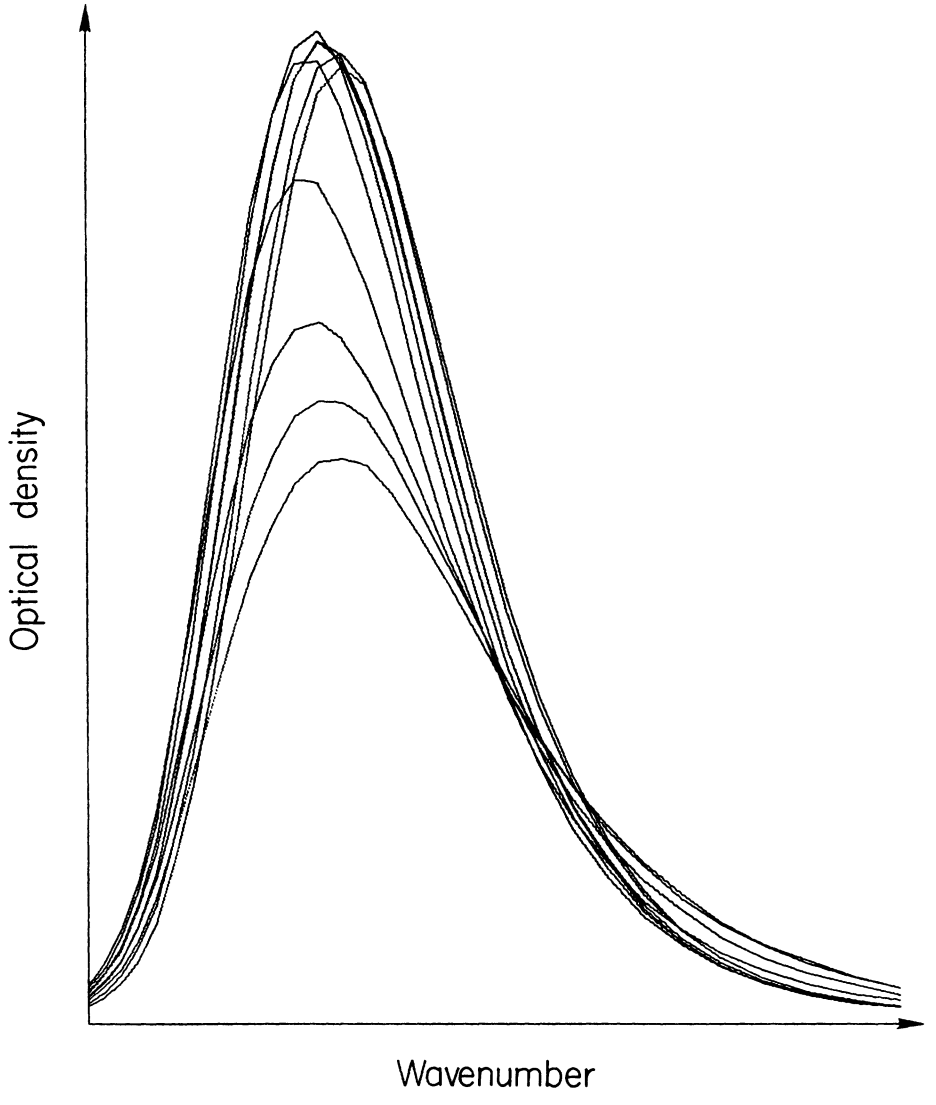

FIGURE 1

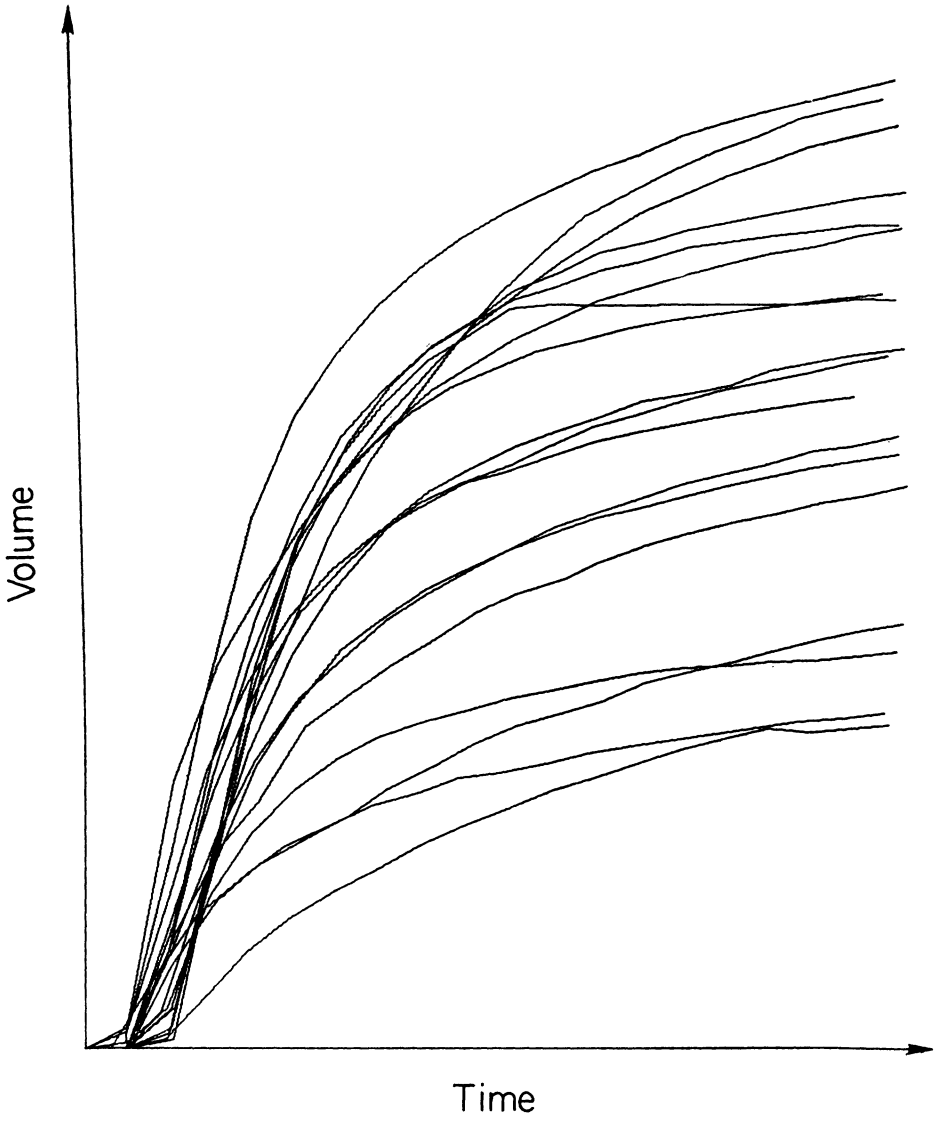

FIGURE 2

is given by

$$Q_i = \sum_{j=1}^{p_i} [y_i(x_{ij}) - f(\hat{\theta}_i ; x_{ij})]^2 \tag{5}$$

where $f(\theta; x)$ is the parametric model used in (4). The average error sum of squares for the population as a whole is then

$$\bar{Q} = \frac{1}{n} \sum_{i=1}^{n} Q_i , \tag{6}$$

and is a convenient measure of the adequacy of the model in fitting the population. $\bar{Q} = 0$ would imply the model fits all the observed curves perfectly.

Column 1 of Table 1 shows Q_i and $\bar{Q}$ for the analysis of the Spirometer curves using this model. It would appear that (4) is a reasonably good model for the population. However, other choices for $g(x)$ might yield a smaller $\bar{Q}$.

If the population of curves satisfies the shape invariant model in (2), then any of the individual curves $y_i(x)$ would provide an acceptable shape function $g(x)$. Select one of these observed curves, say $y_1(x)$, then the shape invariant

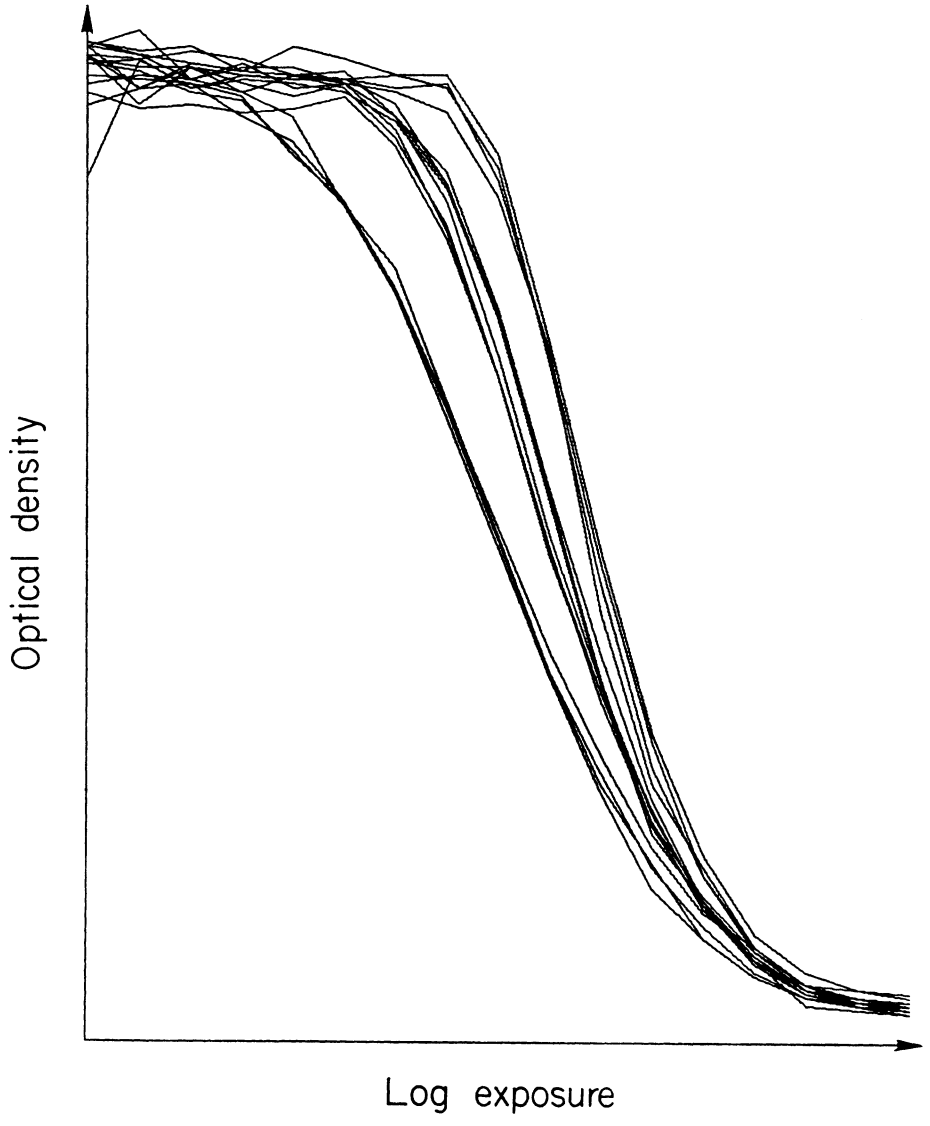

FIGURE 3

TABLE 1

CURVE	MODEL 1 (Eq. 4) Q_i	MODEL 2 (Eq. 7) Q_i	"BEST" MODEL Q_i
1	0.048	0.000	0.012
2	0.105	0.092	0.074
3	0.139	0.022	0.021
4	0.079	0.032	0.023
5	0.086	0.010	0.032
6	0.089	0.106	0.035
7	0.062	0.080	0.058
8	0.033	0.088	0.060
9	0.017	0.032	0.018
10	0.172	0.411	0.159
11	0.082	0.061	0.010
12	0.109	0.111	0.078
13	0.044	0.103	0.030
14	0.010	0.136	0.089
15	0.331	0.221	0.133
16	0.163	0.040	0.037
17	0.129	0.109	0.108
18	0.053	0.022	0.017
$\bar{Q}$	0.097	0.093	0.055

model for the population becomes

$$y_i(x) = f(\boldsymbol{\theta}_i ; x) + \text{error} \tag{7}$$

with

$$f(\boldsymbol{\theta}; x) = \theta_1 y_1([x - \theta_2]/\theta_3). \tag{8}$$

The model for the population is now in terms of one of the members of the population. Once again nonlinear regression will yield $\hat{\boldsymbol{\theta}}_i$, Q_i and $\bar{Q}$ provided we can obtain an analytic expression for $y_1(x)$. This expression need not be some simple algebraic form; all one requires is some rule that associates a value of $y_1(x)$ with each possible value of x.

One possible approach to this problem parallels the use of a function generator in an analog computer. Using the approach described in Bellman [1], one can combine table-lookup with linear interpolation to produce what we shall call an *empirical function*. Bellman calls these polygonal curves; one could also call them 1-st order splines. Such functions will be denoted by a "*". Thus the

empirical function approximating $y_1(x)$ will be denoted by $y_1^*(x)$, and is obtained from $y_1(x)$ in the following manner.

From a plot of $y_1(x)$ select m points from the curve in such a way that $y_1(x)$ is approximately linear between the selected points. Call these points $(x_1^*, y_1(x_1^*)), \cdots, (x_m^*, y_1(x_m^*))$. The empirical approximating function $y_1^*(x)$ is then given by

$$y_1^*(x) = \begin{cases} \sum_{k=1}^{m-1} \left(\frac{y_1(x_{k+1}^*) - y_1(x_k^*)}{x_{k+1}^* - x_k^*} [x - x_k^*] + y_1(x_k^*) \right) \Delta_k(x), & x_1^* < x \leq x_m^* \\ y_1(x_1^*), & x \leq x_1^* \\ y_1(x_m^*), & x > x_m^* \end{cases} \tag{9}$$

where

$$\Delta_k(x) = \begin{cases} 1 & \text{for } x_k^* < x \leq x_{k+1}^* \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$y_1^*(x)$ is, of course, simply the function obtained by connecting the points $(x_k^*, y_1(x_k^*))$ by straight lines within the region $x_1^* \leq x \leq x_m^*$ and setting it constant outside this region. This approximation works well provided the function f being approximated satisfies the condition

$$df(x)/dx \cong 0 \text{ outside some finite interval.} \tag{11}$$

Under these conditions $y_1^*(x)$ will be close to $y_1(x)$ in just about any reasonable norm provided the m points $x_1^*, \cdots, x_m^*$ are reasonably selected (as, for example, in Bellman [1]). One might note in passing that m points $x_1^*, \cdots, x_m^*$ used to define the empirical function $y_1^*(x)$ are not necessarily the same as the p_i points $x_{i1}, \cdots, x_{ip_i}$ used in (5) for nonlinear regression.

Using the model given by (7) and the shape function (9), one can carry out the n nonlinear regressions to determine $\hat{\boldsymbol{\theta}}_i$, Q_i, and $\bar{Q}$. Column 2 of Table 1 shows the Q_i and $\bar{Q}$ obtained with the empirical function model. There is very little difference in the value of $\bar{Q}$ from that in our earlier model.

However, insofar as $y_1(x)$ is subject to error, the model in (8) may not give the best overall fit to the population. The choice of some other observed curve $y_i(x)$ as the basic shape function might yield a smaller $\bar{Q}$. What one would like to have, however, is a best choice for the shape function $g(x)$; that is, the shape function $g(x)$ *which yields the smallest possible* $\bar{Q}$.

3. A Best Possible Shape Function

Let $y_i(x_{ij})$, $i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, p_i$ denote the n observed curves where the i-th curve is read at the p_i points x_{ij}. Assume that these curves satisfy our shape invariant model

$$f(\boldsymbol{\theta}_i ; x) = \theta_{0i} + \theta_{1i} g([x - \theta_{2i}]/\theta_{3i}) \tag{12}$$

and condition (11). Given any particular choice of the shape function g, let $\hat{\boldsymbol{\theta}}_i(g) = (\hat{\theta}_{0i}(g), \cdots, \hat{\theta}_{3i}(g))$ denote the vector of least squares parameter

estimates; that is, the vector $\boldsymbol{\theta}_i$ which minimizes

$$\sum_{j=1}^{p_i} [y_i(x_{ij}) - f(\boldsymbol{\theta}_i ; x_{ij})]^2 \tag{13}$$

with $f(\boldsymbol{\theta}_i ; x)$ given by (12).

The average error sum of squares for the population as a whole is then given by

$$\bar{Q}(g) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p_i} [y_i(x_{ij}) - f(\hat{\boldsymbol{\theta}}_i(g); x_{ij})]^2. \tag{14}$$

For each choice of the function g, there is a corresponding average error sum of squares $\bar{Q}(g)$. Let g_1 and g_2 denote two possible choices for the shape function g. If $\bar{Q}(g_1) < \bar{Q}(g_2)$, then it is clear that the shape invariant model with g_1 is to be preferred over the shape invariant model with g_2. By the same token, the best possible choice for g would be the one which minimizes (14). *No other choice of g could lead to a better shape invariant model.* On the other hand, if $\bar{Q}(g_1) = \bar{Q}(g_2)$, then we shall say that g_1 and g_2 are *equivalent.* Now if $g_\alpha(x) = \alpha_0 + \alpha_1 g([x - \alpha_2]/\alpha_3)$, then the collection of functions

$$\{g_\alpha : \alpha_1 > 0, \alpha_3 > 0, -\infty < \alpha_0 < \infty, -\infty < \alpha_2 < \infty\} \tag{15}$$

are all equivalent to the function g. Different choices of $\boldsymbol{\alpha}$ merely correspond to different linear transformations of the $\boldsymbol{\theta}$ vector in the same shape invariant model.

In searching for the best shape function g, it will be necessary to consider only one function from each equivalence class defined by (15). By specifying a location and scale normalization on g, it is possible to select one function from each of the equivalence classes. There are several ways in which this normalization may be specified. If it is properly specified, the $\boldsymbol{\theta}_i$ in (12) will have easily interpreted graphical meaning.

If the observed curves are unimodal as in Figure 1, then a convenient normalization requires that g satisfy

$$\begin{aligned} \min_x [g(x)] &= 0 \\ \max_x [g(x)] &= 1 \\ x_{\max} &= 0 \\ \Delta x &= 1 \end{aligned} \tag{16}$$

where $x_{\max}$ is the x value for which $g(x)$ takes on its maximum value, and Δx is the width of the unimodal function at half its maximum height.

If, on the other hand, the curves are monotonic as in Figures 2 and 3, then a convenient set of normalization conditions is given by

$$\begin{aligned} \min_x [g(x)] &= 0 \\ \max_x [g(x)] &= 1 \\ x_{\frac{1}{2}} &= 0 \\ g'(x_{\frac{1}{2}}) &= 1 \end{aligned} \tag{17}$$

where $x_{\frac{1}{2}}$ is the x value for which g reaches half maximum height, and $g'(x)$ denotes the derivative of g at x.

Whatever normalization is selected, let G denote the collection of normalized (representative) shape functions. The function $g \in G$ which minimizes (14) will be called the *best shape function subject to* G, and will be denoted by $\hat{g}$.

Let G^* denote the collection of empirical functions

$$g^*(x) = \begin{cases} \sum_{k=1}^{m-1} \left(\frac{g_{k+1} - g_k}{x_{k+1}^* - x_k^*} [x - x_k^*] + g_k \right) \Delta_k(x), & \text{for } x_1^* < x \leq x_m^* \\ g_1 \quad \text{for } x \leq x_1^* \\ g_m \quad \text{for } x > x_m^* \end{cases} \tag{18}$$

which are in G. G^* is contained in G. In practice then we shall determine the best shape function subject to G^* rather than the best shape function subject to G. That is, we shall minimize (14) subject to $g \in G^*$ rather than G. Let $\hat{g}^*$ denote this best empirical shape function. Since the empirical functions defined by (18) are extremely flexible, $\hat{g}^*$ should be very close to $\hat{g}$ in any reasonable norm, provided the x_k^*, $k = 1, \cdots, m$ are properly chosen.

Given the defining points $x_1^*, x_2^*, \cdots, x_m^*$, the function g^* in (18) is completely determined by the vector $\mathbf{g}^* = (g_1, g_2, \cdots, g_m)$. Let $\mathbf{G}^*$ denote the collection of vectors $\mathbf{g}^*$ for which the corresponding empirical function $g^* \in G^*$. The problem of determining the best empirical shape function is then reduced to that of determining the vector $\mathbf{g}^*$ which minimizes (14) subject to the restriction $\mathbf{g}^* \in \mathbf{G}^*$. If one rewrites (14) in terms of the vector $\mathbf{g}^*$, then one has

$$\bar{Q}(\mathbf{g}^*) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p_i} [y_i(x_{ij}) - f^*(\hat{\boldsymbol{\theta}}_i(\mathbf{g}^*); x_{ij})]^2 \tag{19}$$

where

$$f^*(\boldsymbol{\theta}_i ; x) = \theta_{0i} + \theta_{1i} g^*([x - \theta_{2i}]/\theta_{3i}) \tag{20}$$

and $\hat{\boldsymbol{\theta}}_i(g^*)$ is the vector of least squares estimates as before. Let this optimum choice of the vector $\mathbf{g}^*$ be denoted by $\hat{\mathbf{g}}^*$.

The problem of finding the best shape invariant model has been reduced to the problem of determining a best m-dimensional vector $\hat{\mathbf{g}}^*$, and, for each observed curve, the associated best parameter values $\hat{\boldsymbol{\theta}}_i(\hat{\mathbf{g}}^*)$. This simultaneous determination of a best shape function $\hat{g}^*$ and its associated best parameter values will be called *self modeling nonlinear regression.*

The vector $\hat{\mathbf{g}}^*$ and its associated $\boldsymbol{\theta}_i(\hat{\mathbf{g}}^*)$ may be determined by the iterative procedure described in Section 6.

Given a population of n observed curves $y_i(x)$, $i = 1, \cdots, n$ read at the points $x_{i1}, \cdots, x_{ip_i}$, the analysis proceeds in the following manner. Select the points $x_1^*, x_2^*, \cdots, x_m^*$ at which the optimal empirical function $\hat{g}_i$ is to be determined. One then provides an initial guess for the n vectors $\hat{\boldsymbol{\theta}}_i(\hat{\mathbf{g}}^*)$. Each iteration of the procedure uses an m-parameter linear regression and n 4- parameter nonlinear regressions to produce improved estimates of $\hat{\mathbf{g}}^*$ and $\hat{\boldsymbol{\theta}}_i(\hat{\mathbf{g}}^*)$. The iterations continue until the estimate of $\hat{\mathbf{g}}^*$ has converged to the specified number of decimal places. Despite the considerable amount of computation

required, a typical analysis takes only about 30 seconds of computer time on an IBM S/360 MOD65.

It might be of interest to note that this problem of determining a "best" vector and its associated parameter values also arises in principal component analysis. In fact, if we drop the parameters θ_{0i} , θ_{2i} , and θ_{3i} from the model in (20), then we have the simple linear model.

$$f^*(\theta_{1i} ; x) = \theta_{1i} g^*(x). \tag{21}$$

If we then take for $\mathbf{G}^*$ the vectors of unit length, read all the $y_i(x)$ at the same x values x_1 , x_2 , $\cdots$, x_p , and choose the $x_k^* = x_k$, $k = 1, 2, \cdots, m = p$, then $\mathbf{g}^*$ is just the eigenvector associated with the largest eigenvalue of the second moment matrix $[Y'Y]/n$. Here Y is the matrix with elements $y_i(x_j)$. The parameters $\hat{\theta}_{1i}(\hat{\mathbf{g}}^*)$ are simply the principal components.

In the sense that the shape invariant model in (20) is a generalization of the model in (21), one can say that the best shape invariant model is a generalization of a one vector principal component analysis.

4. Applications

In this section the iterative technique of Section 6 is applied to the populations shown in Figures 1, 2, and 3.

Spirometer Application: The $n = 18$ Spirometer curves in Figure 2 are all monotone increasing functions. None of the observed curves exhibit any baseline, and so the parameter θ_0 will be dropped from the model for this application. The shape invariant model becomes

$$y_i(x) = f(\boldsymbol{\theta}_i ; x) + \text{error} \tag{22}$$

where

$$f(\boldsymbol{\theta}_i ; x) = \theta_{1i} g([x - \theta_{2i}]/\theta_{3i}). \tag{23}$$

The normalization in (17) is appropriate and yields

$$\begin{aligned} \theta_{1i} &= \text{Lung volume of the } i\text{-th individual} \\ \theta_{2i} &= \text{Time at which half the air has been expelled from the lungs of } i\text{-th individual} \\ \theta_{3i} &= \text{Rate of explusion at time } \theta_{2i} \text{ for } i\text{-th individual.} \end{aligned} \tag{24}$$

Each observed curve $y_i(x)$ was read at 20 points. The x_{ij} and the associated $y_i(x_{ij})$ are shown in Table 2. The next step is to select the m points x_1^* , $\cdots$, x_m^* at which g^* is to be defined. Perhaps the easiest method of selecting the x_k^* is to normalize one of the observed curves according to the conditions in (17) and then select the x_k^* which yield a good approximation to this normalized curve. The selection may be by eye from a plotted curve or by Bellman's procedure.

The next step is to obtain starting estimates for the $\hat{\boldsymbol{\theta}}_i(\hat{g}^*)$. Since these parameters have the geometric interpretation in (24), starting estimates $\boldsymbol{\theta}_i^{(0)}$ are easily obtained. These starting values and the $n = 18$ curves are fed into the iterative procedure along with the x_k^* . Each iteration produces improved

TABLE 2
Spirometer Curves

1	X	0.0	0.18	0.51	0.81	1.16	1.44	1.76	2.08	2.40	2.70	3.00	3.34	3.64	3.95	4.21	4.57	4.87	5.17	5.49	5.80
	Y	0.0	0.0	0.58	1.05	1.37	1.59	1.75	1.91	2.01	2.13	2.17	2.24	2.32	2.38	2.42	2.47	2.52	2.57	2.59	2.63
2	X	0.0	0.32	0.63	0.96	1.26	1.60	1.91	2.20	2.52	2.81	3.13	3.39	3.74	4.05	4.37	4.68	5.01	5.35	5.64	5.93
	Y	0.0	0.0	0.66	1.17	1.45	1.65	1.83	2.05	2.22	2.39	2.54	2.62	2.73	2.87	2.96	3.03	3.13	3.22	3.28	3.33
3	X	0.0	0.30	0.62	0.93	1.24	1.49	1.86	2.16	2.49	2.77	3.09	3.38	3.72	3.99	4.36	4.63	4.98	5.26	5.60	5.90
	Y	0.0	0.0	0.31	1.43	2.22	2.64	3.14	3.38	3.62	3.76	3.92	4.04	4.17	4.23	4.34	4.41	4.49	4.55	4.61	4.66
4	X	0.0	0.33	0.65	0.98	1.25	1.60	1.90	2.23	2.54	2.85	3.13	3.48	3.73	4.08	4.37	4.70	5.00	5.33	5.63	5.96
	Y	0.0	0.0	0.90	1.59	1.94	2.53	2.77	3.02	3.22	3.43	3.57	3.69	3.82	3.95	4.04	4.13	4.19	4.28	4.33	4.41
5	X	0.0	0.30	0.62	0.95	1.24	1.55	1.85	2.18	2.47	2.81	3.11	3.42	3.70	4.06	4.34	4.66	4.93	5.29	5.56	5.90
	Y	0.0	0.0	0.76	1.67	2.27	2.75	3.05	3.36	3.56	3.77	3.94	4.10	4.20	4.33	4.43	4.51	4.58	4.69	4.72	4.80
6	X	0.0	0.30	0.29	0.61	0.88	1.18	1.49	1.79	2.12	2.41	2.76	3.04	3.38	3.70	4.03	4.30	4.66	4.95	5.28	5.58
	Y	0.0	0.0	0.11	1.19	2.18	2.87	3.40	3.76	4.05	4.23	4.44	4.52	4.65	4.76	4.84	4.90	4.97	5.02	5.07	5.11
7	X	0.0	0.30	0.62	0.94	1.23	1.53	1.85	2.16	2.49	2.76	3.09	3.38	3.71	4.01	4.32	4.68	4.95	5.22	5.59	5.89
	Y	0.0	0.0	0.16	1.55	2.99	3.97	4.64	5.05	5.40	5.58	5.81	5.84	5.83	5.83	5.84	5.84	5.84	5.84	5.88	5.87
8	X	0.0	0.30	0.64	0.92	1.24	1.55	1.86	2.16	2.50	2.80	3.11	3.41	3.76	4.04	4.39	4.68	5.00	5.28	5.62	5.91
	Y	0.0	0.0	0.96	2.29	3.38	4.20	4.80	5.16	5.49	5.69	5.88	5.98	6.11	6.18	6.29	6.33	6.38	6.41	6.46	6.46
9	X	0.0	0.26	0.58	0.89	1.17	1.49	1.79	2.11	2.42	2.74	3.06	3.38	3.68	4.00	4.32	4.62	4.94	5.24	5.56	5.83
	Y	0.0	0.0	0.07	0.43	0.76	1.03	1.23	1.41	1.60	1.75	1.89	2.03	2.13	2.24	2.35	2.43	2.51	2.48	2.51	2.54
10	X	0.0	0.30	0.65	0.95	1.26	1.56	1.91	2.21	2.52	2.86	3.12	3.42	3.74	4.12	4.36	4.67	4.98	5.31	5.64	5.96
	Y	0.0	0.0	0.33	1.51	2.86	4.03	4.75	5.19	5.51	5.74	5.95	6.08	6.25	6.34	6.42	6.48	6.54	6.61	6.67	6.71
11	X	0.0	0.30	0.60	0.94	1.23	1.57	1.88	2.18	2.49	2.81	3.11	3.41	3.75	4.04	4.37	4.67	4.97	5.33	5.59	5.95
	Y	0.0	0.0	0.94	2.05	2.79	3.43	3.79	4.10	4.37	4.56	4.72	4.84	4.96	5.08	5.14	5.20	5.32	5.39	5.44	5.49
12	X	0.0	0.30	0.55	0.87	1.18	1.50	1.80	2.18	2.44	2.74	3.08	3.39	3.69	3.99	4.33	4.63	4.93	5.23	5.54	5.83
	Y	0.0	0.0	0.29	1.33	2.31	3.09	3.59	4.07	4.29	4.44	4.58	4.74	4.88	4.97	5.08	5.17	5.23	5.30	5.38	5.43
13	X	0.0	0.30	0.61	0.91	1.18	1.54	1.76	2.07	2.39	2.67	3.00	3.33	3.63	3.95	4.25	4.59	4.88	5.19	5.48	5.80
	Y	0.0	0.0	0.12	1.48	2.53	3.41	4.00	4.61	5.10	5.49	5.91	6.25	6.53	6.71	6.88	7.04	7.15	7.30	7.38	7.45
14	X	0.0	0.29	0.64	0.92	1.18	1.50	1.77	2.09	2.37	2.66	2.98	3.28	3.59	3.92	4.23	4.55	4.88	5.20	5.48	5.79
	Y	0.0	0.0	0.07	1.47	2.69	3.79	4.35	4.74	4.99	5.19	5.34	5.47	5.55	5.65	5.71	5.76	5.81	5.85	5.89	5.92
15	X	0.0	0.30	0.63	0.87	1.21	1.52	1.83	2.13	2.44	2.74	3.07	3.33	3.70	4.02	4.34	4.63	4.94	5.25	5.58	5.89
	Y	0.0	0.14	1.57	2.76	4.16	4.96	5.46	5.83	6.13	6.36	6.57	6.70	6.89	7.01	7.16	7.26	7.35	7.43	7.52	7.60
16	X	0.01	0.30	0.61	0.92	1.20	1.54	1.84	2.19	2.55	2.80	3.11	3.45	3.76	4.06	4.38	4.67	5.00	5.26	5.60	5.91
	Y	0.0	0.0	0.78	2.14	3.06	3.96	4.46	4.95	5.38	5.64	5.90	6.17	6.38	6.57	6.71	6.85	6.97	7.07	7.16	7.24
17	X	0.0	0.30	0.64	0.97	1.26	1.56	1.87	2.21	2.52	2.82	3.12	3.43	3.73	4.06	4.36	4.67	4.99	5.29	5.62	5.93
	Y	0.0	0.17	2.10	3.03	3.58	4.08	4.50	4.87	5.17	5.37	5.57	5.72	5.87	5.98	6.07	6.16	6.24	6.29	6.38	6.43
18	X	0.0	0.30	0.59	0.91	1.21	1.52	1.83	2.12	2.45	2.76	3.07	3.38	3.70	4.02	4.34	4.66	4.97	5.28	5.60	5.89
	Y	0.0	0.05	0.32	1.21	1.70	2.04	2.28	2.45	2.57	2.66	2.75	2.82	2.89	2.95	2.99	3.01	3.04	3.05	3.08	3.11

estimates of the shape function $\hat{g}$ and the parameters $\hat{\theta}_i$. Figure 4 shows the first two iterations of $\hat{g}$; the best shape function changes very little after the first two iterations. The iterative procedure halts when $\max_x |g^{*(k)} - g^{*(k+1)}| \leq 0.001$ where $g^{*(k)}$ is the best shape function at the k-th iteration.

The Q_i and $\bar{Q}(g^*)$ for this best shape invariant model are shown in Column 3 of Table 1. As expected $\bar{Q}(g^*)$ is smaller than the $\bar{Q}$ of the previous two models of Section 2. If this average squared error is acceptable, then it is possible to represent any of the observed Spirometer curves by a three parameter vector θ_i .

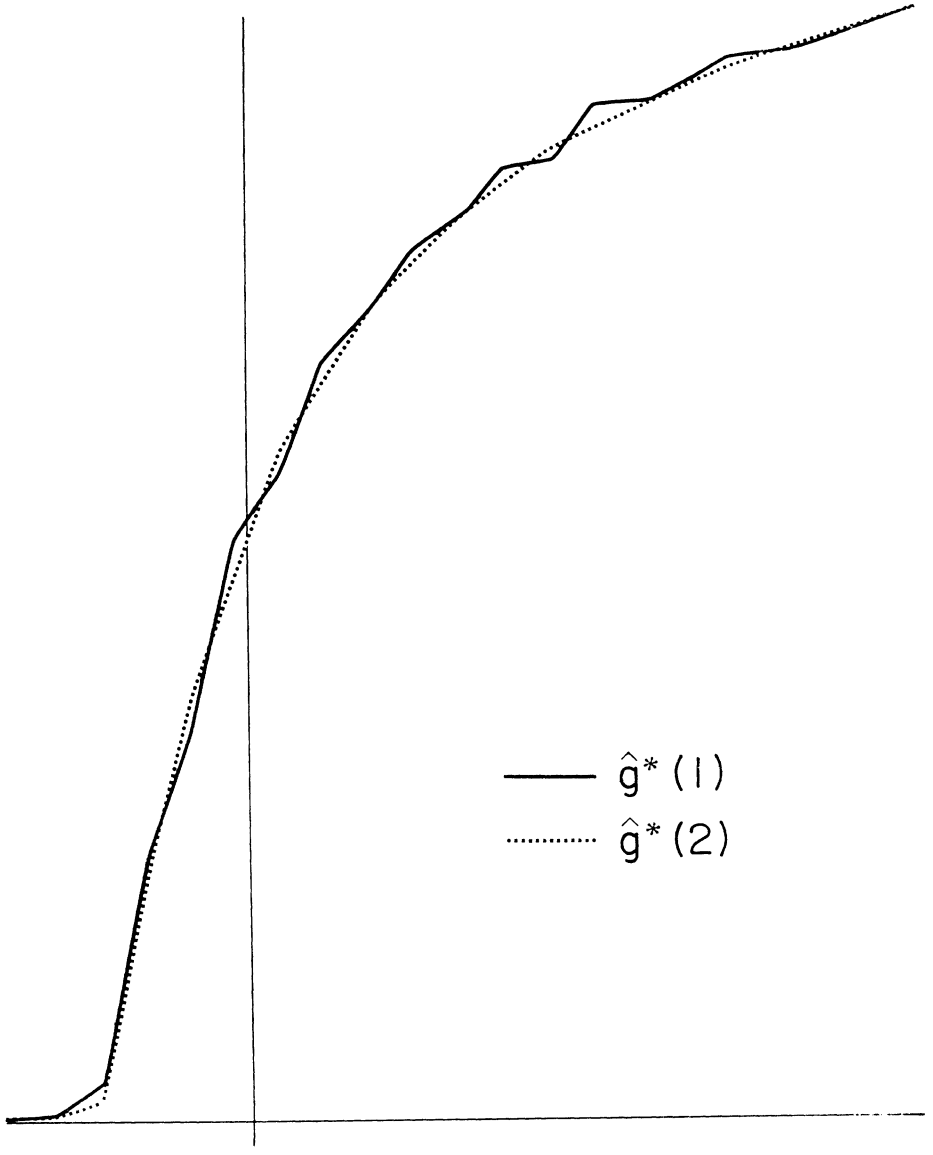

FIGURE 4

Each element of this vector has the simple interpretation given in (24). This empirically derived model was then used on other Spirometer curves, the continued low error in the squared errors indicated the shape invariant model with the g^* of Figure 4 is valid.

The computer time required for the analysis was 25 seconds.

Spectrophotometric Analysis: The $n = 10$ spectrophotometric curves in Figure 1 are all unimodal. The normalization in (16) will be used in this case. Once again we shall omit θ_0 and use the model in (22) and (23). The parameter interpretation is

$$\begin{aligned} \theta_{1i} &= \text{Maximum peak density } i\text{-th curve} \\ \theta_{2i} &= \text{Wavenumber at maximum density for } i\text{-th curve} \\ \theta_{3i} &= \text{Bandwidth at half peak height.} \end{aligned} \tag{25}$$

Each observed curve was read at the same $p = 30$ points. The x_k^* are chosen in the same manner as the last example. Initially $m = 19$ points are selected for the x_k^*. Once again starting estimates are easily obtained because of (25). Only 3 iterations, requiring 14 seconds, are needed in this case. The best shape function is shown in Figure 5. It would appear that the x_k^* were not properly selected. There are too few points in the peak region and in the lefthand tail. The analysis was therefore repeated with 3 additional x_k^*. The results of this second analysis are shown in Figure 6 in which the original data are shown as small dots, and the best shape invariant predictions as continuous curves.

Sensitometric Curve Analysis: The $n = 16$ curves in Figure 3 are monotone decreasing curves and the normalization in (17) is used. In this case the curve minima are not all zero, and so the parameter θ_0 will not be omitted. The full 4-parameter shape invariant model (12) is used. The parameters are

$$\begin{aligned} \theta_{0i} &= \text{Fog level (density due to chemical fogging)} \\ \theta_{1i} &= \text{Density range of the film} \\ \theta_{2i} &= \text{A photographic speed measure} \\ \theta_{3i} &= \text{Slope at } \theta_{2i} \text{ (measure of contrast).} \end{aligned} \tag{26}$$

The analysis required 4 iterations, about 20 seconds of computer time. The error was on the order of instrument measurement error.

5. Resolution Of Shape Invariant Models

Although the shape invariant model has broad application, there are important problems in which one observes the *sum* of two or more shape invariant models. Such problems arise in spectrophotometry, chemical kinetics, and chromatography. Observed curves from such processes have the form

$$y_i(x) = f(\boldsymbol{\theta}_i\,;\, x) + \text{error} \tag{27}$$

with

$$f(\boldsymbol{\theta}_i\,;\, x) = \theta_{0i} + \theta_{1i}g([x - \theta_{2i}]/\theta_{3i}) + \theta_{4i}g([x - \theta_{5i}]/\theta_{6i}) + \cdots \tag{28}$$

where $g(x)$ is the common shape function. Once again our goal is to find the best shape function (bandshape) $\hat{g}(x)$, and estimate the associated least squares vector $\boldsymbol{\theta}_i$.

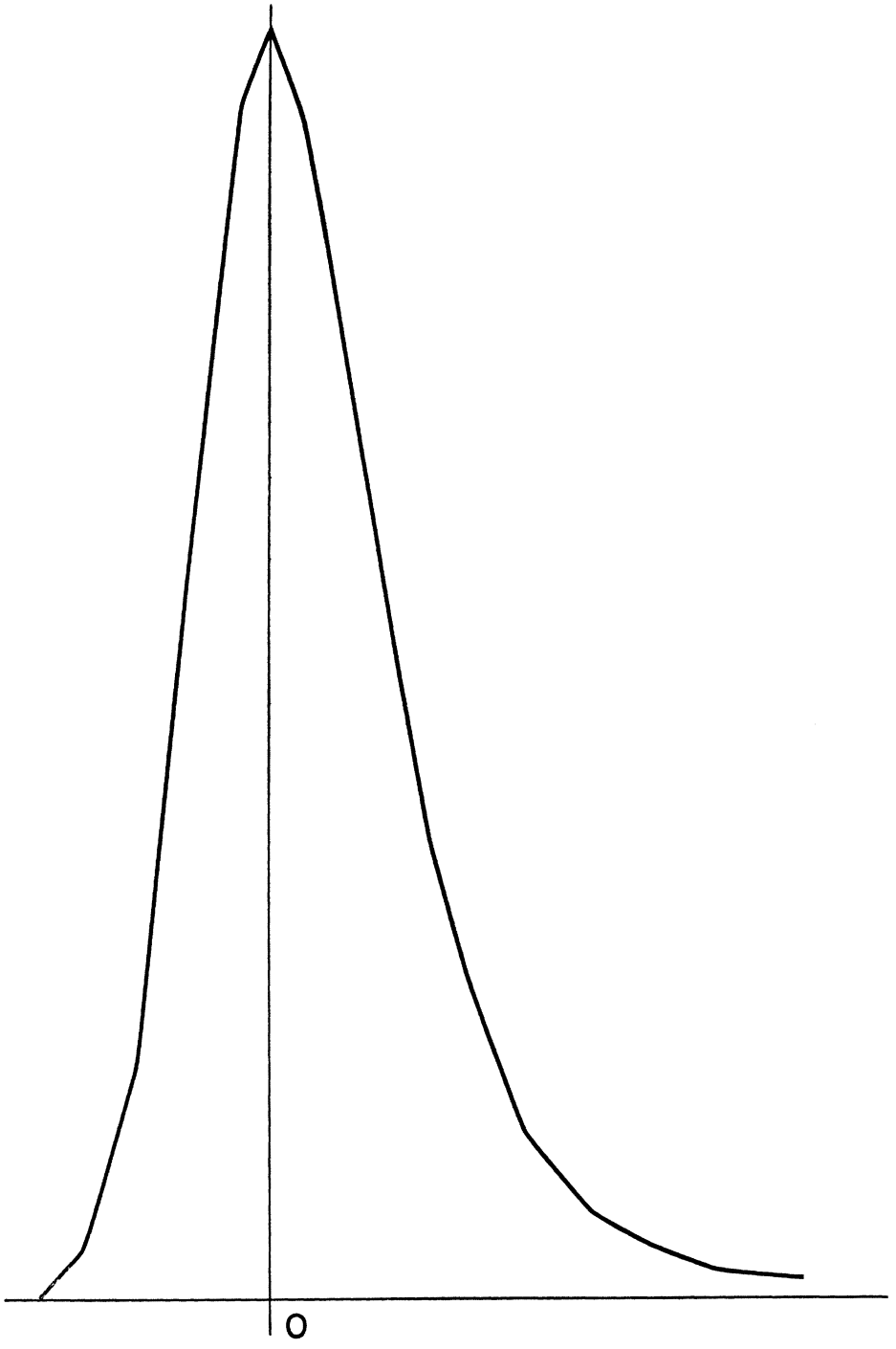

FIGURE 5

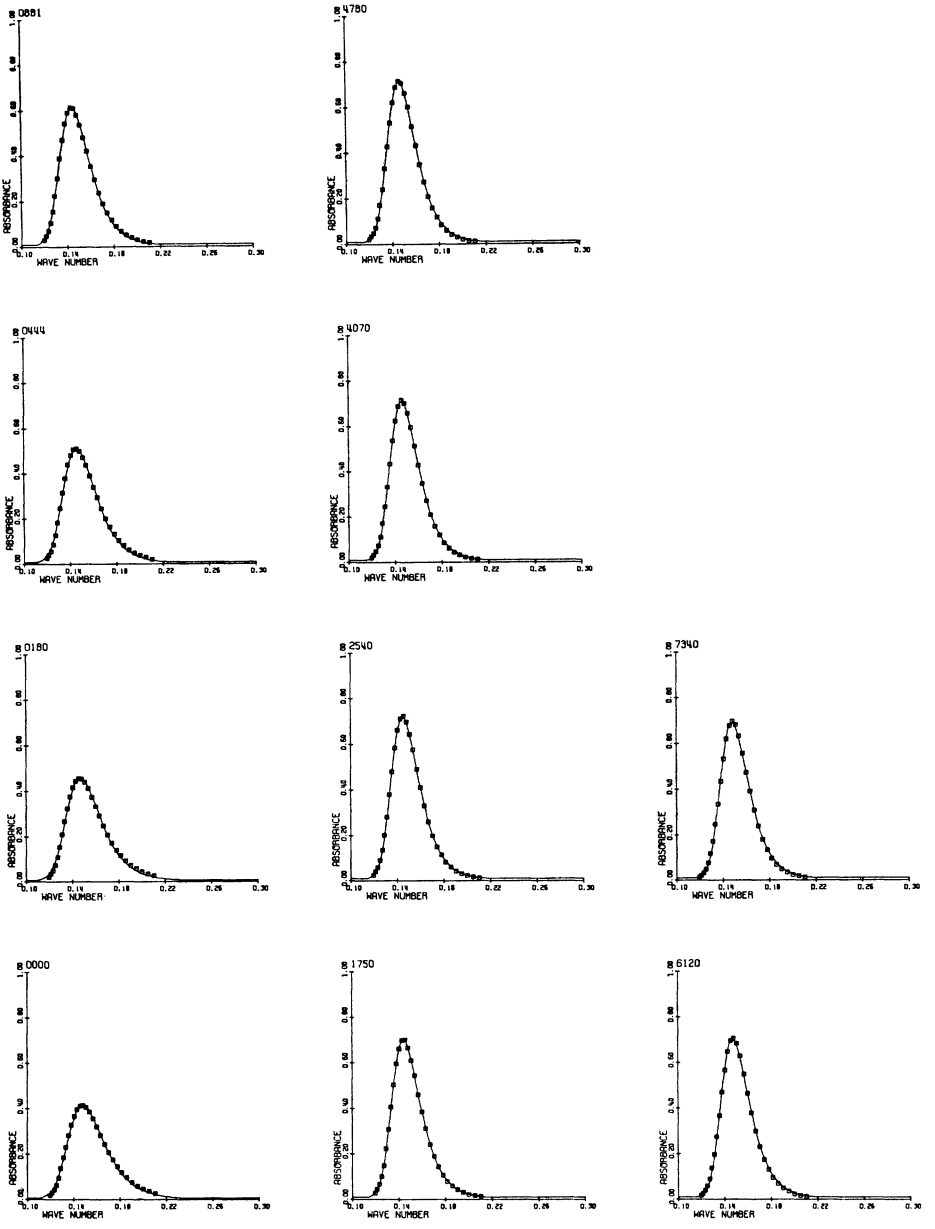

FIGURE 6

For the present we shall restrict ourself to the two component case

$$f(\theta_i\ ;\ x) = \theta_{0i} + \theta_{1i}g([x - \theta_{2i}]/\theta_{3i}) + \theta_{4i}g([x - \theta_{5i}]/\theta_{6i}). \tag{29}$$

We are concerned then, with the familiar curve resolution problem. The usual approach [10] has been to assume that $g(x)$ has some known form (Gaussian, Lorentzian, etc.), and then solve for the unknown parameters by nonlinear

regression. Another possible approach is given in [7]. The latter approach provides estimates of $g(x)$, θ_{1i} , and θ_{4i} with no assumptions about the form of $g(x)$. It does, however, require that the parameters θ_{2i} , θ_{3i} , θ_{5i} , and θ_{6i} be constant for all i.

In the present approach we seek the empirical function $g^* \in G^*$ which minimizes (14) with $f(\boldsymbol{\theta}_i ; x)$ defined as in (29). The iterative technique of Section 6 can again be used to obtain $\hat{g}^*$ and $\hat{\boldsymbol{\theta}}_i(\hat{g}^*)$. The only difference being that we must now perform n 7-parameter nonlinear regressions rather than n 4-parameter nonlinear regressions on each iteration.

Spectrophotometric Curve Resolution: The $n = 12$ spectrophotometric curves shown in Figure 7 show considerable lack of fit when modeled with the simple shape invariant model of (2). There were clear indications that these 12 curves were varying mixtures of two absorptance bands. We next assumed that these curves followed the model in (29) with a Gaussian bandshape. The curves were then resolved using straightforward nonlinear regression with the algorithm in [6]. The resulting resolution still exhibited clear systematic errors indicating

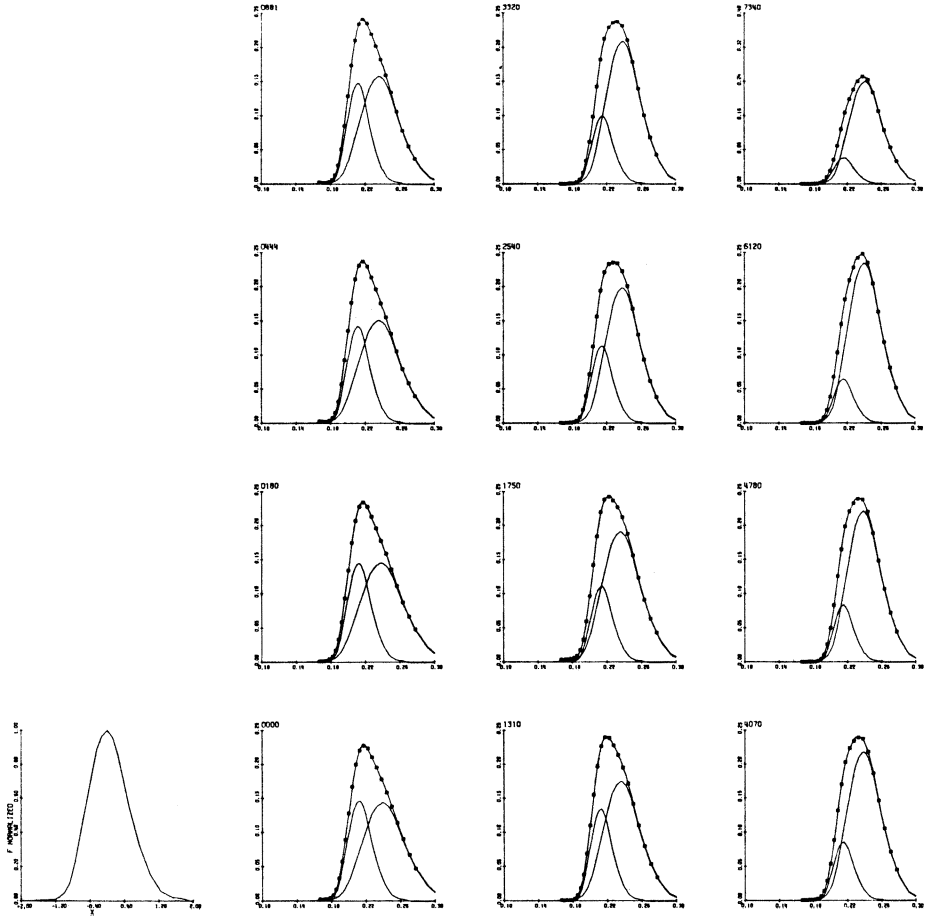

FIGURE 7

that the Gaussian bandshape was not correct. The parameter estimates from the Gaussian resolution were fed into the Self Modeling algorithm and the best bandshape determined. Three iterations were required and produced the analysis shown in Figure 7.

The best bandshape is shown in the lower lefthand plot of Figure 7. This band is clearly non-Gaussian and exhibits a skew toward the right. The raw data are plotted as points, with the resolved components and their sum shown as continuous curves. As you may note there is no observable lack of fit. The standard errors were reduced an order of magnitude over the Gaussian model. More importantly, subsequent data has been analyzed using the bandshape in Figure 7 and the standard errors continue to be an order of magnitude better than the Gaussian model.

6. Iterative Procedure

This section develops the iterative procedure used in the previous sections. The procedure will be described in terms of the simple shape invariant model only. The approach is, however, applicable to the general model

$$f(\alpha, \beta, \gamma; x) = \sum_i \alpha_i m_i(x) + \sum_i \beta_i g(r(\gamma; x)) \tag{30}$$

where the $m_i(x)$ are known functions, and $r(\gamma; x)$ is a known transformation of the x-axis. In our shape invariant model $m_1(x) \equiv 1$ and $r(\gamma; x) = (x - \gamma_1)/\gamma_2$.

For the basic shape invariant model, it was determined in Section 3 that one seeks to minimize (19) subject to $\mathbf{g}^* \in \mathbf{G}^*$. This is equivalent to seeking the $\mathbf{g}^* \in \mathbf{G}^*$ and $\boldsymbol{\theta}_i$ which minimize

$$Q(\mathbf{g}^*, \boldsymbol{\theta}_i) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p_i} [y_i(x_{ij}) - h(\mathbf{g}^*, \boldsymbol{\theta}_i ; x_{ij})]^2 \tag{31}$$

where

$$h(\mathbf{g}^*, \boldsymbol{\theta}_i ; x) = \begin{cases} \theta_{0i} + \theta_{1i} \sum_{k=1}^{m-1} \left(\frac{g_{k+1} - g_k}{x^*_{k+1} - x^*_k} [z_i - x^*_k] + g_k \right) \Delta_k(z_i), & x^*_1 < z_i \leq x^*_m \\ \theta_{0i} + \theta_{1i} g_1 , & z_i \leq x^*_1 \\ \theta_{0i} + \theta_{1i} g_m , & z_i > x^*_m \end{cases} \tag{32}$$

and $z_i = (x - \theta_{2i})/\theta_{3i}$.

This problem can be considered to be a nonlinear regression problem with the model.

$$y_{ij} = \sum_{l=1}^{n} h(\mathbf{g}^*, \boldsymbol{\theta}_l ; x_{lj}) \, \delta_{li} \tag{33}$$

where y_{ij} is $y_i(x_{ij})$ and $\delta_{li} = 1$ if $l = i$, and 0 otherwise.

The model in (33) has a total of $4n + m$ parameters; $\mathbf{g}^*$ accounts for m parameters, and each $\boldsymbol{\theta}_i$, $i = 1, 2, \cdots, n$ accounts for 4 parameters. In the sensitometric curve example of Section 4, $4n + m = 80$. Nonlinear regression analyses

with such a large number of parameters are invariably *ill-conditioned* and unstable. The chances of solving this problem by treating it as a $4n + m$ parameter nonlinear regression would be nil.

At this point the techniques for the elimination of linear parameters discussed in [6] can be applied to advantage. Since the parameters $g_1, g_2, \cdots, g_m$ appear linearly in (33), one can work with the $4n$-parameter *reduced model*

$$y_{ij} = \sum_{l=1}^{n} h(\hat{g}^*(\theta), \theta_l ; x_{lj}) \delta_{li} \tag{34}$$

where $\hat{\mathbf{g}}^*(\theta)$ are the best *companion values* of $\mathbf{g}^*$ given $\theta = (\theta_1, \theta_2, \cdots, \theta_n) = (\theta_{01}, \theta_{11}, \cdots, \theta_{3n})$. The $\hat{\mathbf{g}}^*(\theta)$ are obtained by linear regression using the model in (33) as described in [6]. (34) can be written in the form

$$y_{ij} = \sum_{k=1}^{n} g_k w_{ijk} \tag{35}$$

where the w_{ijk} are known predictor variables once θ is given. The best companion values are then simply $\hat{\mathbf{g}}^*(\theta) = \hat{g}_k$.

It is important to note that the matrix of predictors in (35) may have whole columns of zeros; that is, $w_{ijk} = 0$ for all i and j. This problem occurs when the x_k^* are overspecified.

Let $z_{ij} = (x_{ij} - \theta_{2i})/\theta_{3i}$. If there is some x_k^* such that $z_{ij} \notin (x_{k-1}^*, x_{k+1}^*]$ for all i and j, then $w_{ijk} = 0$ for all i and j. One cannot then determine the coefficient g_k since the original data "provide no information about $\hat{g}^*$ in the region of x_k^*". This problem is overcome by deleting those columns corresponding to the "unused" x_k^*, and estimating the empirical function defined by the *reduced* set of x_k^*; call this empirical function g^*_{reduced}. Now since the z_{ij} which determine whether or not an x_k^* is dropped change from iteration to iteration (because the θ_{2i}, θ_{3i} change), the x_k^* are dropped only for the purposes of performing the linear regression. After the regression analysis, one defines the empirical function $\hat{\mathbf{g}}^*(\theta)$ at all the original x_k^*. This can be done since g^*_{reduced} is defined for all x.

Having eliminated $\mathbf{g}^*$ one is now faced with the $4n$-parameter nonlinear regression in (34). That is, we seek the θ which minimizes

$$Q(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{p_i} \left[y_{ij} - \sum_{l=1}^{n} h(\hat{\mathbf{g}}^*(\theta), \theta_l ; x_{lj}) \delta_{li} \right]^2. \tag{36}$$

The Gauss–Newton approach to this problem replaces the function in (34) by the first order terms of its Taylor expansion in θ about the starting guess $\theta^{(0)}$. This approximation amounts to replacing h in (34) by the approximation

$$h(\mathbf{g}^*(\theta), \theta_l ; x) \cong h(\mathbf{g}^*(\theta^{(0)}), \theta_l^{(0)}; x) + \sum_{l=1}^{n} \sum_{k=0}^{3} \frac{\partial h}{\partial \theta_{kl}} \bigg|_{\theta^{(0)}} (\theta_{kl} - \theta_{kl}^{(0)}), \tag{37}$$

for θ near $\theta^{(0)}$. Using this approximation an "improved" vector $\theta^{(1)}$ is determined. The expansion in (37) is then taken about the improved vector $\theta^{(1)}$, and so forth. If all goes well, $\theta^{(k)} \to \hat{\theta}(\hat{g}^*)$. However, θ is still a $4n$ dimensional vector (typically $4n = 60$), and, as noted earlier, these nonlinear procedures are usually unstable with high dimension parameter vectors.

It is, however, possible to obtain the solution to (36) by using the approximation

$$h(\hat{\mathbf{g}}^*(\boldsymbol{\theta}),\ \boldsymbol{\theta}_l\ ;\ x) \cong h(\hat{\mathbf{g}}^*(\boldsymbol{\theta}^{(0)}),\ \boldsymbol{\theta}_l\ ;\ x) \tag{38}$$

for $\boldsymbol{\theta}$ near $\boldsymbol{\theta}^{(0)}$, rather than the approximation in (37). The approximation in (38) is based on the assumption that the best companion vector $\hat{\mathbf{g}}^*(\boldsymbol{\theta})$ does not change significantly for small changes in $\boldsymbol{\theta}$. As before in the Gauss–Newton approach, one uses the approximation to produce an improved value $\boldsymbol{\theta}^{(1)}$, then uses the expansion in (38) about $\boldsymbol{\theta}^{(1)}$ to produce the next "improved" value $\boldsymbol{\theta}^{(2)}$, etc. If one has reasonable starting values $\boldsymbol{\theta}^{(0)}$, it has been our experience that $\boldsymbol{\theta}^{(k)} \to \hat{\boldsymbol{\theta}}(\hat{g}^*)$.

The approximation in (38) was selected because it reduces the $4n$-parameter nonlinear regression in (36) to n 4-parameter nonlinear regressions. Substituting (38) into (36) one has

$$Q(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{p_i} \left[y_{ij} - \sum_{l=1}^{n} h(\hat{\mathbf{g}}^*(\boldsymbol{\theta}^{(0)}),\ \boldsymbol{\theta}_l\ ;\ x_{lj})\ \delta_{li} \right]^2 \tag{39}$$

which is of the form

$$Q(\boldsymbol{\theta}) = \sum_{i=1}^{n} s_i^2(\boldsymbol{\theta}_i) \tag{40}$$

where

$$s_i^2(\boldsymbol{\theta}_i) = \sum_{j=1}^{p_i} [y_{ij} - h(\hat{\mathbf{g}}^*(\boldsymbol{\theta}^{(0)}),\ \boldsymbol{\theta}_i\ ;\ x_{ij})]^2. \tag{41}$$

The improved estimate $\boldsymbol{\theta}^{(1)}$ is then the vector $\boldsymbol{\theta}$ which minimizes (39). Since $s_i^2(\boldsymbol{\theta}_i)$ depends solely on the parameters $\boldsymbol{\theta}_i = (\theta_{0i}\ ,\ \theta_{1i}\ ,\ \theta_{2i}\ ,\ \theta_{3i})$, each $s_i^2(\boldsymbol{\theta}_i)$ may be minimized independently of the others. It then follows that (39) is minimized by selecting $\boldsymbol{\theta}^{(1)} = (\hat{\boldsymbol{\theta}}_1\ ,\ \hat{\boldsymbol{\theta}}_2\ ,\ \cdots\ ,\ \hat{\boldsymbol{\theta}}_n)$ where $\hat{\boldsymbol{\theta}}_i$ is the vector which minimizes (41).

Notice that (41) may be written in the more familiar form

$$s_i^2(\boldsymbol{\theta}_i) = \sum_{j=1}^{p_i} [y_i(x_{ij}) - \theta_{0i} - \theta_{1i} g_0^*([x_{ij} - \theta_{2i}]/\theta_{3i})]^2 \tag{42}$$

where g_0^* is the empirical function defined by the vector $\hat{g}^*(\boldsymbol{\theta}^{(0)})$. It then follows from (42) that $\boldsymbol{\theta}_i$ can be obtained by a 4-parameter nonlinear regression on the i-th observed curve using the model

$$y_i(x) = \theta_{0i} + \theta_{1i} g_0^*([x - \theta_{2i}]/\theta_{3i}). \tag{43}$$

These 4-parameter nonlinear regressions can be further reduced to 2-parameter nonlinear regressions by elimination of the linear parameters θ_{0i} and θ_{1i} using the methods discussed in [6].

$\boldsymbol{\theta}^{(1)}$ now replaces $\boldsymbol{\theta}^{(0)}$ in (39) and the process is continued until, for some prespecified ϵ,

$$\max_x |\hat{g}^*(\boldsymbol{\theta}^{(k)}) - \hat{g}^*(\boldsymbol{\theta}^{(k+1)})| < \epsilon. \tag{44}$$

7. Summary

The authors have written a FORTRAN program based on this algorithm, and have used this program for a number of analyses. It would appear that the method is about as stable as nonlinear regression using a known shape function, yet it has the advantage of allowing the analysis to synthesize a best shape function from the observed data itself. Practice has shown that these best models are generally at least a factor of 2 better than models based on shape functions with simple algebraic forms. A more complete report on this technique is available from the authors.

It is possible to replace the empirical functions with cubic splines, and reduce the number of x_k^* needed for an analysis. The authors hope to publish a short note on this at a later date. The extension is relatively straightforward.

References

[1] Bellman, R. and Roth, R. (1969). Curve Fitting by Segmented Straight Lines. *J. Amer. Statis. Assoc. 64*, 1079–1084.

[2] Box, G. E. P. and Hunter, W. G. (1962) A Useful Method of Model Building. *Technometrics 4*, 301–318.

[3] Church, A. Jr. (1966). Analysis of Data when the Response is a Curve. *Technometrics 8*, 229–246.

[4] Draper and Smith. (1966). *Applied Regression Analysis*. John Wiley and Sons. New York.

[5] Hartley, H. O. (1961). The Modified Gauss-Newton Method in the Fitting of Nonlinear Regression Functions by Least Squares. *Technometrics 3*, 269–280.

[6] Lawton, W. H. and Sylvestre, E. A. (1971). Elimination of Linear Parameters in Nonlinear Regression. *Technometrics 15*, 461–467.

[7] Lawton, W. H. and Sylvestre, E. A. (1971). Self Modeling Curve Resolution. *Technometrics 15*, 617–633.

[8] Mandel, J. (1969). Partitioning of Interactions in Analysis of Variance. *J. Res. Nat. Bur. Stds. Section B, Math. Sciences.*

[9] Marquardt, D. W. (1963) An Algorithm for Least Squares Estimation of Nonlinear Parameters. *SIAM 2*, 431–441

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