Nonlinear Regression Applied to Non-Newtonian Flow

PAUL J. NIEBERGALL, ROGER L. SCHNAARE, and EDWIN T. SUGITA

Abstract □ Modifications of the Gauss–Newton method are among the most widely used methods for nonlinear regression analysis. One such modification, which the authors found applicable to a wide variety of pharmaceutical systems, is described. Its application in describing the flow behavior of non-Newtonian systems and the FORTRAN IV program utilized are presented. One of the more useful, accurate, and physically significant equations for describing non-Newtonian flow is the structure equation: $F = f + \eta_\infty S - b_v e^{-aS}$, in which F is shear stress, S is shear rate, and the other terms are constants. The equation was originally evaluated through the use of multiple regression, with the constant a assumed to be equal to 0.001, which gave good fit to a variety of flow systems. Since the original equation was developed, nonlinear regression techniques have appeared which make it possible to examine the structure equation in greater detail. It was found, for example, that for dispersions of a wax, consisting of a mixture of polyethoxylated higher fatty alcohols, in water the constant a varied from 0.013 to 0.049 rather than remaining fixed at 0.001. Some of the original data, upon which the structure equation was based, were reevaluated using nonlinear regression analysis. These data were for suspensions of salicylamide (varying concentration) in methylcellulose solutions of varying concentration. The constant a was found to vary from 0.00109 to 0.00172 as the concentration of methylcellulose increased and was independent of salicylamide concentration. In all instances, allowing a to vary as an adjustable parameter gave a better fit to the data than assuming it to be constant at 0.001. The use of nonlinear regression analysis served to emphasize the usefulness of the structure equation.

Keyphrases □ Regression analysis, nonlinear—non-Newtonian flow systems □ Non-Newtonian flow systems—application of nonlinear regression analysis □ NONLIN, subroutine—digital computer program □ Digital computer programs—subroutine NONLIN

Although nonlinear regression analysis techniques (1–3) have been available for a number of years, they have received little attention in the pharmaceutical sciences except for the area of pharmacokinetics (4–7). The techniques of nonlinear regression analysis, except for practical difficulties which may occasionally arise, should be capable of curve fitting any system that may be of interest to pharmaceutical scientists. The authors, for example, successfully applied these techniques to studies of protein binding, the determination of dissociation constants for polyprotic acids, complexation studies, Michaelis–Menten kinetics, and the investigation of rheological systems. One dramatic example of the application of nonlinear regression analysis was the analysis of one of the more useful, accurate, and physically significant equations for describing non-Newtonian flow. This equation, designated as the structure equation (8), is:

$$F = f + \eta_\infty S - b_v e^{-aS} \qquad \text{(Eq. 1)}$$

in which F is the shear stress, S is the shear rate, and the other terms are constants. At the time the equation was developed, the only manner in which it could be analyzed was to assume a value for the constant a and treat the equation as a multiple-regression problem in the form:

$$F = f + \eta_\infty S - b_v X \qquad \text{(Eq. 2)}$$

in which:

$$X = e^{-aS} \qquad \text{(Eq. 3)}$$

An arbitrary value of 0.001 was found to give very good fit to a wide variety of systems. Data from the authors' laboratories, as well as some of the data upon which the equation was based originally (9)[1], were analyzed using the techniques of nonlinear regression analysis, in which the constant a was allowed to be an adjustable parameter. The results and a discussion of the techniques utilized are presented in this article.

Table I—Constants of Structure Equation for Dispersions of a Mixture of Polyethoxylated Higher Fatty Alcohols[a] in Water at 25°

Percent Wax	f	η_∞	b_v	a
1	25.2	0.1040	19.55	0.013
2	41.7	0.3146	17.39	0.044
3	110.4	0.7101	45.68	0.044
4	232.7	1.0350	131.80	0.049

[a] Polawax (Croda).

THEORETICAL

To utilize the techniques of nonlinear regression, initial estimates, $P_1^0, P_2^0, P_i^0, \ldots, P_k^0$, are needed for the k adjustable parameters, and correction vectors defined such that:

$$P_1 = P_1^0 - \Delta P_1 \qquad \text{(Eq. 4a)}$$

$$P_2 = P_2^0 - \Delta P_2 \qquad \text{(Eq. 4b)}$$

$$\vdots$$

$$P_k = P_k^0 - \Delta P_k \qquad \text{(Eq. 4c)}$$

in which ΔP_i represents a correction vector that will enable P_i to be a "better" estimate of the true parameter than P_i^0. The function:

$$Y = f(X; P_1, P_2, P_i, \ldots, P_k) \qquad \text{(Eq. 5)}$$

in which Y is the dependent variable, X is the independent variable, and P_i represents the current estimate of the true parameter, is

[1] The authors are grateful to Dr. Wayne Grim for his permission to use the data from his doctoral thesis. All of the data on salicylamide in methylcellulose referred to in this publication were obtained from this thesis.

Table II—Nonlinear Regression Analysis on 10.29% v/v Salicylamide in 1.59% w/w Methylcellulose, 1500 cps., Solutions

Constant	Value	Standard Deviation	95% Confidence Limits
f	1.4082×10^{3}	3.058×10^{1}	$1.348 \times 10^{3} - 1.468 \times 10^{3}$
η_∞	6.3453×10^{-1}	1.298×10^{-2}	$6.091 \times 10^{-1} - 6.600 \times 10^{-1}$
b_v	1.2887×10^{3}	2.645×10^{1}	$1.237 \times 10^{3} - 1.340 \times 10^{3}$
a	1.8900×10^{-3}	7.037×10^{-5}	$1.752 \times 10^{-3} - 2.028 \times 10^{-3}$

$Y \times 10^{-3}$	Y Calculated $\times 10^{-3}$	Percent Difference
0.2587	0.2846	−10.003
0.3317	0.3414	−2.926
0.4709	0.4695	0.296
0.6235	0.6014	3.552
0.7628	0.7377	3.285
0.8756	0.8610	1.660
1.008	1.006	0.213
1.154	1.142	1.009
1.267	1.273	−0.471
1.393	1.403	−0.700
1.512	1.528	−1.009
1.625	1.631	−0.393
1.711	1.725	−0.809
1.784	1.804	−1.088
1.857	1.872	−0.786
1.930	1.937	−0.350
1.990	1.994	−0.187
2.050	2.047	0.117
2.109	2.097	0.586
2.149	2.147	0.077
2.202	2.196	0.300
2.249	2.243	0.270
2.295	2.285	0.415
2.335	2.327	0.348
2.388	2.374	0.587
2.428	2.425	0.115
2.467	2.466	0.068
2.521	2.515	0.209
2.560	2.561	−0.039
2.613	2.614	−0.020
2.673	2.669	0.139
2.733	2.727	0.220
2.773	2.778	−0.180
2.826	2.829	−0.112
2.865	2.867	−0.054
2.912	2.907	0.178
2.938	2.942	−0.122
2.958	2.975	−0.553

expanded in a Taylor series retaining only the first-order terms to give:

$$Y = \hat{Y} + \Delta P_1 F_1 + \Delta P_2 F_2 + \Delta P_i F_i + \ldots + \Delta P_k F_k \quad \text{(Eq. 6)}$$

in which:

$$F_i = \partial Y / \partial P_i \quad \text{(Eq. 7)}$$

and $\hat{Y}$ represents the theoretical value of Y calculated from the current set of estimated parameters.

Equation 6 is linear in the correction vectors and can be treated using the techniques of multiple regression in which Eq. 6 can be represented by:

$$Y = a_1 X_1 + a_2 X_2 + a_i X_i + \ldots + a_k X_k + \epsilon \quad \text{(Eq. 8)}$$

in which ϵ represents the residual for a given point. Taking the partial derivative of ϵ^2 with respect to each ΔP_i, setting each resulting equation equal to zero to minimize the sum of squares of residuals, and summing over the entire set of experimental points yield a set of "normal equations":

$$(\Sigma F_1 F_1)\Delta P_1 + (\Sigma F_1 F_2)\Delta P_2 + (\Sigma F_1 F_i)\Delta P_i + \ldots + (\Sigma F_1 F_k)\Delta P_k = \Sigma F_1 F \quad \text{(Eq. 9a)}$$

$$(\Sigma F_2 F_1)\Delta P_1 + (\Sigma F_2 F_2)\Delta P_2 + (\Sigma F_2 F_i)\Delta P_i + \ldots + (\Sigma F_2 F_k)\Delta P_k = \Sigma F_2 F \quad \text{(Eq. 9b)}$$

$$(\Sigma F_i F_1)\Delta P_1 + (\Sigma F_i F_2)\Delta P_2 + (\Sigma F_i F_i)\Delta P_i + \ldots + (\Sigma F_i F_k)\Delta P_k = \Sigma F_i F \quad \text{(Eq. 9c)}$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$(\Sigma F_k F_1)\Delta P_1 + (\Sigma F_k F_2)\Delta P_2 + (\Sigma F_k F_i)\Delta P_i + \ldots + (\Sigma F_k F_k)\Delta P_k = \Sigma F_k F \quad \text{(Eq. 9d)}$$

in which:

$$F = Y - \hat{Y} \quad \text{(Eq. 10)}$$

and the subscripted F_i's refer to the partial derivatives as given by Eq. 7. The initial estimates, $P_1^0, P_2^0, P_i^0, \ldots, P_k^0$, are used to solve for $F, F_1, F_2, F_i, \ldots, F_k$ at each data point: the sums of the products are inserted into the set of normal equations (Eqs. 9). This set of simultaneous equations can be solved in a wide variety of ways for $\Delta P_1, \Delta P_2, \Delta P_i, \ldots, \Delta P_k$. The Crout reduction method (10) is used here. These correction vectors are then substituted into Eqs. 4 to

Table III—Reduction in Sums of Squares of Residuals between Values Obtained Using Multiple Regression with $a = 0.001$, and Nonlinear Regression in Which a Was Considered an Adjustable Parameter. Data for Varying Concentrations of Salicylamide in 1.59 % w/w Methylcellulose, 1500 cps., Solutions

Percent Salicylamide, v/v	Sums of Squares of Residuals $\times 10^{-4}$ — Multiple Regression	Nonlinear Regression
0.00	1.00	0.26
5.40	3.65	0.37
10.29	4.57	0.45
14.59	4.58	0.52
19.69	4.26	1.00
24.21	6.49	0.93
28.07	11.91	2.91
34.39	24.03	5.41

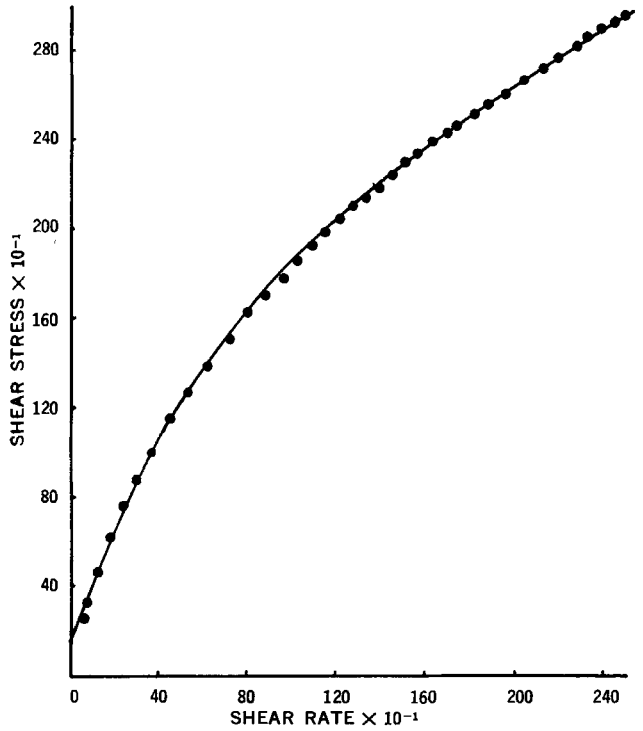


Figure 1—*Comparison between calculated and experimental values of shear stress. The solid line represents the calculated values of shear stress, while the closed circles represent the actual experimental points. The system represented is 10.29% v/v salicylamide in 1.59% w/w methylcellulose, 1500 cps.*

give new estimates of the parameters P_1, P_2, P_i, ..., P_k, which together with the experimental values of Y and X are used to solve for F, F_1, F_2, F_i, ..., F_k for each data point. The sums of squares are again tabulated, and the set of Eqs. 9 is solved for the correction vectors. This iterative procedure is continued until the desired degree of accuracy is obtained.

Although the set of Eqs. 4 can be used successfully for fitting many functions, convergence may not be obtained in others. An alternate procedure, which converges more frequently and which is used in this computer program, is as follows.

1. After each titration, correct the estimated parameters:

$$P_1 = P_1^0 - R\,\Delta P_1 \quad \text{(Eq. 11}a\text{)}$$

$$P_2 = P_2^0 - R\,\Delta P_2 \quad \text{(Eq. 11}b\text{)}$$

$$P_i = P_i^0 - R\,\Delta P_i \quad \text{(Eq. 11}c\text{)}$$

$$\cdots \quad \cdots \quad \cdots$$

$$P_k = P_k^0 - R\,\Delta P_k \quad \text{(Eq. 11}d\text{)}$$

using values of $R = 0$, 0.5, and 1 on each of the parameters.

2. Calculate the sum of squares of residuals (ΣF^2) using the three estimates of P_1, P_2, P_i, ..., P_k, designating S_0 as the result for $R = 0$, $S_{0.5}$ for the set in which $R = 0.5$, and S_1 as the result for $R = 1$.

3. Calculate $R_{\min.}$:

$$R_{\min.} = 0.5 + \frac{(S_0 - S_1)}{4(S_0 - 2S_{0.5} + S_1)} \quad \text{(Eq. 12)}$$

4. Now correct each parameter:

$$P_1 = P_1^0 - R_{\min.}\,\Delta P_1 \quad \text{(Eq. 13}a\text{)}$$

$$P_2 = P_2^0 - R_{\min.}\,\Delta P_2 \quad \text{(Eq. 13}b\text{)}$$

$$P_i = P_i^0 - R_{\min.}\,\Delta P_i \quad \text{(Eq. 13}c\text{)}$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots$$

$$P_k = P_k^0 - R_{\min.}\,\Delta P_k \quad \text{(Eq. 13}d\text{)}$$

Table IV—Differences between Constants Obtained *via* Multiple Regression with $a = 0.001$, and Nonlinear Regression in Which a Was Considered an Adjustable Parameter. Data for Varying Concentrations (v/v%) of Salicylamide in 1.59 % w/w Methylcellulose, 1500 cps., Solutions

Percent Salicylamide	f	η_∞	b_v	$a \times 10^3$	Yield Value ($f - b_v$)
0.00%:					
Literature[a]	1634.0	0.2800	1507.0	1.00	127.0
This work	1086.0	0.4638	1021.0	1.72	65.0
Percent difference	33.5	65.6	32.2	72.0	48.8
5.40%:					
Literature	1981.0	0.2965	1806.0	1.00	175.0
This work	1356.0	0.5046	1248.0	1.65	18.0
Percent difference	31.5	70.2	30.9	65.0	38.3
10.29%:					
Literature	2205.0	0.3624	1985.0	1.00	220.0
This work	1408.0	0.6345	1289.0	1.89	119.0
Percent difference	36.1	75.1	35.1	89.0	45.9
14.59%:					
Literature	2766.0	0.3611	2509.0	1.00	257.0
This work	1736.0	0.7144	1611.0	1.93	125.0
Percent difference	37.2	97.8	35.8	93.0	33.2
19.69%:					
Literature	3511.0	0.3937	3199.0	1.00	312.0
This work	2416.0	0.7579	2220.0	1.64	196.0
Percent difference	31.2	92.5	30.6	64.0	37.2
24.21%:					
Literature	3773.0	0.5492	3415.0	1.00	358.0
This work	2477.0	0.9884	2271.0	1.79	206.0
Percent difference	34.3	80.0	33.5	79.0	42.5
28.07%:					
Literature	5021.0	0.7281	4528.0	1.00	493.0
This work	3333.0	1.2976	3036.0	1.76	297.0
Percent difference	33.6	78.2	33.0	76.0	39.8
34.39%:					
Literature	7454.0	0.9837	6754.0	1.00	700.0
This work	4969.0	1.8206	4552.0	1.74	417.0
Percent difference	33.3	85.1	32.6	74.0	40.4

[a] Reference 9.

Table V—Listing for the Main Program

```
      IMPLICIT REAL*8 (A-H,O-Z)
      INTEGER NAME(60)
      COMMON/SET1/P,NAME
      COMMON/SET2/Y(100),YCALC(100),X(100),F(10)
      DIMENSION P(10)
    1 READ(5,10)NAME
   10 FORMAT(60A1)
      READ,N,NOCON,TS
      READ,(P(I),I-1,NOCON)
      READ,(X(I),I=1,N)
      READ,(Y(I),I=N)
      NO = N
      WRITE(6,15)NAME
   15 FORMAT('1',T3,60A1,///)
      CALL NONLIN(NO,NOCON,TS)
      READ,MORE
      GO TO(1,20),MORE
   20 STOP
      END
```

5. Continue the iteration procedure as before.

If the right side of the set of Eqs. 9 is replaced by the unit matrix (one in which all of the diagonal elements are equal to 1 and all other elements are equal to zero), the covariance matrix can be obtained by solving Eqs. 9 using each column of the unit matrix in succession. The diagonal elements of the covariance matrix, designated as V_{11}, V_{22}, V_{tt}, . . . , V_{kk}, are used to obtain estimates of the standard deviation for each parameter as follows:

$$\text{VAREXT} = S_0/(N - \text{NOCON}) \qquad \text{(Eq. 14)}$$

in which VAREXT is an external estimate of the unit variance, N is the number of data points, and NOCON is the number of parameters being fitted. The standard deviation for each parameter is defined as follows:

$$\text{STNDV}_1 = (V_{11} \cdot \text{VAREXT})^{1/2} \qquad \text{(Eq. 15}a\text{)}$$

$$\text{STNDV}_2 = (V_{22} \cdot \text{VAREXT})^{1/2} \qquad \text{(Eq. 15}b\text{)}$$

$$\text{STNDV}_i = (V_{ii} \cdot \text{VAREXT})^{1/2} \qquad \text{(Eq. 15}c\text{)}$$

It is necessary at times to evaluate a constant, for example G, which is a function of the adjustable parameters of a particular equation:

$$G = f(P_1, P_i, \ldots, P_n) \qquad \text{(Eq. 16)}$$

in which n may be equal to or less than k, the total number of adjustable parameters of a particular equation. The variance, STNDV_g^2, for G can be obtained from:

$$\text{STNDV}_g^2 = \text{VAREXT} \sum_{i=1}^{n} V_{ii}F_i^2 + 2\text{VAREXT} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \times V_{ij}F_iF_j \qquad \text{(Eq. 17)}$$

in which F_i represents the partial derivative of G with respect to P_i.

An estimate of the variance for each calculated value of the dependent variable, $\hat{Y}$, as calculated from the adjusted parameters can be obtained from:

$$\sigma_{\hat{Y}}^2 = \text{VAREXT}\left(\sum_{i=1}^{k} V_{ii}F_i^2 + 2\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} V_{ij}F_iF_j\right) \qquad \text{(Eq. 18)}$$

and the 95% confidence interval for $\hat{Y}$ would be:

$$\text{limits} = \hat{Y} \pm t\sigma_{\hat{Y}} \qquad \text{(Eq. 19)}$$

ANALYSIS OF THE STRUCTURE EQUATION

To use the computer program (*Appendix*), the structure equation, Eq. 1, has to be converted into terms of $\hat{Y}$, X, and P_4:

$$\hat{Y} = P_1 + P_2X - P_3e^{-P_4X} \qquad \text{(Eq. 20)}$$

The appropriate partial derivatives, as described by Eq. 7, are:

$$F_1 = \partial\hat{Y}/\partial P_1 = 1.0 \qquad \text{(Eq. 21}a\text{)}$$

$$F_2 = \partial\hat{Y}/\partial P_2 = X \qquad \text{(Eq. 21}b\text{)}$$

$$F_3 = \partial\hat{Y}/\partial P_3 = -e^{-P_4X} \qquad \text{(Eq. 21}c\text{)}$$

$$F_4 = \partial\hat{Y}/\partial P_4 = XP_3e^{-P_4X} \qquad \text{(Eq. 21}d\text{)}$$

The conversion of Eqs. 20 and 21 into the FORTRAN IV language used in the computer program is described in the *Appendix*.

According to Eq. 1, when the shear rate approaches zero:

$$F = f - b_v = \text{yield value} \qquad \text{(Eq. 22)}$$

The estimated variance for the yield value, which is obtained from parameters P_1 and P_3 of Eq. 20, can be obtained using Eq. 17 to give:

$$\text{variance for yield value} = \text{VAREXT}(V_{11}F_1^2 + V_{22}F_2^2) + 2\text{VAREXT}(V_{12}F_1F_2) \qquad \text{(Eq. 23)}$$

RESULTS AND CONCLUSIONS

A series of dispersions, containing varying amounts of a mixture of polyethoxylated higher fatty alcohols[2], in water at 25° was analyzed using an Epprecht-Rheomat 15. Initial estimates of P_1 and P_2 in Eq. 20 were obtained from the approximately linear portion of the plot of shear stress *versus* shear rate. The standard "feathering" technique was used to obtain initial estimates of P_3 and P_4. The data were analyzed using the computer program given in the *Appendix* on an IBM 360-75 system using the WATFOR compiler. The results are shown in Table I. The initial estimated parameters differed from the final adjusted parameters by 30–50% in all instances. Since the values of the constant a of Eq. 1 were 13–49 times greater than the value of 0.001, the data of Grim (9) were reanalyzed in an effort to determine whether the wax dispersions showed abnormal behavior or whether the nonlinear regression analysis was indeed indicating that the constant a was capable of being treated as an adjustable parameter.

The results for a typical set of data are shown in Table II, which was adapted from the computer printout. The same set of data is shown in Fig. 1 in an effort to demonstrate the close agreement between the theoretical line drawn from the computer-adjusted constants and the experimental data, In a least-squares regression analysis, the line giving the best fit to a set of experimental data is that line for which the sum of squares of residuals, SS, is a minimum in which:

$$SS = \Sigma(Y - \hat{Y})^2 \qquad \text{(Eq. 24)}$$

and $\hat{Y}$ represents the theoretical value of Y, calculated from the adjusted parameters obtained in the least-squares analysis, and the experimental values of X. The structure equation, Eq. 1, was used to calculate the SS for the set of data represented in Table II, using the constants obtained by multiple regression (9) and those obtained by nonlinear regression analysis, respectively. It was found that the SS obtained using nonlinear regression analysis was 32.6% lower than the SS obtained using multiple regression and assuming a equal to 0.001.

Table III shows the marked reduction in SS for a series containing varying amounts of salicylamide suspended in 1.59% w/w methylcellulose, 1500 cps. It should be noted that both sets of constants, that obtained assuming a was equal to 0.001 and that obtained with nonlinear regression analysis, gave good fit to the expeimental data. The nonlinear regression analysis simply expands the inherent usefulness of the structure equation by now allowing it to be analyzed in such a manner as to take full advantage of its ability to fit non-Newtonian flow data. The differences obtained between the constants obtained using both techniques for the set of data shown in Table III are shown in Table IV. Although the values of the constants changed markedly in some instances, the overall conclusions initially drawn (9) as to effects of concentration, *etc.*, on the constants are fully substantiated, with an even greater degree of con-

[2] Polawax (Croda).

Table VI—Necessary Input Data; This Program Uses Unformatted Input

Input Record	Reference Line Number	Input Variable
1	6	NAME = any 60 characters for identification of the problem
2	8	N = number of data points NOCON = number of constants to be fitted TS = Student's t for N − NOCON degrees of freedom
3	9	P(I) = initial estimates of constants $P_1 \ldots P_{NOCON}$
4	10	X(I) = independent variable. Enter $X_1, X_2 \ldots, X_N$
5	11	Y(I) = dependent variable. Enter $Y_1, Y_2 \ldots, Y_N$
6	16	MORE = code for running more than one set of data. Enter 1 if you wish to run another set, and the number 2 if you wish to stop.

Table VII—Input Program in Which Raw Data Are Transformed into the Master Variables for Curve Fitting and in Which the Initial Estimates of the Constants Are Taken from the Raw Data

```
1                    IMPLICIT REAL*8 (A-H,O-Z)
2                    INTEGER NAME(60)
3                    COMMON/SET1/P,NAME
4                    COMMON/SET2/Y(100),YCALC(100),X(100),F(10)
5                    DIMENSION P(10)
6           1        READ(5,10)NAME
7           10       FORMAT(60A1)
8                    READ,N,NOCON,TS
9                    READ,TAU,ATAU,P(4)
10                   READ,(X(I),I=1,N)
11                   READ,(Y(I),I=1,N)
12                   DO 20 I = 1,N
13                   X(I) = ATAU*X(I)
14                   Y(I) = TAU*Y(I)
15          20       CONTINUE
16                   P(2) = (Y(N)-Y(N-1))/(X(N)-X(N-1))
17                   P(1) = Y(N)-P(2)*X(N)
18                   P(3) = P(1) − (Y(1)-(Y(2)-Y(1))/(X(2)-X(1))/(X(2)-X(1))*X(1))
19                   WRITE(6,30)NAME
20          30       FORMAT('1',T3,60A1,///)
21                   WRITE(6,40)
22          40       FORMAT(5X,'SHEAR STRESS',5X,'SHEAR RATE',/)
23                   DO 50 I = 1,N
24                   WRITE(6,45)Y(I),X(I)
25          45       FORMAT(5X,1PD10.3,12X,D10.3)
30          50       CONTINUE
31                   WRITE(6,60)
32          60       FORMAT(//5X,'INITIAL ESTIMATES OF CON. FOR STRUCTURE EQUATION'/)
33                   WRITE(6,70)(P(I),I=1,NOCON)
34          70       FORMAT(5X,'F=',1PD12.5,2X,'ETA INF. =',D12.5,2X,'BV=',D12.5,
35                   *2X,'A=',D14.7,//)
36                   NO = N
37                   CALL NONLIN(N,NOCON,TS)
38                   READ,MORE
39                   GO TO(1,80),MORE
40          80       STOP
41                   END
```

fidence using the nonlinear regression estimates for the parameters of the structure equation.

In summary, the structure equation was found to give excellent fit to a variety of systems, and a nonlinear regression analysis of the equation showed that the constant a of Eq. 1 is an adjustable parameter. In systems containing varying amounts of salicylamide suspended in solutions of varying concentrations of methylcellulose, the constant a varied from 0.00109 to 0.00172 as the w/w% of methylcellulose increased, but it seemed to be relatively insensitive to changes in salicylamide concentration. The full physical significance of the constant a cannot be ascertained at this time, due to the relatively small amount of data analyzed. Studies are in progress to determine the rheological significance and potential usefulness of this parameter.

APPENDIX

The double-precision FORTRAN IV program for nonlinear regression analysis developed by the authors has proven to be easy to use, readily adaptable to a number of digital computer systems (IBM 360-75, Sigma 7, and Burroughs 5500), and rather inexpensive. Although the cost of running the program will vary from system to system, the IBM 360-75 system[3] will run five sets of data, each with 30–50 experimental points, for a total cost of about $0.55 if the WATFOR compiler is used. The exact same material will cost approximately $0.85 using the G level standard compiler. The entire nonlinear regression analysis program is written in three sections. The only section that needs revision from problem to problem is the third section, a subroutine called FUNC.

Main Program—This program is used for entering all of the required data. The basic routine is shown in Table V. The line numbers on the extreme left of the listing are not part of the program *per se* but are included as reference points into the program. The necessary input data are described in Table VI. This basic program can be readily modified to manipulate input data into master variables for the curve-fitting procedure; it may be used to obtain initial estimates for some or all of the parameters to be fitted; and it may be modified to include additional printout of information.

The modification statements would normally be placed between lines 12 and 13 of the basic routine. One such modification that the authors currently are using for evaluation of the structure equation

[3] At the University of Pennsylvania.

Table VIII—Listing of Subroutine NONLIN

```
1.000          SUBROUTINE NONLIN(NO,NOCON,TS)
2.000          IMPLICIT REAL*8(A-H,O-Z)
3.000          INTEGER NAME (60)
4.000          DIMENSION P(10),SP(10),STNDV(10),DEL(10),A(10,21),
5.000         *B(10,21),V(10,10),SB(10,21),PRCNT(100),SYCALC(100),
6.000         *PONE(10),PHALF(10),HILIM(10),HILIMY(100)
7.000          REAL*8 LOLIM(10),LOLIMY(100)
8.000          COMMON/SET1/P,NAME
9.000          COMMON/SET2/Y(100),YCALC(100),X(100),F(10)
10.000         SIGMA = 0.0
11.000    5    NUM = (2*NOCON)+ 1
12.000         N = NOCON + 1
13.000         L = NOCON + 2
14.000    C INITIALIZE ARRAYS
15.000         DO 10 I = 1,NOCON
16.000         DO 10 J = 1,NUM
17.000         A(I,J) = 0.0
18.000    10   CONTINUE
19.000         DO 30 I = 1,NOCON
20.000         J = I + N
21.000         A(I,J) = 1.0
22.000    30   CONTINUE
23.000    C CALCULATE VALUES FOR THE "A" ARRAY
24.000         DO 35 I = 1,NO
25.000         CALL FUNC(I,P,N)
26.000         DO 35 K = 1,N
27.000         DO 35 J = 1,N
28.000         A(K,J) = A(K,J) + F(K)*F(J)
29.000    35   CONTINUE
30.000    C CALCULATE "B" MATRIX
31.000         SUM = 0.0
32.000         DO 85 I = 1,NOCON
33.000         DO 85 J = 1,NUM
34.000         IF(I - J) 60,40,40
35.000    40   L = J - 1
36.000         IF(L)65,55,65
37.000    55   B(I,J) = A(I,J) - SUM
38.000         GO to 80
39.000    60   L = I - 1
40.000         IF(L)65,75,65
41.000    65   DO 70 K = 1,L
42.000         SUM = SUM + B(I,K)*B(K,J)
43.000    70   CONTINUE
44.000         IF(I - J)75,55,55
45.000    75   B(I,J) = (A(I,J) - SUM)/B(I,I)
46.000    80   SUM = 0.0
47.000    85   CONTINUE
48.000    C CALCULATE CORRECTION VECTORS
49.000         DO 130 I = 1,NOCON
50.000         J = N - I
51.000         IF(J - NOCON)115,125,125
52.000    115  L = J + 1
53.000         DO 120 K = L,NOCON
54.000         SUM = SUM + B(J,K)*DEL(K)
55.000    120  CONTINUE
56.000    125  DEL(J) = B(J,N) - SUM
57.000         SUM = 0.0
58.000    130  CONTINUE
59.000         DO 135 I = 1,NOCON
60.000         PONE(I) = P(I) - DEL(I)
61.000         PHALF(I) = P(I) - 0.5*DEL(I)
62.000    135  CONTINUE
63.000         RONE = 0.0
64.000         RHALF = 0.0
65.000         RZERO = 0.0
66.000         DO 137 I = 1,NO
67.000         CALL RESID(I,PHALF)
68.000         RHALF = RHALF + F(N)*F(N)
69.000         CALL RESID(I,PONE)
70.000         RONE = RONE + F(N)*F(N)
71.000         CALL RESID(I,P)
72.000         RZERO = RZERO + F(N)*F(N)
73.000    137  CONTINUE
74.000         RMIN = 0.5 + (0.25*(RZERO - RONE))/
75.000        *(RONE - 2.0*RHALF + RZERO)
76.000         DO 145 I = 1,NOCON
77.000         PRINT 140,I,P(I)
78.000    140  FORMAT(2X,'P(',I2,') = ',1PD14.7)
79.000    145  CONTINUE
80.000         PRINT 150,RZERO
81.000    150  FORMAT(2X,'RESIDUAL = ',1PD14.7,//)
82.000         IF(SIGMA - 0.0)155,160,155
83.000    155  RATIO = RZERO/SIGMA
84.000         IF(ABS(RATIO) - 0.999999)160,185,270
```

Table VIII—(*Continued*)

```
85.000                C SAVE "B" MATRIX, "P" MATRIX AND "YCALC" MATRIX
86.000           160      DO 165 I = 1,NOCON
87.000                    DO 165 J = 1,NUM
88.000                    SB(I,J) = B(I,J)
89.000           165      CONTINUE
90.000                    DO 170 I = 1,NOCON
91.000                    SP(I) = P(I)
92.000           170      CONTINUE
93.000                    DO 172 I = 1,NO
94.000                    SYCALC(I) = YCALC(I)
95.000           172      CONTINUE
96.000                    SRZERO = RZERO
97.000                C CALCULATE NEW VALUES FOR THE PARAMETERS
98.000                    DO 175 I = 1,NOCON
99.000                    P(I) = P(I) - RMIN*DEL(I)
100.000          175      CONTINUE
101.000          180      SIGMA = RZERO
102.000                   GO TO 5
103.000               C CALCULATE INVERSE MATRIX
104.000          185      NUM = NOCON + 1
105.000                   SUM = 0.0
106.000                   DO 205 M = 1,NOCON
107.000                   KO = NUM + M
108.000                   DO 205 I = 1,NOCON
109.000                   J = NUM - I
110.000                   IF(J - NOCON)190,200,200
111.000          190      L = J + 1
112.000                   DO 195 K = L,NOCON
113.000                   SUM = SUM + B(J,K)*V(K,M)
114.000          195      CONTINUE
115.000          200      V(J,M) = B(J,KO) - SUM
116.000                   SUM = 0.0
117.000          205      CONTINUE
118.000               C CALCULATE STANDARD DEVIATION ON EACH PARAMETER
119.000                   VAREXT = RZERO/(FLOAT(NO) - FLOAT(NOCON))
120.000                   DO 210 I = 1,NOCON
121.000                   J = I
122.000                   STNDV(I) = DSQRT(V(I,J)*VAREXT)
123.000          210      CONTINUE
124.000                   PRINT 211, NAME
                  a
125.000          215      FORMAT(///,T3,'CONSTANT NO.',T18,'STNDRD. DEV.',
126.000                 *T32,'95% LOW LIM.',T47,'95% HI LIM.')
127.000               C CALCULATE AND PRINT CONFIDENCE LIMITS
128.000                   DO 225 I = 1,NOCON
129.000                   ABLE = TS*STNDV(I)
130.000                   HILIM(I) = P(I) + ABLE
131.000                   LOLIM(I) = P(I) - ABLE
132.000                   PRINT 220,I,P(I),STNDV(I),LOLIM(I),HILIM(I)
133.000          220      FORMAT(T2,0PI1,T5,1PD12.5,T19,D10.3,T32,
134.000                 *D10.3,T47,D10.3)
135.000          225      CONTINUE
136.000                   PRINT 230,RZERO
137.000          230      FORMAT(///,' RESIDUAL = ',1X,1PD14.7)
138.000                   PRINT 235,VAREXT
139.000          235      FORMAT(////,' VAREXT = ',1X,1PD14.7)
140.000               C CALCULATE CONFIDENCE LIMITS ON YCALC
141.000                   DO 250 I = 1,NO
142.000                   CALL FUNC(I,P,N)
143.000                   SUM1 = 0.0
144.000                   SUM2 = 0.0
145.000                   DO 240 L = 1,NOCON
146.000                   J = L
147.000                   SUM1 = SUM1 + V(L,J)*F(L)**2
148.000          240      CONTINUE
149.000                   KO = NOCON - 1
150.000                   DO 245 L = 1,KO
151.000                   M = L + 1
152.000                   DO 245 J = M,NOCON
153.000                   SUM2 = SUM2 + 2.0*(V(L,J))*F(L)*F(J)
154.000          245      CONTINUE
155.000                   BAKER = DABS(VAREXT*(SUM1*SUM2))
156.000                   DOG = TS*DSQRT(BAKER)
157.000                   HILIMY(I) = SYCALC(I) + DOG
158.000                   LOLIMY(I) = SYCALC(I) - DOG
159.000                   DIF = Y(I) - SYCALC(I)
160.000                   PRCNT(I) = (DIF/Y(I))*100.0
161.000          250      CONTINUE
162.000                   PRINT 255
163.000          255      FORMAT(///,T6,'Y',T15,'YCAP',T27,'% DIFF.',
164.000                 *T39,'LOLIM',T50,'HILIM')
165.000                   DO 265 I = 1,NO
166.000                   PRINT 260,Y(I),SYCALC(I),PRCNT(I),LOLIMY(I),
167.000                  *HILIMY(I)
```

(*Continued*)

Table VIII—(*Continued*)

```
168.000                 260    FORMAT(T2,1PD10.3,2X,D10.3,2X,D10.3,2X,D10.3,
169.000                       *2X,D10.3)
170.000                 265    CONTINUE
171.000                 C CALCULATE AND PRINT THE CORRELATION COEFFICIENT
172.000                        SUMY = 0.0
173.000                        DO 267 I = 1,NO
174.000                 267    SUMY = SUMY + Y(I)**2
175.000                        R2 = (SUMY - RZERO)/SUMY
176.000                        R = DSQRT(R2)
177.000                        PRINT 268, R2,R
178.000                 268    FORMAT(///,T3,'R SQUARED =',1X,D12.3,3X,
179.000                       *'R =',1X,D12.3,//)
180.000                 C PRINT THE CO-VARIANCE MATRIX
181.000                        DO 269 I = 1,NOCON
182.000                        DO 269 J = 1,NOCON
183.000                        PRINT 271,I,J,V(I,J)
184.000                 271    FORMAT(2X,'V(',I1,',',I1,') = ',2X,1PD14.7)
185.000                 269    CONTINUE
186.000                        GO TO 285
187.000                 C RECALL "B" MATRIX AND "P" ARRAY FROM STORAGE
188.000                 270    IF(SIGMA - RZERO)272,185,185
189.000                 272    NUM = (2*NOCON) + 1
190.000                        DO 275 I = 1,NOCON
191.000                        DO 275 J = 1,NUM
192.000                        B(I,J) = SB(I,J)
193.000                 275    CONTINUE
194.000                        DO 280 I = 1,NOCON
195.000                        P(I) = SP(I)
196.000                 280    CONTINUE
197.000                        RZERO = SRZERO
198.000                        GO TO 185
199.000                 285    RETURN
200.000                        END
```

[a] At this point the following two lines should be added:

```
125.000        211     FORMAT (///,T3,60A1;//)
126.000                PRINT 215
```

All subsequent lines should then be renumbered.

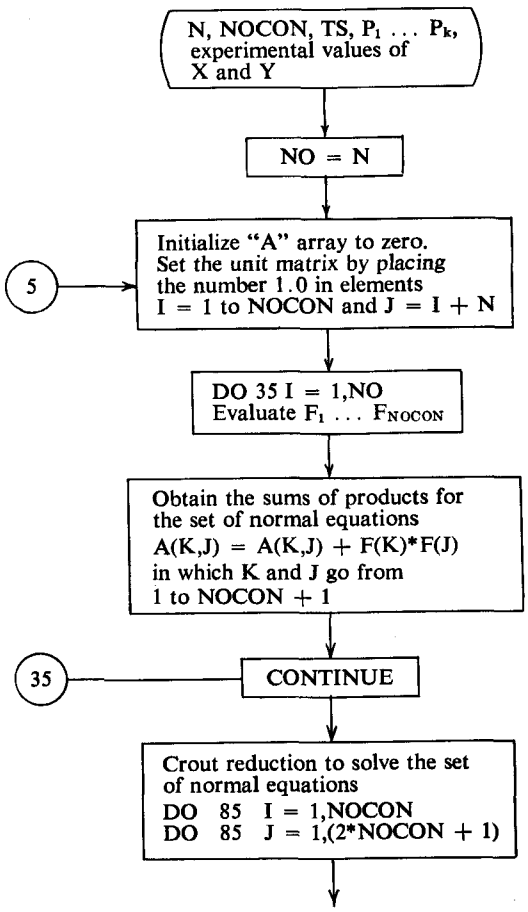
is given in Table VII. Instead of inputting initial estimates of the parameters in reference line 9, the variables TAU and ATAU are input, along with an estimate of P (4) which corresponds to the constant a of Eq. 1. The variables TAU and ATAU are used to convert the rheometer readings into shear stress and shear rate, respectively. This is done in reference lines 12–15. Initial estimates of the constants P_1, P_2, and P_3 of Eq. 20 are obtained in lines 16–18. Reference lines 21–35 are used for printing out additional information generally not included in the basic program.

Subroutine NONLIN—This subroutine is used for the actual curve-fitting procedure and is shown in Table VIII. It does not have to be altered and remains constant regardless of the function to be analyzed. It does call the subroutine FUNC, which must be defined for each function that is to be analyzed.

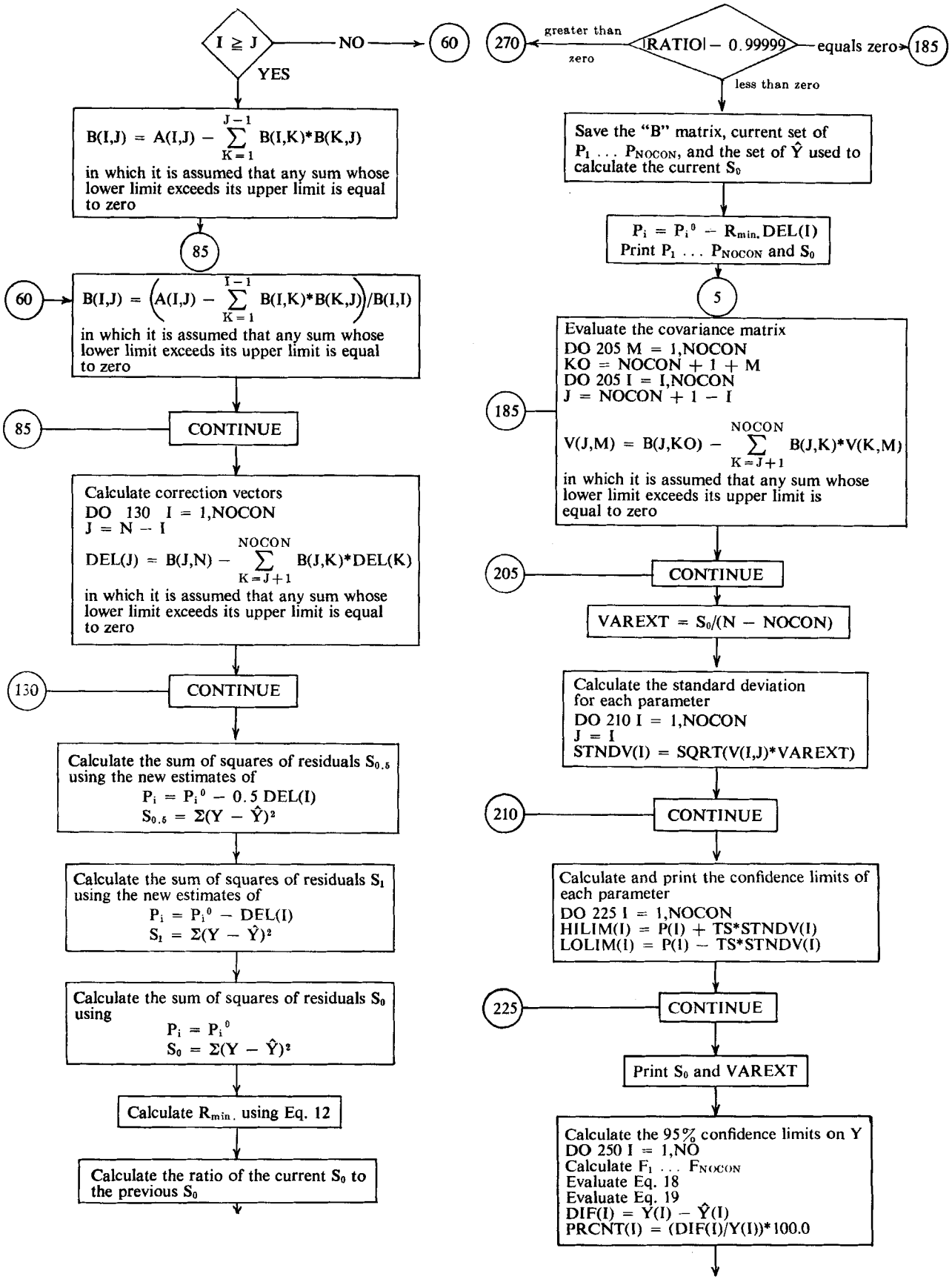
Subroutine FUNC—This subroutine calculates the values of the partial derivatives, as defined in Eq. 7, for each set of experimental data points. The listing of this subroutine is given in Table IX. Reference lines 1–4 should remain unchanged from function to function. The statement immediately following the statement on reference line 9 must contain the definition for YCALC (that is, the function to be fitted must be defined mathematically). The partial derivatives of YCALC with respect to P_1, P_2, *etc.*, are defined starting with reference line 5. This particular listing for subroutine

Table IX—Program Listing for Subroutine FUNC, Which Converts Eqs. 18 and 19 into Computer Language

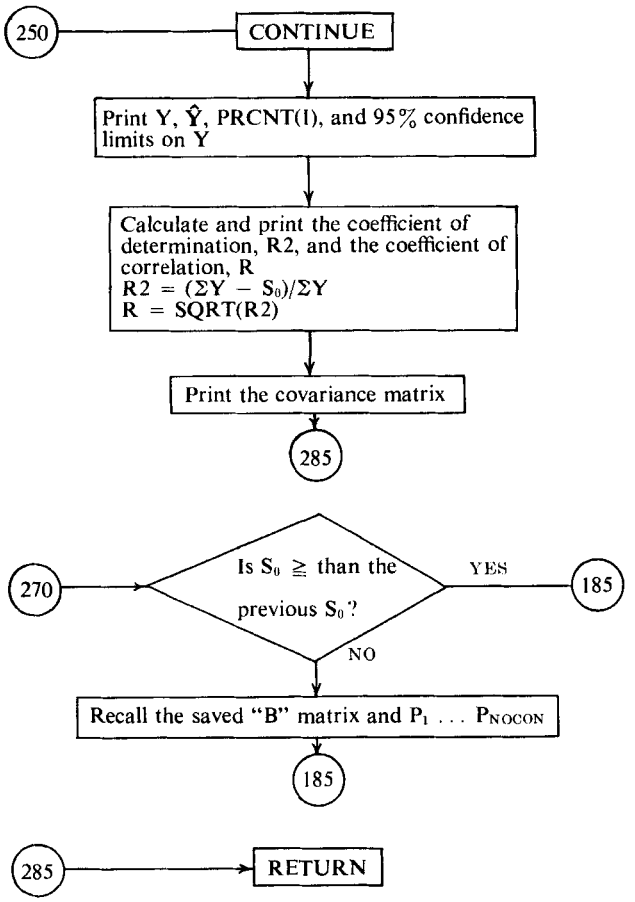
```
 1  SUBROUTINE FUNC(I,P,N)
 2  IMPLICIT REAL*8 (A-H,O-Z)
 3  COMMON/SET2/Y(100),YCALC(100),X(100),F(10)
 4  DIMENSION P(10)
 5  F(1) = 1.0
 6  F(2) = X(I)
 7  F(3) = -DEXP(-P(4)*X(I))
 8  F(4) = X(I)*P(3)*DEXP(-P(4)*X(I))
 9  ENTRY RESID(I,P)
10  YCALC(I) = P(1) + P(2)*X(I) - P(3)*DEXP(-P(4)*X(I))
11  F(N) = Y(I) - YCALC(I)
12  RETURN
13  END
```



Scheme I—Flow Diagram for Subroutine NONLIN



Scheme I—(Continued)



Scheme I—(Continued)

FUNC shows the appropriate definitions for the structure equation and converts Eqs. 20 and 21 into computer language.

Computer Printout—The first set of printout data gives the estimates of each parameter and the sum of squares of residuals for each iteration. When the iteration process is satisfied, the printout skips to the next page and prints the title of the problem, the final value for each constant, the estimated standard deviation for each constant, and the 95% confidence interval for each constant.

The next section of printout gives the final sum of squares of residuals and the estimate of the unit variance, VAREXT. This section is followed by one which gives the experimental and calculated values for the dependent variable (in this instance, shear stress), the relative difference between the two, and the 95% confidence interval for the calculated values. This section is followed by an estimate of the coefficient of determination, which gives the fraction of the variance accounted for by regression, and the correlation coefficient.

The final section of printout gives the elements of the covariance matrix. These elements are useful for determining the variance on any constant that is a function of the parameters adjusted through the nonlinear regression analysis (Eqs. 16 and 17).

A flow diagram for the subroutine NONLIN as presented in this article is given in Scheme I. This should prove useful in the event modifications are needed for a particular computer system.

REFERENCES

(1) W. E. Deming, "Statistical Adjustment of Data," Wiley, New York, N. Y., 1943, and Dover Publications, New York, N. Y., 1964.

(2) H. O. Hartley, *Technometrics*, **3**, 269(1961).

(3) H. O. Hartley and A. Booker, *Ann. Math. Statist.*, **36**, 638 (1965).

(4) J. G. Wagner, *Clin. Pharmacol. Ther.*, **8**, 201(1967).

(5) J. G. Wagner and C. M. Metzler, *J. Pharm. Sci.*, **56**, 658 (1967).

(6) *Ibid.*, **58**, 87(1969).

(7) W. Lowenthal and B. L. Vitsky, *J. Pharm. Sci.*, **56**, 169(1967).

(8) R. Shangraw, W. Grim, and A. M. Mattocks, *Trans. Soc. Rheol.*, **V**, 247(1961).

(9) W. Grim, "Rheology of Pharmaceutical Suspensions," Ph.D. dissertation, University of Michigan, Ann Arbor, Mich., 1959.

(10) K. L. Nielsen, "Methods in Numerical Analysis," Macmillan, New York, N. Y., 1964, pp. 185–197.

ACKNOWLEDGMENTS AND ADDRESSES

Received February 10, 1971, from the *Department of Pharmacy, Philadelphia College of Pharmacy and Science, Philadelphia, PA 19104*

Accepted for publication May 21, 1971.

Presented to the Basic Pharmaceutics Section, APhA Academy of Pharmaceutical Sciences, San Francisco meeting, March 1971.