한국통계학회논문집 제2권 2호
1995년 12월 pp. 135-145

Statistical Estimation and Algorithm in Nonlinear Functions[1]

Jea-Young Lee[2]

Abstract

A new algorithm was given to successively fit the multiexponential function/nonlinear function to data by a weighted least squares method, using Gauss-Newton, Marquardt, gradient and DUD methods for convergence. This study also considers the problem of linear-nonlinear weighted least squares estimation which is based upon the usual Taylor's formula process.

1. Introduction

Expressions of quantitative physiological data are based on either an empirically derived equation(s) which summrizes the phenomenon under study or a theoretically derived statistical model. Of course, they contain parameters with unknown values.

We consider the following nonlinear regression model

$$y(t)=f(x_t,\theta^*)+\varepsilon_t, \qquad t=1,2,...,N, \tag{1.1}$$

where $f^t(x_t,\theta^*)=[f(x_1,\theta^*),......,f(x_N,\theta^*)]$ is the response function, $\theta^*=M\times 1$ is an M dimensional vector of unknown parameters, and the represent unobservable observational or experiment errors.

We will assume at first that these errors $\varepsilon^t=(\varepsilon_1,......,\varepsilon_N)$ are independently and normally distributed with mean zero and unknown variance σ^2. This kind of nonlinear parameter estimation problem is an important step in compartmental modeling, i.e., fitting $y(t)=\sum_{i=1}^{n}A_ie^{\lambda_it}$.

Let $F(\theta)$ be the N x M matrix with elements $\frac{\partial}{\partial\theta_j}f(x_t,\theta)$, where time t indexes the rows and j indexes the columns of the matrix .

1) This research was supported by the Yeungnam University Research Grants in 1994.
2) Department of Statistics, Yeungnam University, Gyongsan, 712-749, KOREA.

Now, $y = (\theta^*) + \varepsilon$

$\cong f(\theta_0) + F(\theta_0)(\theta^* - \theta_0) + \varepsilon$ for initial value θ_0

$= f(\theta_0) + F(\theta_0)\theta^* - F(\theta_0)\theta_0 + \varepsilon$

Then $y - f(\theta_0) + F(\theta_0)\theta_0 \cong F(\theta_0)\theta^* + \varepsilon$

So the estimate of θ^* in $y = f(\theta^*) + \varepsilon$ is approximately equal to the estimate of θ^* in $z = F(\theta_0)\theta^* + \varepsilon$ where $z \cong y - f(\theta_0) + F(\theta_0)\theta_0$.

The nonlinear regression (cf. Bard, 1974; Box, 1971) uses an iterative process : an initial value for θ^* is chosen and continually improved until the errors sum of squares $\varepsilon'\varepsilon$(SSE) is minimized. This study considers the problem of linear-nonlinear weighted least squares estimation which is based upon the usual Taylor's formula process in Section 2. In Section 3, a new algorithm was given to successively fit the multiexponential function/nonlinear function to data by a weighted least squares method, using Gauss-Newton, Marquardt, gradient and DUD(Doesn't Use Derivatives) methods for convergence.

2. Weighted Nonlinear Least Squares Estimators

Given observation $y(t)$, $t=1,2,\ldots,N$ any vector $\hat{\theta}$ that minimizes the sum of squares function

$$SSE(\theta) = [y - f(\theta)]^t [y - f(\theta)] \tag{2.1}$$

will be called a least squares estimator (LSE) for θ^*

The statistical behavior in large samples is suggested by the approximations

$$\hat{\theta} \cong \theta^* + (F^t F)^{-1} F^t \varepsilon \quad \text{and} \tag{2.2}$$

$$s^2 = \frac{\varepsilon^t [I - F(F^t F)^{-1} F^t] \varepsilon}{N - M} \tag{2.3}$$

where s^2 is the estimate of the variance of the ε_t and $F = (\frac{\partial}{\partial \theta_j} f(x_t, \theta))$ is N x M matrix of derivatives where t is the row index and j is the column index. These equations indicate that in large samples the random variable $\hat{\theta}$ has a M-dimensional multivariate normal distribution with mean θ^* and variance-covariance matrix $\sigma^2 (F^t F)^{-1}$, and that $\frac{(N-M)s^2}{\sigma^2}$ is independently distributed as a chi-squared variate with $N-M$ degrees of freedom. Then

$\frac{\hat{\theta}_i - \theta^*_i}{\sqrt{\hat{\sigma}^2 \hat{c}_{ii}}}$ is distributed as a t-distribution with $N-M$ degrees of freedom where is i-th diagonal element of $[F^t(\hat{\theta})F(\hat{\theta})]^{-1}$. A $(1-\alpha)100$ % confidence interval for θ^*_i is $\hat{\theta}_i \pm \sqrt{s^2 \hat{c}_{ii}}$.

The hypothesis $H_0 : \theta_i = \theta_{i0}$ can be tested at the α level of significance using

$$\text{test statistic} \qquad |\ \tilde{t}_i\ | = \frac{|\ \hat{\theta}_i - \theta_{i0}\ |}{\sqrt{s^2 \hat{c}_{ii}}} \sim t_{\frac{\alpha}{2}}(N-M),$$

where H_0 is rejectd if $|\ \tilde{t}_i\ | > t_{\frac{\alpha}{2}}(N-M)$.

The classical linear or the nonlinear regression theory is based on the assumption that error terms of the model under consideration are sequentially independent,
that is,

$$Var(\varepsilon) = \sigma^2 I \tag{2.4}$$

In many practical(medical) problems, however, this assumption is not appropriate (cf. Dell et. al.,1973; DiStefano et. al.,1984; Landaw et. al.,1984). In this section, the assumption (2.4) will be replaced in the nonlinear model (1.1) with an error term ε , for which

$$E(\varepsilon\varepsilon^t) = \sigma^2 V \tag{2.5}$$

where V is an $N \times N$ known positive definite matrix. If V is a diagonal matrix, but with unequal elements, then the errors (thus, the observations) are uncorrelated with unequal variances, while if some of the off-diagonal elements of V are non-zero then the errors are correlated. Examples of correlated errors arise in time series models where observations occur at successive intervals of time and in numerous engineering applications. Then errors can be accounted for by using weighted least squares.

For example, if the error terms ε_t's satisfy the differences

$$\sum_{i=0}^{q} a_i \varepsilon_{t-i} = u_t , \qquad t = 1,2,\ldots\ldots \tag{2.6}$$

where $a_0 = 1$, q is a known positive integer and $\{ u_t \}$ is a pure random process (i.e., white noise) with zero mean and variance $\sigma^2 > 0$, then the process $\{ \varepsilon_t \}$ is called an autoregressive process of order q (AR(q)).

Let $y = f(\theta^*) + \varepsilon$, with ε independently and normally distributed with mean zero and

unknown variance $\sigma^2 V$, where V is an $N \times N$ known positive matrix.
The general idea in estimating nonlinear parameters is :

$$Q^{-1}y = Q^{-1}f(\theta^*)+Q^{-1}\varepsilon \text{ , where } QQ^t \text{ for some nonsingular Q.}$$

Then $Q^{-1}\varepsilon \sim N(0,\sigma^2 I)$.

So, the LSE of θ using $Q^{-1}y$ is the $\hat{\theta}$ that minimizes

$$\begin{aligned} SSE(\theta) &= [Q^{-1}y-Q^{-1}f(\theta)]^t[Q^{-1}y-Q^{-1}f(\theta)] \\ &= (y-f(\theta))^t(Q^{-1})^tQ^{-1}(y-f(\theta)) \\ &= (y-f(\theta))^tV^{-1}(y-f(\theta)) \end{aligned} \tag{2.7}$$

A vector $\hat{\theta}$ which minimizes the generalized least squares function

$$[(y-f(\theta))^tV^{-1}(y-f(\theta))]$$

will be called a *generalized least squares estimator* (GLSE) (cf. Christensen ,1990, 1987; Berger et. al.,1992) for, and it depends on y, where
$y^t=(y(1),....,y(N))$ and $f(\theta^*)=[f(x_1,\theta^*),......,f(x_N,\theta^*)]$

In particular, if V is a diagonal matrix then $\hat{\theta}$ is a weighted least squares estimates (WLSE) for θ^*. If V = I , then $\hat{\theta}$ is an ordinary least squares estimator (LSE).

The iterative algorithm for obtaining $\hat{\theta}$ is based on Taylor's expansion theorem

$$f(\theta) \cong f(\theta_0)+F(\theta_0)(\theta^*-\theta_0) \text{ for the initial value } \theta_0 \tag{2.8}$$

and $F=(\frac{\partial}{\partial\theta_j}f(x_t,\theta))$ is an $N \times M$ matrix.

Then $SSE(\theta)=\| Q^{-1}[y-f(\theta_0)]-Q^{-1}F(\theta_0)(\theta^*-\theta_0)\|^2$

So the next θ_1 that minimizes SSE(θ) is

$$\begin{aligned} \theta_1-\theta_0 &\cong [Q^{-1}F(\theta_0)^t(Q^{-1}F(\theta_0)]^{-1}(Q^{-1}F(\theta_0))^t[Q^{-1}y-Q^{-1}f(\theta_0)] \\ &= [F^t(\theta_0)V^{-1}F(\theta_0)]^{-1}F^t(\theta_0)V^{-1}[y-f(\theta_0)] \end{aligned}$$

We get $\theta_1=\theta_0+(F^tV^{-1}F)^{-1}F^tV^{-1}r$, where $r=y-f(\theta_0)$.

To obtain the asymptotic distribution theory, Equation (2.8) is repeated expanding at $\theta=\theta^*$. Then we get

$$\hat{\theta}=\theta^*+(F^tV^{-1}F)^{-1}F^tV^{-1}\varepsilon \quad \text{where} \quad F=F(\theta^*) \text{ .} \tag{2.9}$$

The estimate of the variance for ε corresponding to the least squares estimator $\hat{\theta}$ is

$$s^2 = \frac{\varepsilon^t(I-A)^t V^{-1}(I-A)\varepsilon}{N-M} \quad \text{where } A = F(F^t V^{-1} F)^{-1} F^t V^{-1}.$$

These two equations, assuming $\varepsilon \sim N(0,\sigma^2 V)$ with V known positive definite, indicate that in large samples the vector $\hat{\theta}$ has a M-dimensional multivariate normal distribution with mean θ^* and variance-covariance matrix $\sigma^2(F^t V^{-1} F)^{-1}$, and that $\frac{(N-M)s^2}{\sigma^2}$ independently distributed as a chi-squared variate with $N-M$ degrees of freedom (i.e., $\chi^2(N-M)$). Other possible tests include likelihood ratio tests. The asymptotic properties of some of these have been investigated by Gallent (1975). Halperin (1963), and Hartley (1961) show in setting that it is possible to construct exact tests for certain hypothesis in nonlinear regression.

Consider a general solution form represented by an n-compartment linear model (cf. Landaw et. al., 1984), with observations in a single compartment following an impulse input or a constant step input. We have seen that the model output may be represented by the time-dependent function

$$y(t) = \sum_{i=1}^{n} A_i e^{\lambda_i t} \qquad \lambda_i < \lambda_{i+1} \leq 0 \tag{2.10}$$

Here y(t) is in effect the prediction of the data by the model at the time t. For this model, we have the parameter vector $\theta = [A_1, \lambda_1, \ldots\ldots, A_n, \lambda_n]^t$ as an ordered set of the $2n$ parameter constants. Fitting multiexponential form is nonlinear regression and can be obtained by several packages (i.e., SAS, BMDP etc.). But some specific improvement of sums-of-exponentials in algorithm properties can be achieved by exploiting the observation that the model is only nonlinear in half the parameters (see next section). The following is a physiology example of fitted parameters with specific weight.

Example 2.1 (Landaw & DiStefano(1984)). Suppose a set of noise data $y(t_1), y(t_2), \ldots, y(t_N)$ collected from a system $y(t) = z(t,p) + \varepsilon(t)$ has been fitted to the biexponential function Equation (2.10), using a weighted least squares procedure. After a rapidly applied dose D=100, the noise is thought to have 10 % coefficient of variation. Table 2.1 summarizes observed values $y(t_i)$ collected from this system at nine times t_i, plus their weights w_i used in performing a WLS analysis. Table 2.2 is a summary of output after performing a WLS fit of multiexponential models of orders 1, 2, and 3 to the data in Table 2.1. The two-exponential fit shows a significant improvement(p = 0.04) but the three-exponential fit doesn't (p = 0.99). This is an evidence that the two-exponential model fits well enough.

Table 2.1 Sample times, observed values, and weights

Run Title and Date			
Datum Point	Time	Observed (y) Value	Data w_i
1	0.0000	102.3000	0.00956
2	0.5000	71.7000	0.01945
3	1.0000	41.4000	0.05834
4	2.0000	35.5000	0.07935
5	3.0000	18.0000	0.30864
6	4.0000	13.0000	0.59172
7	6.0000	8.07000	1.53551
8	8.0000	3.64000	7.54740
9	10.000	1.97000	25.76722

All data is weighted by 1/(error variance). Constant Coef. of variation = 10.00%.

Table 2.2 Parameter estimate summary for sum of N exponentials

	Number of Exponentials		
	1	2	3
A_1	67.4410	53.1601	27.4852
λ_1	-0.3674	-1.6705	-2.3654
A_2		50.33452	27.7052
λ_2		-0.3241	-1.1799
A_3			48.6972
λ_3			-0.3203

Summary Statistics

Final weighted RSS	30.8456	8.6417	8.5923
df	7	5	3
WMSE (WRSS/df)	4.4066	1.7283	2.8641
F ratio (2 vs. 1, 3 vs. 2)		6.4238	0.0086
P value of F test		0.0415	0.9914

3. Numerical Algorithm of Weighted Linear-Nonlinear Regression

An algorithm of weighted nonlinear regression is presented in which all the parameters to be estimated can be regarded as nonlinear (the traditional approach) or reclassified as linear-nonlinear. The theoretical basis for the reexpression approach is given and an example is presented which allows a comparison of the all nonlinear to the linear-nonlinear method employing widely used iterative techniques such as Hartley (modified Gauss-Newton) (1975,1961), Marquardt(1963), Gradient (1989), and DUD (1979).

The weighted nonlinear least squares problem may be defined as the estimation of parameters in a model by the method of least squares when the parameters enter into the model nonlinearly. Let $y = f(\theta^*)+\varepsilon$, where ε represents the $N \times 1$ observational error vector arising from independent multivariate normal distributions with mean 0 and variance $\sigma^2 V$, V known positive definite and θ is the $M \times 1$ vector of parameters . Also, we define $f(\theta) = [f(x_1,\theta),\ldots,f(x_N,\theta)]$as an $N \times 1$ matrix of predicted values . Then the parameters in the equation $y = f(\theta^*)+\varepsilon$ can be logically expressed in linear and nonlinear form :

$$y = \begin{bmatrix} g_1(x_1;\beta), & g_2(x_1;\beta), & . & . & . & , g_q(x_1;\beta), \\ g_1(x_2;\beta), & g_2(x_2;\beta), & . & . & . & , g_q(x_2;\beta), \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ g_1(x_N;\beta), & g_2(x_N;\beta), & . & . & . & , g_q(x_N;\beta), \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ . \\ \alpha_q \end{bmatrix} + \varepsilon. \qquad (3.1)$$

where $\alpha_1,\alpha_2,\ldots,\alpha_q$ are q linear parameters, β is the r $\times 1$ vector of nonlinear parameters.

Then the equation $y = f(\theta^*)+\varepsilon$ takes on the multiple regression form , $G\alpha+\varepsilon$, where G is the above matrix. So we have $E(Y) = G\alpha$, where the expected value of a vector y of N observations is equal to a linear combination of the q columns of the $N \times q$ G matrix of linear independent functions $g(x_i;\beta)$ with α as the $q \times 1$ vector of linear parameters.

Now, Equation (3.1) yields $SSE(\theta) = [y-f(\theta)]^t V^{-1}[y-f(\theta)]$and the estimated parameter $\hat{\theta}$ is the WLSE such that minimizes $SSE(\theta) = [y-f(\theta)]^t V^{-1}[y-f(\theta)]$ or $\sum_{h=1}^{N} w_h(y_h-f(x_h;\theta))^2$, for particular choice w_h.

Therefore, applying Equation (3.16),

$$\begin{aligned} SSE(\theta) &= [y-f(\theta)]^t V^{-1}[y-f(\theta)] \\ &= [y-G\alpha]^t V^{-1}[y-G\alpha] \\ &\equiv K(\alpha,\beta). \end{aligned}$$

The problem becomes one of calculating q $\hat{\alpha}$'s and r $\hat{\beta}$'s using weighted least squares estimates of α and β to yield the minimum value of $K(\alpha,\beta)$.
So we have

$$\hat{\alpha} = G^t_\beta (G^t_\beta V^{-1} G_\beta)^{-1} G^t_\beta V^{-1} y \ , \tag{3.2}$$

and then

$$SSE(\beta \mid \hat{\alpha}) = y^t(I-G_\beta(G^T_\beta V^{-1} G^T_\beta)^{-1} G^T_\beta V^{-1})y, \tag{3.3}$$

which is a function of the r nonlinear parameters of β which reaches its minimum at $\hat{\beta}$. Of course, the parameter space for the nonlinear part has been reduced in dimension from $M = q + r$ to $M = r$. The algorithm to be presented employs the above technique of reexpressing the nonlinear parameters into linear-nonlinear forms, calculating the linear parameters by linear regression, and using variations of the various methods discussed before (e.g. Gauss-Newton, etc.) to calculate the least square estimates of the nonlinear parameters.

Proposition 3.1. $SSE(\theta) = (r-F\delta)^t W(r-F\delta)$, where W is a weight function,

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \cdot \\ \cdot \\ \cdot \\ r_N \end{bmatrix} = \begin{bmatrix} y_1 - f(x_1;\theta^{0)} \\ y_2 - f(x_2;\theta^{0)} \\ \cdot \\ \cdot \\ \cdot \\ y_N - f(x_N;\theta^{0)} \end{bmatrix}, \ \delta = \begin{bmatrix} \theta_1 - \theta^0_1 \\ \theta_2 - \theta^0_2 \\ \cdot \\ \cdot \\ \cdot \\ \theta_M - \theta^0_M \end{bmatrix},$$

$$F = \begin{bmatrix} \frac{\partial f(x_1;\theta)}{\partial \theta_1} & \frac{\partial f(x_1;\theta)}{\partial \theta_2} & \cdots & \frac{\partial f(x_1;\theta)}{\partial \theta_M} \\ \frac{\partial f(x_1;\theta)}{\partial \theta_1} & \frac{\partial f(x_1;\theta)}{\partial \theta_2} & \cdots & \frac{\partial f(x_1;\theta)}{\partial \theta_M} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \frac{\partial f(x_N;\theta)}{\partial \theta_1} & \frac{\partial f(x_N;\theta)}{\partial \theta_2} & \cdots & \frac{\partial f(x_N;\theta)}{\partial \theta_M} \end{bmatrix}.$$

Proof) Let $y = f(\theta^*) + Q^{-1}\varepsilon$, where ε is $n \times 1$ vector of $N(0,\sigma^2 V)$ and V is known positive definite , taking inverse matrix above,

$\leftrightarrow$ $Q^{-1}y = Q^{-1}f(\theta^*) + Q^{-1}\varepsilon$, where $Q^{-1}\varepsilon \sim N(0,\sigma^2 I)$.

and we have $SSE(\theta) = (y-f(\theta))^t V^{-1}(y-f(\theta))$

$$= \sum_{h=1}^{N} w_h[y_h - f(x_h;\theta)]^2, \text{ for special weight } w_h ,$$

$$\simeq \sum_{h=1}^{N} w_h[y_h - f(x_h;\theta_0) - F(\theta_0)(\theta-\theta_0)]^2$$

$$= \sum_{h=1}^{N}[y_h - f(x_h;\theta_0) - \sum(\theta_i - \theta^0_i)\frac{\partial f(x_h;\theta)}{\partial \theta_i} \mid_{\theta_0}]^2.$$

Then

$$w^{1/2}(r-p\delta) = w^{1/2}\left[\begin{bmatrix} y_1 - f(x_1;\theta^0) \\ \cdots\cdots \\ y_N - f(x_N;\theta^0) \end{bmatrix} - \begin{bmatrix} \frac{\partial f(x_1;\theta)}{\partial \theta_1} & \cdots & \frac{\partial f(x_1;\theta)}{\partial \theta_M} \\ \cdots & \cdots & \cdots \\ \frac{\partial f(x_N;\theta)}{\partial \theta_1} & \cdots & \frac{\partial f(x_N;\theta)}{\partial \theta_M} \end{bmatrix}_{(\theta=\theta_0)} \begin{bmatrix} \theta_1 - \theta^0_1 \\ \cdots \\ \theta_M - \theta^0_M \end{bmatrix}\right]$$

$$= w^{1/2}\begin{bmatrix} y_1 - f(x_1;\theta^0) - \sum_{i=1}^{M}(\theta_i - \theta^0_i)\frac{\partial f(x_1;\theta)}{\partial \theta_i}\Big|_{\theta^0} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ y_N - f(x_N;\theta^0) - \sum_{i=1}^{M}(\theta_i - \theta^0_i)\frac{\partial f(x_N;\theta)}{\partial \theta_i}\Big|_{\theta^0} \end{bmatrix}$$

Therefore , we proved $(r-F\delta)^t W(r-F\delta) = [W^{1/2}(r-F\delta)]^t[W^{1/2}(r-F\delta)] = SSE(\theta)$

Proposition 3.2. $A\delta = v$ where $A = F^t V^{-1} P$ and $v = F^t V^{-1} r$.

Proof :

From proposition 3.1 and $W = V^{-1}$, we know that

$$SSE(\theta) = (r-F\delta)^t W(r-F\delta)$$
$$\simeq \sum_{h=1}^{N} w_h[y_h - f(x_h;\theta_0) - F(\theta_0)(\theta-\theta_0)]^2.$$

By taking derivative with respect to δ and equating to 0, i.e. $\frac{\partial SSE(q)}{\partial \delta_i} = 0$, for $i = 1,2,..,M$. we obtain $A\delta = v$.

Theorem 3.3. If reexpress the nonlinear parameters in $A\delta = v$ in linear-nonlinear form, then

$$\delta_2 = (A_{22} - A_{21} A_{11}^{-1} - A_{12})^{-1} v_2, \quad (3.4)$$

where $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, $\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, and δ_1, v_1, are the linear terms and δ_2, v_2 are the nonlinear terms.

Proof) With the choice of the WLSE of α^0 given β^0, then $v_1 = 0$ and so

$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ v_2 \end{bmatrix}$. Therefore , $\delta_2 = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} V_2$.

Equation (3.4) is used by the program at each iteration to obtain the correction terms for the nonlinear parameters. Th equation essentially replaces the equation in proposition 3.2 which would be used when all the parameters are considered as nonlinear. The algorithm allows the option of using the Gauss-Newton, Marquardt, Gradient or DUD technique to obtain the solution to Equation (3.4).

In summary, the algorithm schemes are :

(a) Choose initial values , β^0 (nonlinear parameters) , and calculate initial linear parameters , $\alpha^0 = G^t_\beta (G^t_\beta V^{-1} G_\beta) G^t_\beta V^{-1} y$.

(b) Compute iteration $\delta_2 = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} V_2$. for next using Gauss-Newton, Marquardt , DUD, or Gradient methods, and then calculate α.

(c) Check for convergence .

(d) Stop.

Based on the previous physiological Example 2.1 and considering several multiexponential models and the linear-nonlinear algorithms as described above four Gauss-Newton, Marquardt, DUD or Gradient methods, we make the following conclusions.

4. Advantages of Linear-Nonlinear Algorithm and Results

In many cases tried, it appeared that the proposed method was the more stable method in comparison with the usual method and one conclude the following in regard to new algorithm presented:

(a) Easier to choose initial values.

Models	Nonlinear Model	Linear-Nonlinear Model
1 Exponential form	Choose 2 parameters	Choose 1 parameters (only)
2 Exponential form	Choose 4 parameters	Choose 2 parameters (only)
3 Exponential form	Choose 6 parameters	Choose 3 parameters (only)

(b) One step regression (for linear part) is much better (exact) than that of the graphical curve-peeling method.

(c) If we use Gradient or DUD method , Linear-Nonlinear is particularly better than Nonlinear Model .

(d) If we use Gauss-Newton or Marquardt , the number of iterations need for convergence in Linear-Nonlinear method is greater than the number in Nonlinear method, but the total cpu time is likely to be less.

References

[1] Bard, Jonathan (1974). *Nonlinear Parameter Estimation*, New York, Academic Press.

[2] Berger, R. L. and Cassella, George (1992). Deriving generalized means as least squares and maximum likelihood estimates, *The American Statistician,* Vol. 46, No. 4, 279-282.

[3] Box, M. J. (1971). Bias in nonlinear estimation, *Journal of the Royal Statistical Society,* Vol. B32, 171-201.

[4] Christensen, R. (1990). *Linear Models for Multivariate, Time-series, and Spatial Data,* Springer-Verlag, New-York.

[5] Christensen, R. (1987). *Plane Answers to Complex Questions,* Springer-Verlag, New York.

[6] Cook, R. D. and Tsai, C. (1985). Residuals in nonlinear regression, *Biometrika,* Vol. 72, No. 1, 23-29.

[7] Dell, R. B., Sciacca, R., Lieberman, K., Case, D. B., and Cannon, P. J. (1973). A weighted least-squares technique for the analysis of kinetic data and its application to the study of renal xenon washout in dogs and man, *Circulation Research,* Vol. 32, 71-84.

[8] DiStefano III., J. J. and Landaw, E. M. (1984). Multiexponential, multicompartmental, and noncompartmental modeling. I. Methodological limitations and physiological interpretations, *Americian Journal of Physiology,* Vol. 246, R651-R664.

[9] Gallant, A. R. (1975). Nonlinear Regression, *Americian Statistician,* Vol. 29, No. 2, 73-81.

[10] Hartley, H. O. (1961). Modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares, *Technometrics,* Vol. 3, 269-280.

[11] Landaw, E. M. and DiStefano III., J. J. (1984). Multiexponential, multicompartmental, and noncompartmental modeling II. Data analysis and statistical considerations, *Americian Journal of Physiology,* Vol. 246, R665-R677.

[12] Lehmann E. L. (1987). *Testing Statistical Hypothesis,* 2nd Ed., John Wiley & Sons.

[13] Marquardt, D. W. (1963). An algorithm for least squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics,* Vol. 11, 431-441.

[14] Ralston, M.L. and Jennrich R.I. (1979). DUD, a derivative-free algorithm for nonlinear least squares, *Technometrices,* Vol. 1, 7-14.

[15] Villegas, C. (1969). On the least squares estimation of nonlinear relations, *Annals of Mathematical Statistics,* Vol. 40, No. 2, 462-466.