Talanta, Vol. 40, No. 2, pp. 269–277, 1993
Printed in Great Britain.

0039-9140/93 $6.00 + 0.00

MULTIPARAMETRIC CURVE FITTING XIV*

MODUS OPERANDI OF THE LEAST-SQUARES ALGORITHM MINOPT

JIŘÍ MILITKÝ
Department of Textile Materials, Technical University, CS-461 17 Liberec, Czech Republic

MILAN MELOUN
Department of Analytical Chemistry, University of Chemical Technology, CS-532 10 Pardubice, Czech Republic

(*Received* 9 *December* 1991. *Revised* 29 *June* 1992. *Accepted* 29 *June* 1992)

Summary–Hybrid least-squares algorithm MINOPT for a nonlinear regression is introduced. MINOPT from CHEMSTAT package combines fast convergence of the Gauss–Newton method in a vicinity of minimum with good convergence of gradient methods for location far from a minimum. Quality of minimization and an accuracy of parameter estimates for six selected models are examined and compared with different derivative least-squares methods of five commercial regression packages.

In literature many regression algorithms and program packages for non-linear regression are described and classified.[1] According to their practical applicability in the chemical laboratory the program's modus operandi may be elucidated using a block structure classification:[2,3] regression program may be divided into functional blocks as INPUT, RESIDUAL SUM OF SQUARES, MINIMIZATION, STATISTICAL ANALYSIS, DATA SIMULATION, ADDITIONAL SUBROUTINES, *etc*. An amount of useful information achieved from program application, efficiency and reliability of results can be deduced from

(i) a numerical point-of-view which concerns ability to reach a minimum of the regression criterion (subroutines of a MINIMIZATION block);

(ii) a statistical point-of-view which concerns quality of statistical information (subroutines of STATISTICAL ANALYSIS block).

According to these two blocks the commonly used programs are not always reliable. Due to a great variability of regression models, regression criteria and data the effective algorithms enabling sufficiently fast convergence to a global extreme are not available. Some algorithms and programs often fail, *i.e.*, converge very slow or diverge.

In this paper we concentrate on procedures of derivative methods for the least-squares (LS) criterion which represents a very large group of methods today.[4] Some numerical aspects of the algorithm MINOPT are presented. Its numerical quality is examined and compared with other derivative methods on selected mathematical models usually found in problems of reaction kinetics and solution equilibria studies.

RESIDUAL SUM OF SQUARES BLOCK

In the classical setting the additive model of measurements is adopted

$$y_i = f(x_i; \beta) + \epsilon_i, \quad i = 1, \ldots, n \qquad (1)$$

In model (1) the y_i is the response (experimental quantity), x_i are non-stochastic explanatory variables (without detriment to generality, x is supposed to be scalar), $f(x_i, \beta)$ is a regression model containing the $(m \times 1)$ parameter vector β and ϵ_i is the so called (experimental) error.

The main task of regression is to find estimators, $\hat{b}$, of an unknown parameter vector β. A process of parameter estimation is based on assumptions about errors ϵ: classical presumption requires the errors ϵ to be independent and identically distributed random variables having normal distribution $N(0, \sigma^2)$ with zero mean and constant variance σ^2. Based on these

*Part XIII, *Talanta*, 1988, **35**, 981.

assumptions the sufficient estimates $\hat{b} = \{b_1, \ldots, b_m\}$ can be obtained minimizing the least-squares criterion

$$U(\hat{b}) = \sum_{i=1}^{n} [y_i - f(x_i; \hat{b})]^2, \quad n \geqslant m \qquad (2)$$

MINIMIZATION BLOCK

For minimization of $U(\hat{b})$ criterion a lot of various derivative and non-derivative algorithms exist.[4-8] Derivative algorithms are useful for all model functions which are twice differentiable. In a sequel we concentrate on derivative methods and LS criterion only.

The main disadvantage of derivative methods is a local convergence which depends on a choice of an initial guess $\hat{b}^{(0)}$. All algorithms of this group are of iterative nature. In the i-th iteration a procedure starts from the estimates $\hat{b}^{(i)}$ to which a suitable increment vector $d^{(i)}$ is added:

$$\hat{b}^{(i+1)} = \hat{b}^{(i)} + d^{(i)} \qquad (3)$$

The vector $d^{(i)}$ is considered acceptable if

$$U(\hat{b}^{(i)} + d^{(i)}) \leqslant U(\hat{b}^{(i)}) \qquad (4)$$

Here, the increment vector can be expressed by relation

$$d^{(i)} = \alpha_i V \qquad (4a)$$

where V is directional vector and α is scalar. Some algorithms admit equality or even a small increase of $U(\hat{b}^{(i+1)})$ against $U(\hat{b}^{(i)})$. Procedure of a search of minimum $U(\hat{b})$ consists of the following four steps:

1. *Determination of initial guess of parameters* $\hat{b}^{(0)}$

This step is decisive for many algorithms for successful minimization. From a good initial guess $\hat{b}^{(0)}$ the simple algorithms usually converge. For a very poor initial guess a minimum cannot be found by any methods of this group.

2. *Determination of direction vector* V

Derivative of a LS criterion function $U(\hat{b})$ in a point $(\hat{b} + \alpha V)$ according a scalar α has form

$$\frac{\delta U(\beta)}{\delta \alpha} = \left[\frac{\delta U(\beta)}{\delta \beta}\right]^T \frac{\delta \beta}{\delta \alpha} \qquad (5)$$

For $\alpha \to 0$ we get from equation (5) so called directional derivative

$$S_D = \left.\frac{\delta U(\beta)}{\delta \alpha}\right|_{\alpha \to 0} = g^T V \qquad (6)$$

where g is the gradient vector of $U(\hat{b})$ whose elements g_j are equal to $\delta U(\hat{b})/\delta b_j$. The steepest decrease of a criterion function is in the direction $-g$. The condition of acceptability of the directional vector V requests that the directional derivative is not positive. Any direction for which an inequality $g^T V > 0$ holds is therefore inconvenient. Moreover, if the directional vector V is acceptable the positive regular definite matrix $\mathbf{R}$ exists so that

$$V = -\mathbf{R}g \qquad (7)$$

The directional derivative S_D is then equal to

$$S_D = -g^T \mathbf{R} g \qquad (8)$$

For a positive definite matrix $\mathbf{R}$ their quadrative forms are always positive so that S_D in equation (8) is negative.

3. *Calculation of minimization step* αV

For calculation of the minimization step (also called the optimal increment or the correction vector) $d = \alpha V$ in direction V the approximation of $U(\hat{b})$ by the Taylor series up to a quadratic term can be used. It leads to form

$$U(\hat{b} + \alpha V) \approx U(\hat{b}) + \alpha g^T V + \frac{\alpha^2}{2} V^T \mathbf{H} V \qquad (9)$$

where $\mathbf{H}$ is symmetric Hessian (matrix) having as elements the second derivatives of $U(\hat{b})$. Equation (9) assumes α to be approximately quadratic so that the optimal value of α may be estimated by putting the first derivative $U(\hat{b} + \alpha V)$ according to α to zero. Solving this equation will give

$$\alpha^* = -\frac{\delta U(\hat{b})}{\delta \alpha} \Big/ \frac{\delta^2 U(\hat{b})}{\delta \alpha^2} = -g^T V [V^T \mathbf{H} V]^{-1} \qquad (10)$$

and after substitution from equation (8) we obtain the so called Raleigh coefficient

$$\alpha^* = g^T \mathbf{R} g \, [g^T \mathbf{R}^T \mathbf{H} \mathbf{R} g]^{-1} \qquad (11)$$

The suitability of Raleigh coefficient α^* is restricted for a region in which the approximation (9) can be used.

For LS criterion $U(\hat{b})$ the gradient g can be expressed in the form

$$g = 2\mathbf{J}^T \hat{e} \qquad (12)$$

and the Hessian $\mathbf{H}$ in the form

$$\mathbf{H} = 2[\mathbf{J}^T \mathbf{J} - \mathbf{W}^T \hat{e}] = 2[\mathbf{J}^T \mathbf{J} - \mathbf{B}] \qquad (13)$$

Here $\hat{e}$ is the residual vector having components

$$\hat{e}_i = y_i - f(x_i; \hat{b}) \qquad (14)$$

J is the Jacobian (matrix) of dimension $(n \times m)$ with elements

$$J_{ij} = \frac{\delta f(x_i, \hat{b})}{\delta b_k}, \quad j = 1, \ldots, n; \quad k = 1, \ldots, m \quad (15)$$

and **W** is a three-dimensional array of dimension $(n \times m \times m)$ which is composed from n layers where the ith one is formed by the matrix $\mathbf{W}_i$ having elements

$$W_{i(j,k)} = \frac{\delta^2 f(x_i, \hat{b})}{\delta b_j \, \delta b_k} \quad (16)$$

4. *Termination of iteration process*

The natural criterion of an optimal estimate $\hat{b}$ is a zero value of the gradient g. Many methods of a minimum search terminate the iterative process when the norm of gradient

$$\|g\|^2 = \sum_{j=1}^{m} g_j^2 \quad (17)$$

is sufficiently small. It is possible to select a critical value of this norm, for example, equal to 10^{-4} *i.e.*, the limit under which the point $\hat{b}^{(i)}$ is considered as a local extreme. Often iterations terminate when too small changes of parameter estimates appear. None of these criteria enable a termination in a minimum. Minimization may terminate less heuristically. From the geometry of LS we get termination criterion as follows: the residual vector $\hat{e}$ is approximately perpendicular on columns of the matrix **J**. This is equal to condition $\mathbf{J}^T\hat{e} = 0$. For cosines of angle α_j between the residual vector $\hat{e}$ and the j-th column J_j of a matrix **J** a simple relation is valid

$$\cos \alpha_j = \hat{e}^T J_j [J_j^T J_j \hat{e}^T \hat{e}]^{-1/2} \quad (18)$$

When a maximal value of $\cos \alpha_j$ is sufficiently small, *e.g.*, smaller than 10^{-9} it is supposed that a minimum $U(\hat{b})$ was reached. Some other termination criteria may be found in Ref. 7.

The following derivative algorithms seem to be dominant in nonlinear regression analysis today:

(a) Gauss–Newton methods;
(b) Marquardt methods;
(c) dog-leg method.

Gauss–Newton methods

For determination of a convenient directional vector V the quadratic approximation of a criterion function $U(\hat{b})$ may be used which also corresponds to equation (9) for $\alpha = 1$. From

$$\frac{\delta U(\hat{b} + V)}{\delta V_i} = 0 \quad (19)$$

the optimal direction vector $V_i = N_i$ in the form

$$N_i = -\mathbf{H}^{-1} g = (\mathbf{J}^T\mathbf{J} + \mathbf{B})^{-1}\mathbf{J}^T\hat{e} \quad (20)$$

is evaluated. Substituting into equation (11) we estimate that $\alpha^* = 1$. Therefore N_i is directly a minimization step d_i and the method is called the Newton–Raphson method. It is obvious that when the criterion $U(\hat{b})$ is a quadratic function (*i.e.*, an elliptic paraboloid) the minimum $\hat{b}$ will be reached in one step. For other forms of criterion function $U(\hat{b})$ and estimates $\hat{b}^{(0)}$ far from β, this method does not converge too fast. Moreover it requires knowledge of an array of second derivatives $\mathbf{W}_i$ for a determination of a matrix **B** in equation (13).

Neglecting matrix **B** is equivalent to a linearization of regression model and is theoretically acceptable for a case when a residual vector $\hat{e}$ is negligible. The corresponding directional vector L_i has the form

$$L_i = (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}\hat{e} \quad (21)$$

and methods are called Gauss–Newton methods. They belong to the simple and the most frequently used procedures of nonlinear regression. When $\mathbf{H} \approx (\mathbf{J}^T\mathbf{J})$ is supplied into equation (11) it leads to $\alpha^* = 1$. From the practical point it is important that the Gauss–Newton method will work well, if some of the following conditions are fulfilled:

I. Residuals $\hat{e}_i = y_i - f(x_i, \hat{b})$ are small.

II. The model function $f(x, \beta)$ is nearly linear *i.e.*, the Hessian **H** has a small norm and its elements are nearly zero.

III. Residuals $\hat{e}_i$ have alternate signs so that **B** is approximately a zero matrix. It is valid in a vicinity of optimum $\hat{b}$.

Extending a region of convergence of this very simple method is possible to reach by different ways:

(a) The technique of an inversion of the matrix $\mathbf{J}^T\mathbf{J}$ and solution of a set of linear equations

$$(\mathbf{J}^T\mathbf{J})L = \mathbf{J}^T\hat{e} \quad (22)$$

(b) Improving a matrix $(\mathbf{J}^T\mathbf{J})$ in order to be close to Hessian **H**.

(c) Choice of a suitable length of the step α.

Marquardt methods

The natural selection of a directional vector V_i is the direction of steepest descent $-g$. It corresponds to a matrix option $\mathbf{R} = \mathbf{E}$. For optimal coefficient α^* in this direction it is from equation (11) that

$$\alpha^* = g^T g [g^T \mathbf{H} g]^{-1} \approx g^T g [g^T (\mathbf{J}^T \mathbf{J})^{-1} g]^{-1} \quad (23)$$

The minimization step $d_i = -\alpha^* g$ corresponds to the gradient method.

The gradient methods converge often slowly in a vicinity of an optimum. On the other hand, in cases when $\hat{b}^{(i)}$ is far from β it enables a direction leading to a minimum to be found. It is effective to use a combination of directions of the Newton method N_i or a direction of linearization L_i together with a direction $-g$ to a construction of the more robust procedures which are also called the hybrid procedures. Known representative is here the Marquardt method which calculates the directional vector $V_i(\lambda)$ by relation

$$V_i(\lambda) = (\mathbf{J}^T \mathbf{J} + \lambda \mathbf{D}_i^T \mathbf{D})^{-1} \mathbf{J}^T \hat{e} \quad (24)$$

where λ is the parameter and $\mathbf{D}_i$ is the diagonal matrix which eliminates an influence of various magnitudes of components of the matrix $\mathbf{J}$. Usually the diagonal elements D_{ii} are equal to diagonal elements of matrix $(\mathbf{J}^T \mathbf{J})$. Convenient selection of a parameter λ ensures:

(1) positive definiteness of a matrix $\mathbf{R} = (\mathbf{J}^T \mathbf{J} + \lambda \mathbf{D}^T \mathbf{D})$ which is necessary for its invertibility;

(2) a shortening step $V_i(\lambda)$ moving from a direction of linearization L_i;

(3) a possibility of a selection between a direction L_i and approximate direction $-g$. Step length in direction $-g$ is however equal to zero;

(4) a restriction of a magnitude of the incremental vector V_i to the certain "admissible" region in a vicinity of $\hat{b}^{(i)}$.

The necessity of repeated matrix inversion for each λ is a disadvantage of this procedure which is rather time-consuming. Moreover a situation may happen that for large λ a magnitude V_i is too small. Therefore the maximal magnitude of λ is limited. Individual modifications of the Marquardt method differ especially in strategy of the adaptive setting of parameter λ.

Generally it is valid that methods of Marquardt type are for their robustness a standard part of library programs of most computer packages.

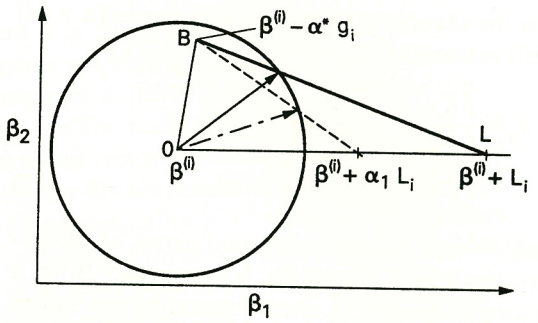

Fig. 1. Geometrical interpretation of dog–leg strategy. The circle shows admissible range of increments. Solid hypotenuse is $V(\mu)$ for $\alpha_1 = 1$ and dotted hypotenuse is $V(\mu)$ for $\alpha_1 = 1$.

Dog-leg methods

Among the main disadvantages of the Marquardt method are:

(a) a necessity of matrix inverse at change of parameter λ;

(b) a small length of vector $V(\lambda)$ for a large λ.

Both these disadvantages are removed in hybrid methods when the optimal directional vector $V(\mu)$ is the convex combination of vectors L and the vector $-\alpha^* g_i$. It holds that

$$V(\mu) = \hat{b}^{(i)} + (1 - \mu) L_i \alpha_1 - \mu \alpha^* g_i \quad (25)$$

Here α^* is estimated from equation (23) and condition $0 \leqslant \mu \leqslant 1$ is valid. The function $V(\mu)$ for cases $\alpha_1 = 1$ and $\alpha_1 < 1$ on Fig. 1 hypotenuses of right angle triangles with dotted line for $\alpha_1 < 1$ and solid line for $\alpha_1 = 1$. Classical strategy of the Powell dog–leg method estimates an optimal vector $V_i(\mu)$ on the abscissa TB of a triangle defined by vertices $0 = \hat{b}^{(i)}$; $T = \hat{b}^{(i)} + L_i$; $B = \hat{b}^{(i)} - \alpha^* g_i$ where α^* is defined by equation (23).

It is obvious that for $\mu = 0$ the vector $V(\mu)$ is identical with a linearization direction L_i and for $\mu = 1$ with a direction of negative gradient $-g$. The magnitude of a total increment in direction $-g$ correspond to the optimal value α^*.

Dennis and Mei[10] used the "shorter" vector $\alpha_1 L_i$ instead of a vector L_i. The parameter α_1 is determined that the increment in a linearization direction approximately corresponds to a Raleigh point, *cf.* Ref. 10.

$$\alpha_1 = 0.2 + 0.8 \, \|g_i\|^4 \times [g_i^T (\mathbf{J}^T \mathbf{J})^{-1} g_i g_i^T (\mathbf{J}^T \mathbf{J}) g_i]^{-1} \quad (26)$$

From Fig. 1 it is obvious that shortening $L_i \alpha_1$ leads to a directional vector $V_i^*(\mu)$ which is

closer to a linearization direction than the vector $V(\mu)$ calculated at option $\alpha_1 = 1$. MINOPT algorithm[11] uses $V_1^*(\mu)$ directional vector. For solution of matrix inverse problems a rational rank technique (*i.e.*, special pseudoinversion) is adopted. A special heuristic strategy for constraining a maximum step length based on quality of quadratic approximation of $U(\hat{b})$ is used here.

Other blocks as STATISTICAL ANALYSIS, GOODNESS-OF-FIT TEST, DATA SIMULATION, *etc.* will be described in the next contributions of this series.

Software

Program MINOPT from CHEMSTAT package carries out the numerical and statistical analysis of a non-linear regression model $f(x;\beta)$ with use of modified "double dog–leg" strategy. Input consists from the experimental data (x_i, y_i), $i = 1, \ldots, n$, and the initial guess of parameters estimates $\hat{b}^{(0)}$. The user supplies the regression model. All required derivatives are calculated numerically.

Program CHEMSTAT is available from authors on request.

RESULTS AND DISCUSSION

Comparison of some commercial packages for nonlinear regression

In a study of reaction kinetics and solution equilibria, the regression analysis of frequently used nonlinear models requires an estimation of unknown parameters of exponentials or parameter powers. To examine the reliability of MINOPT algorithm six testing problems have been chosen. Models I, II, and III are selected from literature. Models IV and VI are based on simulated data and Model V is based on experimental data. Testing models with their data and available initial guess of parameters are summarized below. To compare parameter estimates $\hat{b}$ and $U(\hat{b})$, no restart or repeated determination with new initial guess of parameters in divergence or failing were allowed. Commercial packages BMDP (*i.e.*, BMDP PC-90), SAS (*i.e.*, SAS version 6.03), SYSTAT (*i.e.*, SYSTAT version 5.01), SPSS (*i.e.*, SPSS PC+ version 3.1), ASYST (*i.e.*, ASYSTANT+ version 1.5), STATGR (*i.e.*, STATGRAPHICS version 5.0) and CHEMSTAT (*i.e.*, CHEMSTAT version 1.25) were used,[11,12] *cf.* Table 3.

Six tested models with data:

Model I. $y = \beta_1 + \beta_2 \exp(\beta_3 x)$

x	1	5	10	15	20	25	30	35	40	50
y	16.7	16.8	16.9	17.1	17.2	17.4	17.6	17.9	18.1	18.7

Model II. $y = \exp(\beta_1 x) + \exp(\beta_2 x)$

x	1	2	3	4	5	6	7	8	9	10
y	4	6	8	10	12	14	16	18	20	22

Model III. $y = \beta_1 \exp\left[\dfrac{\beta_2}{\beta_3 + x}\right]$

x	50	55	60	65	70	75	80	85
y	34780	28610	23650	19630	16370	13720	11540	9744

90	95	100	105	110	115	120	125
8261	7030	6005	5147	4427	3820	3307	2872

Model IV. $y = \beta_1 \exp(\beta_3 x) + \beta_2 \exp(\beta_4 x)$

x	7.448	7.448	7.969	8.176	9.284	9.439	7.552
y	57.544	53.546	19.498	16.444	4.305	3.006	45.290

7.877	8.552	9.314	7.607	7.847	8.176	8.523
27.952	11.803	4.764	51.286	31.623	21.777	13.996

8.903	9.314
7.727	4.999

Model V. $y = \beta_1 x^{\beta 3} + \beta_2 x^{\beta 4}$

x	12	13	14	15	16	17	18	19	20
y	7.31	7.55	7.80	8.05	8.31	8.57	8.84	9.12	9.40

21	22	23
9.69	9.99	10.30

Model VI. $y = \beta_1 [\exp(-\beta_2 x_1) + \exp(\beta_3 x_2)]$

x_1	0	0.6	0.6	1.4	2.6	3.2	0.8	1.6	2.6	4.0
x_2	0	0.4	1.0	1.4	1.4	1.6	2.0	2.2	2.2	2.2
y	40	10	5.0	2.5	2.5	2.0	1.0	0.7	0.8	0.7

1.2	2.0	4.6	3.2	1.6	4.2	4.2	3.2	2.8
2.6	2.6	2.8	3.0	3.2	3.4	3.4	3.8	4.2
0.4	0.4	0.3	0.22	0.22	0.1	0.05	0.07	0.03

4.2	5.4	5.6	3.2
4.2	4.4	4.8	5.0
0.03	0.03	0.02	0.01

Table 1. Initial guess of parameters estimated for six tested models

Model	$b_1^{(0)}$	$b_2^{(0)}$	$b_3^{(0)}$	$b_4^{(0)}$	$U(b^{(0)})$
I	1	1	1	—	$2 \cdot 10^{43}$
II	0.3	0.4	—	—	$4 \cdot 10^{3}$
III	0.02	4000	250	—	$1.7 \cdot 10^{9}$
IV	10^5	10^5	−1.679	−1.31	$1.12 \cdot 10^{4}$
V	100	0.1	2	10	$2.68 \cdot 10^{3}$
VI	12	1.0	25	—	226.9

Table 2. Best estimate of parameters of six various tested models

Model	b_i	b_2	b_3	b_4	$U(b)$
I	15.67	0.994	0.0222	—	$5.98 \cdot 10^{-3}$
II	0.2578	0.2578	—	—	124.34
III	0.005618	6180	345.2	—	87.9
IV	$8.315 \cdot 10^{7}$	$5.088 \cdot 10^{3}$	−1.95	−0.7786	134
V	3.802	$4.141 \cdot 10^{-3}$	0.223	2.061	$2.98 \cdot 10^{-5}$
VI	31.5	1.51	19.9	—	1.25

Table 3. Results of six analyzed models

Model I

Program	Method	Solution	Note	RSS
BMPD	3R-Gauss	False	Local minimum	3.68
	AR(DUD)	False	Local minimum	3.68
SAS	Gauss–Newton	False	Local minimum	1.903
	Marquardt	o.k.		5.987E-03
	Gradient	False	Local minimum	1.903
	DUD	False	Local minimum	2.036
SPSS	Marquardt	o.k.		5.986E-03
STATGR	Marquardt	Aborted	Overflow	4.011
ASYST	Gauss–Newton	False	Local minimum	67.76
	Var. metric	False	Divergence	
	Hybrid. method.	Aborted	System error	
SYSTAT	Var. metric	False	Local minimum	3.68
	Simplex	False	Local minimum	3.68
CHEMSTAT MINOPT		o.k.	28 iterations	5.986E-03

Model II

Program	Method	Solution	Note	RSS
BMDP	3R-Gauss	False	Local minimum	259.28
	AR(DUD)	False	Local minimum	1063.0
SAS	Gauss–Newton	False	Local minimum	3400
	Marquardt	o.k.	10 iterations	124.36
	Gradient	False	Very slow converg.	245.4
	DUD	DUD	Nearly o.k.	127.0
SPSS	Marquardt	o.k.		124.4
STATGR	Marquardt	o.k.	10 iterations	124.36
ASYST	Gauss–Newton	False	Program error	
	Var. metric	False	Program error	
	Hybrid. method	o.k.		124.36
SYSTAT	Var. metric	False	Local minimum	2000
	Simplex	o.k.	5 iterations	124.36
CHEMSTAT MINOPT		o.k.	10 iterations	124.36

continued

Table 3—*continued*

Model III

Program	Method	Solution	Note	RSS
BMDP	3R-Gauss	o.k.	11 iterations	87.95
	AR(DUD)	o.k.	160 iterations	87.95
SAS	Gauss–Newton	False	No convergence	1.6E + 09
	Marquardt	False	No convergence	6.9E + 06
	Gradient	False	No convergence	6.9E + 06
	DUD	o.k.	2.66 iterations	87.95
SPSS	Marquardt	o.k.		87.95
STATGR	Marquardt	False	Local minimum	9.0E + 04
ASYST	Gauss–Newton	False	No convergence	6.9E + 06
	Var. metric	False	Program error	
	Hybrid. method	False	Program error	
SYSTAT	Var. metric	False	Slow converg. err.	1.7E + 03
	Simplex	o.k.	160 iterations	87.95
CHEMSTAT MINOPT		o.k.	47 iterations	87.95

Model IV

Program	Method	Solution	Note	RSS
BMDP	3R-Gauss	False	No convergence	1.8E + 04
	AR(DUD)	False	Stack overflow	
SAS	Gauss–Newton	False	Local minimum	9.59
	Marquardt	False	No convergence	1.8E + 04
	Gradient	False	No convergence	1.3E + 04
	DUD	False	No convergence	1.8E + 04
SPSS	Marquardt	o.k.		3.18E − 04
STATGR	Marquardt	o.k.	28 iterations	3.179E − 04
ASYST	Gauss–Newton	False	No convergence	6.9E + 06
	Var. metric	False	Program error	
	Hybrid. method	False	Program error	
SYSTAT	Var. metric	o.k.	44 iterations	3.179E − 04
CHEMSTAT MINOPT		o.k.	37 iterations	3.179E − 04

Model V

Program	Method	Solution	Note	RSS
SPSS	Marquardt	False	Underflow error	
CHEMSTAT MINOPT		o.k.	47 iterations	128.98

Model VI

Program	Method	Solution	Note	RSS
SPSS	Marquardt	False	Very slow converg.	97.8
CHEMSTAT MINOPT		o.k.	51 iterations	2.98E − 05

Table 4. Performance index PI for tested packages

Package	PI[%] (tests 1–4)	PI[%] (tests 1–4)
BMDP	25	—
SAS	25	—
SYSTAT	37.5	—
STATGR	50	—
ASYST	8.3	—
SPSS	100	66.6
CHEMSTAT	100	100

Initial guess of parameters (Table 1), parameters estimates (Table 2) and results of convergence (Table 3) for six tested models are summarized. Detailed results may be found in the forthcoming[12] textbook or from the authors. For overall comparison of packages the Performance Index PI was computed

$$PI = \frac{100 * (\text{number of correct results})}{T * (\text{number of used methods in package})}$$

where T is the number of tests. From a numerical viewpoint the greater value of PI indicates the better package. Performance index PI for all tested packages are summarized in Table 4.

CONCLUSION

From this comparative study it can be deduced that the best results have been obtained using MINOPT procedure. Even this comparison may disappoint some users of standard statistical packages as it indicates that errors due to a false optimum, saddle points or a flat $U(b)$ function can often cause failure of the whole regression analysis.

REFERENCES

1. D. A. Ratkowsky, *Nonlinear Regression Modelling*, Marcel Dekker New York, 1983.
2. M. Meloun and M. Javůrek, *Talanta* 1985, **32,** 973.
3. M. Meloun, J. Havel and E. Högfeldt, *Computation of Solution Equilibria*, Ellis Horwood, Chichester, 1988.
4. P. E. Gill, W. Murray, M. M. Wright, *Practical Optimization*, Academic Press, London, 1981.
5. R. Schmidt, *Advances in Nonlinear Parameter Optimization*, Springer, Berlin, 1982.
6. A. R. Gallant, *Nonlinear Statistical Models*, Wiley, New York, 1987.
7. Y. Bard, *Nonlinear Parameter Estimation*, Academic Press, New York, 1974.
8. D. M. Bates and D. G. Watts, *J. Roy. Stat. Soc.* 1980, **B24,** 1.
9. J. E. Dennis and H. H. W. Mei, *J. Opt. Theor. Appl.*, 1979, **28,** 453.
11. J. Militký and J. Čáp, *Proc. Conf. CEF 87*, Taormina, Sicilia, May 1987.
12. M. Meloun, J. Militký and M. Forina, *Chemometrics in Instrumental Analysis*, Vol. 1, Solved Problems by IBM PC, Vol. 2. Interactive Model Building on IBM PC, Ellis Horwood, Chichester, 1992.