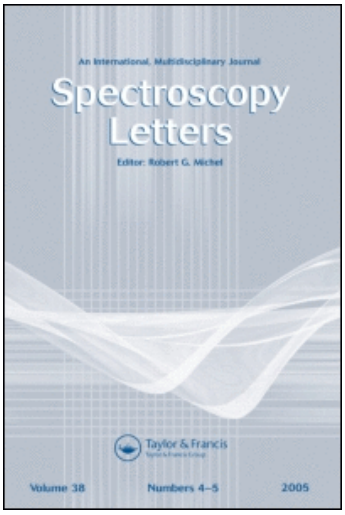
This article was downloaded by:
On: *30 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Automatic Analysis of Raman Spectra

L. A. Slavić[a]; S. Š Ašić[a]
[a] Department of Physical Chemistry, University of Belgrade, Belgrade, Yugoslavia

To cite this Article Slavić, L. A. and Ašić, S. Š(1995) 'Automatic Analysis of Raman Spectra', Spectroscopy Letters, 28: 5, 783 — 794
To link to this Article: DOI: 10.1080/00387019508009919
URL: http://dx.doi.org/10.1080/00387019508009919

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

SPECTROSCOPY LETTERS, 28(5), 783-794 (1995)

AUTOMATIC ANALYSIS OF RAMAN SPECTRA

I.A. Slavić, S. Šašić

Department of Physical Chemistry, University of Belgrade
P.O.BOX 137 Belgrade Yugoslavia

Downloaded At: 03:43 30 January 2011

Abstract. In this paper the automatic analysis of Raman spectra is described. The method applied is so-called SRIP, i.e. "Standard Representation Independent Parameter" fitting, which has been given previosly and proven to be steadily converging, if the model has been chosen properly. The approach is extended here by introducing the interactive continuation of the analysis: some of peaks may be rejected or left in the model with fixed up parameters of the position and full width at half maximum .

The main improvements of the algorithm is to be applied to the Raman spectra analysis include the mixed mathematical model as well as the enhanced sensitivity in discovering of peaks. That is achieved by the analysis of smoothed residuals at the steps of the peak revelation.

This approach enables successful analysis of spectra with low signal to noise ratio, without limitation in number of the peaks. The reliability of Raman band analysis was proven using overlapped spectra of solution of noninteracting liquids: carbontetrachloride (CTC), dichloromethane (DCM) and 1,2 dichloroethane (DCE) with different molar ratio. Solution spectra can be calculated from component spectra and compared with analysis by described method.

Introduction

Recently (1), the SRIP-method (*Independent Parameter in Standard Representation)* has been applied successfully to the automatic analysis of soft X-ray emission spectra. The rigorous mathematical theory, developed by Slavic and Slavic (2), given for linear and nonlinear case, has been given in Appendix A and Appendix B respectively; the uncorrected print errata in those parts forced us to repeat the Appendices in the present paper.

Copyright © 1995 by Marcel Dekker, Inc.

Let us present an original spectrum data by the set

$$(V_i, W_i) \in R^2, \quad i \in N\,[0,m], \tag{1}$$

where $M = m + 1$ is the number of data pairs for which we seek a model in original space; The model is given by sum of the J (Gaussian+Lorentzian-type) -band functions and polynomial of the third degree, describing the background, hence

$$W_i(V_i) = \sum_{j=1}^{J} \{ G_j + L_j \} + P_3(V_i) \pm E_i , \tag{2}$$

where

$$G_j = b_{4j-3}\exp\{ -\ln 2[2(V_i - b_{4j})/b_{4j-1}]^2\}, \tag{2'}$$

$$L_j = b_{4j-2}/\{ 1 + \ln 2[2(V_i - b_{4j})/b_{4j-1}]^2\}. \tag{2"}$$

Note that the b_{4j-1} parameter for the j-th peak means really full width at half maximum (FWHM) only if the band will be done by pure Gaussian.

Downloaded At: 03:43 30 January 2011

General Procedure

The procedure starts by the transformation of data described in details in Appendix A and dealing onwards with the (x_i, y_i) set of data in SR^2 space.

The maximum value of the spectrum in SR^2 space is given value 1 and represents the height of the greatest available peak.The fourth power of this spectrum gives a chart where the main peak is separated with the FWHM nearly equal to a half of the original peak - the other peak heights become suppressed (3,4). The analysis of this chart yields the parameters of the "first", i.e. the greatest peak. Multiplying the obtained FWHM by two we get the peak in the starting set, therefrom we strip the peak, smooth the residuals and continue the search for the next peak, and so on.The more peaks added to the model, the next observed peak is smaller until the desired sensitivity of the "peak discovering" is reached.

The mathematical model in SR^2 space is given by

$$Y_i(x_i) = \sum_{j=1}^{J} \{ G_j + L_j \} + P_3(x_i) \pm \in_i , \tag{3}$$

where

$$G_j = a_{4j-3}\exp\{ -[(x_i - a_{4j})/a_{4j-1}]^2\}, \tag{3'}$$

and

$$L_j = a_{4j-2}/\{ 1 + [(x_i - a_{4j})/a_{4j-1}]^2\}. \tag{3"}$$

Notice that the model is simple, i.e. without constants in order to avoid redundant multiplications during the terative fitting procedure by the algorithm described in Appendix B.

The transformation (A5) returns the parameters into R^2 as

(1) Gaussian height $b_{4j-3} = a_{4j-3} (W_{max} - W_{min})$,

(2) Lorentzian height $b_{4j-2} = a_{4j-2} (W_{max} - W_{min})$,

(3) FWHM $b_{4j-1} = a_{4j-1} (V_{max} - V_{min})*1.66524$,

(4) position $b_{4j} = a_{4j} (V_{max} - V_{min}) + V_{min}$,

where $j = 1, 2, \ldots, J$ is the current peak index for the total number of J peaks. Thus we calculate

(5) $Area_j = \{ 1.0645* b_{4j-3} + 1.8867*b_{4j-2} \} * b_{4j-1}$.

Downloaded At: 03:43 30 January 2011

Results and Discussion

The purpose of this paper was mainly to prove reliability of the method proposed. Therefore we used different mixtures of the carbontetracloride (CTC), dicloromethane (DCM) and 1,2 dicloroethane (DCE), with the known band characteristics (5) in the range of 250 - 380cm^{-1} (284.7 cm^{-1} for DCM, 300.9 cm^{-1} for DCE, and 313.7 cm^{-1} for CTC). The similarity of the physical and chemical properties of these substances results in their very weak mutual interaction. In order to study the program response in the cases of low signal to noise ratio, we made the same spectra with the small laser power as well as with more narrow slit.

Raman spectra has been obtained at room temperature with SPEX 1401 spectrometer by excitation line 514.5 nm of an Ar^+ ion laser Spectra Physics model 2020. The power of incident beam were 200 mW (slit width 300 μm) and 40 mW (slit width 100 μm) for the first and the second series of spectra respectively.

Figures 1 and 2 show some illustrative spectra obtained by the different ratio of the components. Tables I and II show their parameters in comparison to the pure supstances respectively.

We would like to emphasize that, for the fig.1 , discrepancies in the position parameter reach not more than 1 cm^{-1} and differences in the FWHM not more than 1.8 cm^{-1} regardless of the quality of the spectra or the component ratio. The results are something poorer in a cases where smaller peaks have about ten times lower intensity than most intensive peak in spectra. Such example is shown on fig. 2.

Conclusion

Extensive number of known Raman spectra analysis has shown that the method proposed can be applied with a high reliability for studying complex bands, where the lower peaks have the heights even less than 10 % of

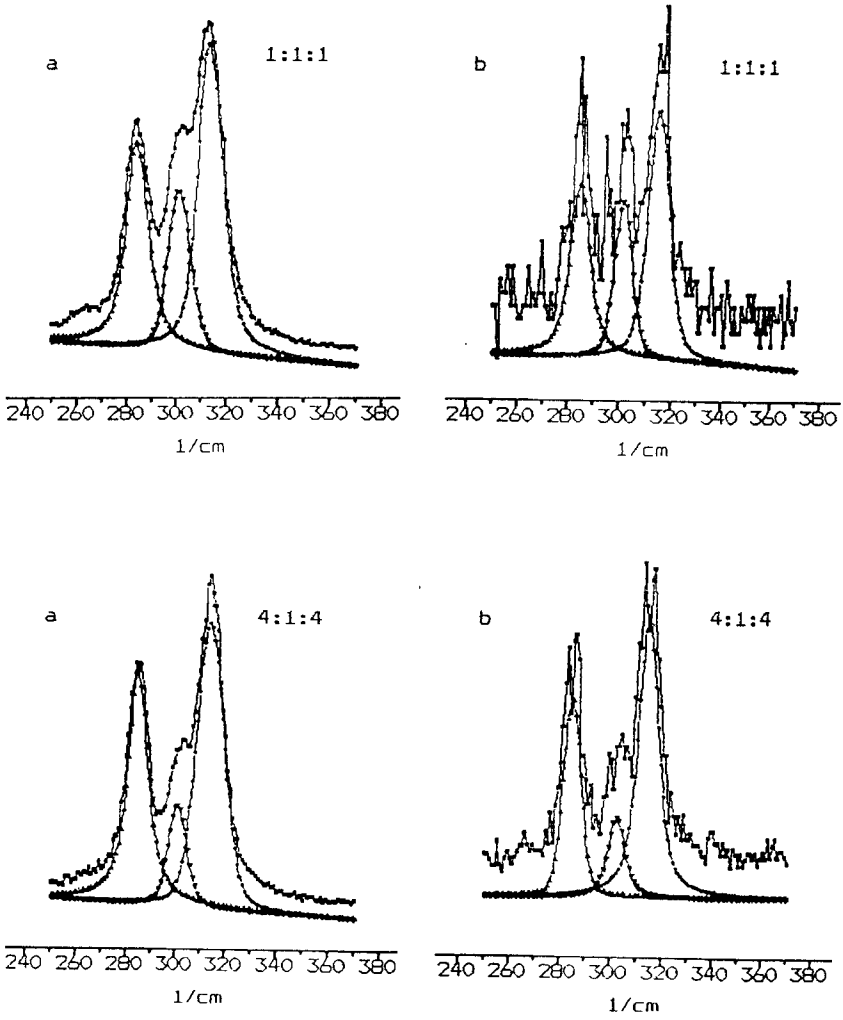


Figure 1. Spectra of solution in 240 - 380 cm^{-1} region. The digits in upper right corner indicate DCM:DCE:CTC volume ratio: a) spectra recorded by incident power 200 mW and slit width 300 μm, b) spectra recorded by 40 mW and 100 μm, ■ - experimental curve, ● - CTC, ▼ - DCE, ▲ - DCM.

Downloaded At: 03:43 30 January 2011

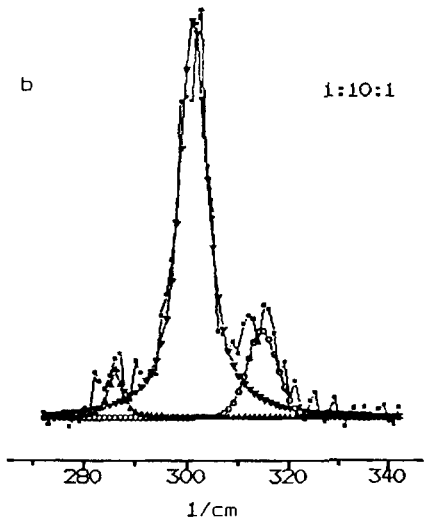


Figure 2. The spectra of solution DCM:DCE:CTC with volume ratio 1:10:1 between 270 - 340 cm^{-1}.

The labels are same like in previous figure.

Downloaded At: 03:43 30 January 2011

Table I: Parameters of peaks recorded by incident power 40 mW and slit width 100 μm.

DCM:DCE:CTC		max.	pure sup.	FWHM	pure sup.
1 : 1 : 1	DCM	284.4	284.9±0.2	8.3	8.2±0.5
	DCE	301.2	300.9±0.4	7.8	7.5±0.3
	CTC	314.6	313.7±0.6	9.5	10.7±0.7
4 : 1 : 4	DCM	284.6	284.9±0.2	8.7	8.2±0.5
	DCE	301.9	300.9±0.4	8.1	7.5±0.3
	CTC	314.2	313.7±0.6	9.2	10.7±0.7
1 : 10 : 1	DCM	286.1	284.9±0.2	3.2	8.2±0.5
	DCE	301.2	300.9±0.4	6.9	7.5±0.3
	CTC	314.4	313.7±0.6	8.51	10.7±0.7

Table II: Parameters of peaks recorded by incident power 200 mW and slit widht 300 μm

DCM:DCE:CTC		max.	pure sup.	FWHM	pure sup.
1 : 1 : 1	DCM	284.4	284.0±0.2	9.61	9.8±0.5
	DCE	301.5	300.9±0.4	10.8	9.8±0.5
	CTC	314.3	313.7±0.6	11.2	12.7±0.2
4 : 1 : 4	DCM	284.7	284.9±0.2	8.0	9.8±0.5
	DCE	301.2	300.9±0.4	8.7	9.8±0.5
	CTC	314.6	313.7±0.6	12.8	12.7±0.2

the greatest peak. Program has been written in C language enabling comfortable interactive analysis with graphics and spending an CPU time of about 0.5 min/peak on PC 386 (25 MHz).It provides successful analysis up to 10 peaks what can be extended if needed.

Downloaded At: 03:43 30 January 2011

APPENDIX A

Polynomial Approximation in Standard Representation (PASR)

If we have a discrete data set

$$(V_i, W_i) \in R^2, \quad i \in N[0,m], \tag{A1}$$

for which we would like to find out the functional relation

$$W(V) \approx F(V), \tag{A2}$$

we can obtain a new data set, i.e.

$$(x_i, y_i) \in SR^2 \equiv R^2[0,1], \; i \in N[0,m] \tag{A3}$$

given by the following simple transformation:

$$x_i = \frac{V_i - V_{min}}{V_{max} - V_{min}}, \quad V_{min} = \min_i \{V_i\}, \; V_{max} = \max_i \{V_i\}, \tag{A4a}$$

and

$$y_i = \frac{W_i - W_{min}}{W_{max} - W_{min}}, \quad W_{min} = \min_i \{W_i\}, \; W_{max} = \max_i \{W_i\}. \tag{A4b}$$

Notice here that

$$0 \le x_i \le 1, \quad 0 \le y_i \le 1, \quad 0 \le i \le m. \tag{A4c}$$

The new data set (x_i,y_i) is said to be the *standard representation* (SR) of the starting data set (V_i,W_i); the transformation returning SR chart to the starting field of values is given by

$$V_i = V_{min} + (V_{max} - V_{min})x_i \tag{A5a}$$

$$W_i = W_{min} + (W_{max} - W_{min})y_i \tag{A5b}$$

The SR has some valuable properties that enable easy solving ofboth interpolating or fitting (smoothing approximation) problems. This is done by parameters disclosing of a linear or nonlinear functional form, which approximates successfully the given SR data, i.e.

$$y \approx f(x), \tag{A2'}$$

Hence, the searched function F,eq.(A2), will be known, implicitly by (A2') , or explicitly expressed by simple returning transformations (A5).

IPF Polynomial Approximation in *SR*

The polynomial approximation appears to be the most simple one in further application.Therefore we consider the case when the SR (A3) can be mathematically identified by the n-th degree polynomial, i.e.

$$y_i(x_i) = P_n(x_i) \equiv \sum_{j=0}^{n} a_j x_i^j, \quad 0 \le i \le m. \tag{A6}$$

In SR the natural constraints appear as follows:

$$P_n(x_0=0) \equiv a_0 = y_0, \tag{A7}$$

$$P_n(x_m=1) \equiv a_0 + a_1 + \ldots a_n + = y_m, \tag{A8}$$

and

$$I = \int_0^1 P_n(x)dx \equiv a_0 + a_1/2 + \ldots a_n/(n+1). \tag{A9}$$

If

$$M = m + 1 \tag{A10}$$

is the total number of available data points and $\Delta x \approx 1/M$, then

$$I = \lim_{M \to \infty} [\Sigma y_i/M] \equiv \lim_{M \to \infty} [y] \tag{A9'}$$

Downloaded At: 03:43 30 January 2011

The total number of polynomial parameters is given by

$$N = n + 1 \qquad \text{(A11)}$$

Lemma 1. If the SR set (A3) is represented by the polynomial (A6),and if it can be found the best agreement of the constraint (A8) , i.e. if it can be found the polynomial degree n for which

$$|a_0 + a_1 + \ldots + a_n - y_m| = \min, \qquad \text{(A8')}$$

then, this n is said to be the optimum degree (n_{opt}).

Proof. No matter how parameters of the polynomial (A6) have been found, the parameters make a sequence

$$a_0^{(n)} + a_1^{(n)} + \ldots + a_n^{(n)} - y_m \equiv \delta(n);$$

therefore it can be determined

$$\delta_{min}(n_{opt}) = \min|\delta(n)|.$$

Let us consider the following obviously chain of inequalities:

$$\Sigma(\partial Y_i/\partial a_j)^2 \leq \Sigma(\partial Y_i/\partial a_j)(\Delta Y_i/\Delta a_j) \leq (\Delta Y_i/\Delta a_j)^2, \qquad \text{(A12)}$$

where, in our case

$$Y_i \equiv P_n(x_i), \quad \Delta Y_i \equiv \epsilon_i = y_i - P_n(x_i), \qquad \text{(A13)}$$

From (A12) we derive the Independent Parameter Estimator (IPE) for the (k+1)-st iterative step of the j-th parameter optimization

$$\Delta a_j^{(k+1)} = D_j \sum_i x_i^j \epsilon_i^{(k)}, \qquad \text{(A14)}$$

Here the "weighting value" is given by

$$D_j = 1 / \sum_i x_i^{2j} \qquad \text{(A15)}$$

and should not be computed by (A15) at each step, except the very first one.

Theorem 1. The simple iterative IPF method, for the SR polynomial, given by

$$a_j^{(k+1)} = a_j^{(k)} + \Delta a_j^{(k+1)}, \quad 0 \leq j \leq n, \qquad \text{(A16)}$$

with the IPE (A14) and (A15),is steady converging.

Downloaded At: 03:43 30 January 2011

Proof. *Sufficiency.* We shall prove the sufficient condition: $|IPE| < 1$. Let us consider (A12) as

$$\Sigma x_i^{2j} \leq (1/\Delta a_j)\, \Sigma x_i^j \epsilon_i \leq (1/\Delta a_j)^2\, \Sigma \epsilon_i^2. \qquad \text{(A12')}$$

Let it be

$$|\epsilon_i|/|\Delta a_j| = r_{ij} \leq r, \quad r = \max_{i,j}\{r_{ij}\}, \qquad \text{(A17)}$$

Hence

$$\Sigma x_i^{2j} \leq r\, \Sigma x_i^j \leq \Sigma r^2. \qquad \text{(A12'')}$$

It follows from (A4c) that:

$$\Sigma x_i^n < \Sigma x_i^{n-1} < \ldots < \Sigma x_i^2 < \Sigma x_i < \Sigma \equiv M \qquad \text{(A4d)}$$

for the nontrivial case $M > 1$. This can be fulfilled only for $r \geq 1$ and therefore

$$|\Delta a_j| \leq |\epsilon_i|. \qquad \text{(A17')}$$

Taking into account condition $|\epsilon_i| \leq y_i$ as well as (A4c), the final conclusion is that

$$|IPE| \equiv |\Delta a_j| < 1 \qquad \text{(A17'')}$$

Necessity. Let it be $r < 1$, say $r = 1/M$; by (A12'') it follows that

$$\Sigma x_i^{2j} \leq (1/M)\, \Sigma x_i^j \leq (1/M), \text{ i.e. } \Sigma x_i^{2j} \leq (1/M),$$

what is in contradiction with the chain inequalities (A4d). ■

APPENDIX B

Nonlinear Approximation in Standard Representation

A mathematical model $F(x)$ approximating a set of data (x_i, y_i) satisfies the relation

$$y_i \equiv y(x_i) = F(x_i) + \epsilon_i \qquad \text{(B1)}$$

where ϵ_i are some errors appearing due to any possible reason. The model $F(x)$ can be accepted ganerally as a good approximation under the following condition

$$|\epsilon_i| << y_i \qquad \text{(B2a)}$$

e.g.

$$|\epsilon_{max}| < \epsilon_{allowed} \qquad \text{(B2b)}$$

or

Downloaded At: 03:43 30 January 2011

$$q(n) \equiv \Sigma\, \epsilon_i^2(n) \rightarrow \min. \qquad \text{(B2c)}$$

The number of data pairs M is given by (A10) and number of model parameters N by (A11); then one can define the model approximation generally as

(A) *analytical form* if:

$$\epsilon_i \equiv 0, \quad 0 \leq i \leq m \quad \text{and} \quad N < M, \qquad \text{(B3)}$$

(I) *interpolating form* if:

$$\epsilon_i \equiv 0, \quad 0 \leq i \leq m \quad \text{and} \quad N = M, \qquad \text{(B4)}$$

(F) *fitting form* if:

$$\epsilon_i \neq 0, \text{ and } N < M, \qquad \text{(B5)}$$

and

(O) *over-rank* parameter form if:

$$\epsilon_i \neq 0, \text{ and } N \geq M. \qquad \text{(B6)}$$

One can yield the over-rank polynomial parameters only by the so-called Independent Parameter Fitting (IPF) iterative optimization method. The IPF method appears also especially useful in SR, where it is steadyly converging; the property has been proved in Appendix A for the linear model fitting in SR.

SR and Nonlinear Model Approximation

Even if a model for exact solution exists, and is denoted under "analytical form" (B3), this rarely could be found in practice, mainly because of the statistical fluctuations of the data. The functional connection between experimental data variables is almost given by a fitted form of a model which may combine linear and nonlinear part (1).

Depending on the aim, it can be convenient to search either linear or nonlinear model approximation;e.g. pure linear model is convenient for the differentiating and integrating purposes (2); the nonlinear model can extract important investigating constituents (e.g. the Gaussian peaks in X-ray, γ-ray, optical end others spectra analysis (3)) or it can be more efficient in the smoothing - with a smaller number of parameters - than a linear model.

Let it be, in SR, the general mathematical model (including linear and nonlinear part) given by

$$Y_i = f(x_i; b_1, \ldots, b_N), \qquad \text{(B7)}$$

Downloaded At: 03:43 30 January 2011

where N unknown parameters have to be estimated. Let they be ordered by their significance (this means that the same relative change of the b_1 parameter produces more model dislocation than the corresponding change of b_2 parameter, etc.).

We write Taylor series expansion for the model value in the vicinity of the experimental i-point (y_i) at (k + 1)-st step of the iterative procedure as

$$y_i = Y_i^{(k)} + [\partial Y_i^{(k)}/\partial b_1]\Delta b_1^{(k+1)} + \ldots + [\partial Y_i^{(k)}/\partial b_N]\Delta b_N^{(k+1)} + \ldots \qquad \text{(B8)}$$

This linearization brings, in matrix notation, the following system of equations

$$U^{(k)}\Delta B^{(k+1)} = E^{(k)}, \qquad \text{(B9)}$$

and, after normalization (without weighting matrix)

$$\tilde{H}^{(k)}\Delta B^{(k+1)} = \tilde{U}^{(k)}E^{(k)}, \qquad \text{(B10)}$$

where

$$\tilde{H}^{(k)} = \tilde{U}^{(k)}U^{(k)}, \qquad \text{(B11)}$$

where $\Delta B = \{\Delta b_j\}$ is the N-dimensional parameter increment vector to be found,

$U = \{u_{ij} \equiv \partial Y_i/\partial b_j\}$ (i = 0,1,..,m; j = 1,2,..,N)

denotes the (M,N) matrix of the partial N-parameter derivatives for M points;

$\tilde{U}$ indicates the transposed matrix and

$E = \{\epsilon_i \equiv y_i - Y_i\}$ is the M-dimensional error vector.

Further, the fitting is usually realized by the modified (damped) Gauss-Newton method (MGN), as the solution of eq.(B10), i.e.

$$\Delta B^{(k+1)} = p^{(k)}[\tilde{H}^{(k)}]^{-1}\tilde{U}^{(k)}E^{(k)}, \qquad \text{(B12)}$$

where p is an arbitrary constant influencing the speed and convergency of the iterative process; this constant is given in (1) as

$$p = \begin{cases} 1, & \text{if converging,} \\ 0.5p^{(k-1)}, & \text{if diverging.} \end{cases} \qquad \text{(B13)}$$

Downloaded At: 03:43 30 January 2011

Note that the MGN algorithm derived here has the simpler form (omitting the weighting diagonal matrix D (1)); the alternative solution may be used too, if the procedure speed is not important.

We assume the following:

(a) mathematical model chosen suits to the general data pattern;

(b) initial parameter vector $B^{(0)}$ is sufficiently good, and

(c) the model parameters are ordered by their significance.

Under these assumptions the convergency is guaranteed ($p \equiv 1$) and the solution can be found by the IPF method; therefore, we calculate the parameter increments simply by

$$\Delta b_j^{(k+1)} = \{\Sigma(\partial Y_i^{(k)}/\partial b_j)(\in_i^{(k)}/y_i)\} / \{\Sigma(\partial Y_i^{(k)}/\partial b_j)^2/y_i\},$$

$$j = 1,\ldots,N. \qquad \text{(B14)}$$

Note that the algorithm (B14) is the alternative one, i.e. it is given with the $1/y_i$ data point weights.

References

1. Slavić A.I.; Slavić I.J.; Gržetić A.I.; Pavićević K.M. Mikrochim.Acta (1992)[Suppl.] 12: 161-171.

2. Slavić A.I.; Slavić I.J. presented on the 2-nd Intern.Symp of Application of Mathematical Methods and Computers in Geology, Mining and Metallurgy, Belgrade 1991.

3. Slavić A.I. Nucl.Instrum.& Methods (1973), 112, 253-260.

4. Slavić A.I.; Bingulac P.S. Data processing in Experimental Physics (Series of Selected Papers in Physics) Physical Society Japan 1971, pp. 101-108.

5. Strommen P.D.; Nakamoto K. Laboratory Raman Spectroscopy; John Wiley & Sons,Inc: New York 1984.

Date Received: February 14, 1995
Date Accepted: March 17, 1995

Downloaded At: 03:43 30 January 2011