

Available online at www.sciencedirect.com



Chemometrics and intelligent laboratory systems

Chemometrics and Intelligent Laboratory Systems 66 (2003) 127-139

www.elsevier.com/locate/chemolab

Successive Bayesian estimation of reaction rate constants from spectral data

Alexey L. Pomerantsev*

Institute of Chemical Physics of RAS, Polycert Ltd., 4 Kosygin St., 119991 Moscow, Russia

Received 15 May 2001; received in revised form 17 January 2003; accepted 7 February 2003

Abstract

The successive estimation of regression parameters is an effective technique applied in nonlinear regression analysis. In the paper, this method is used for obtaining kinetic information from spectral data without any pure component spectra. In order to study the way kinetics may be determined by 'hard' model without separate calibration, the simulated example of a two-step reaction was constructed. With the help of real-world example, this approach is compared with known methods of kinetic modelling. All calculations have been performed with the help of new nonlinear regression software, Fitter. Such approach is valuable in case the concentrations of pure components are not available or if 'soft' calibration methods are inexact or time consuming.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Kinetic parameter estimation; Successive Bayesian estimation; Spectroscopy; Fitter; Nonlinear regression

1. Introduction

Nowadays, spectral data are often used to determine the kinetic parameters. Computerized spectroscopy provided us with a rapid on-line method of measurements. These spectra contain relevant information of the kinetics, thus challenging us to develop a complex method for obtaining that information from the spectral data. There are many methods available to solve this problem [1-7] where the employed approach consists of performing separate "soft" calibration between spectra and concentration units. The calibrated concentration values are used further for the "hard" parameters' fitting. Sometimes,

"soft" methods fail to provide the proper accuracy of results. On the other hand, "hard" kinetic models built on the main principles of chemical physics give the strong basement for data analysis. Such methods can be easily applied when spectra of pure components are known. However, when they are unknown, some essential problems arise. In this work, we suggest a new way of "hard" analysis of spectral data, which is suitable even in cases when pure component spectra are unknown. Successive Bayesian estimation (SBE) technique [8] is known to be helpful for complex large dimension nonlinear regression systems [9]. This method converts the whole problem to a sequence of regressions that have smaller dimensions. Now it is implemented for estimation of reaction rate constants from spectral data, and the results are reported here.

^{*} Tel.: +7-95-939-7483; fax: +7-95-283-4757.

E-mail address: polycert@chph.ras.ru (A.L. Pomerantsev).

There are two main aims in the paper. The first one is to show the feasibility of SBE in application to the spectral data. The second aim is to demonstrate that this algorithm can be easily carried out with the help of Fitter software [9,10], designed for nonlinear regression analysis. To show all the advantages of the suggested method, both simulated and real-world data are used. Simulated data let us compare all predicted values with known 'true' values, and thus validation of the proposed SBE method becomes absolutely clear. Real-world data let us compare this approach with well-known methods of kinetic constants estimation.

2. Experimental. Simulated data

The model for the kinetics of spectral data may be expressed as a function of time t and wavelength xdepending on unknown kinetic rate constants k

$$y(t, x, \boldsymbol{k}) = \sum_{i=1}^{l} c_i(t, \boldsymbol{k}) p_i(x).$$
(1)

Here, y is a spectral signal, c_i are component concentrations, p_i are pure component spectra, and l is the number of reaction components. In discrete case, when the spectra are separated into m wavelengths and time is presented by n points, the matrix notation for this equation can be used

$$Y = CP + E. \tag{2}$$

Here, Y is the $(n \times m)$ measured data matrix, C is the $(n \times l)$ concentration matrix depending on unknown kinetic parameters, and P is the $(l \times m)$ unknown matrix of pure component spectra. Besides, the $(n \times m)$ error matrix E is involved in model (2). Of course, Eq. (1) is valid only in ideal cases. In practice, the spectra contain drifts, baseline errors, interactions between species, etc. Sometimes, these errors can be corrected by suitable spectral pretreatment methods, but not always. In general, the SBE method does not utilize the linearity in Eq. (2) and therefore may be applied to more complicated non-ideal models.

2.1. Concentration matrix C

This matrix can be obtained as a solution to a kinetic model. In this work, we consider an example of two-step kinetics

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C, \tag{3}$$

which may be presented as a system of ordinary differential equations

$$\frac{\mathrm{d}A}{\mathrm{d}t} = -k_1 A; \qquad A(0) = A_0$$

$$\frac{\mathrm{d}B}{\mathrm{d}t} = k_1 A - k_2 B; \qquad B(0) = B_0$$

$$\frac{\mathrm{d}C}{\mathrm{d}t} = k_2 B; \qquad C(0) = C_0$$
(4)

This system has analytical solution

$$A = A_0 \exp(-k_1 t)$$

$$B = \frac{k_1 A_0}{k_1 - k_2} [\exp(-k_2 t) - \exp(-k_1 t)] + B_0 \exp(-k_2 t)$$

$$C = A_0 + B_0 + C_0 + \frac{A_0}{k_1 - k_2} [k_2 \exp(-k_1 t) - k_1 \exp(-k_2 t)] - B_0 \exp(-k_2 t)$$
(5)

Here we use the same notation for reaction components A, B, C as well as for their concentration values [A]=A, [B]=B, [C]=C. This notation is not accurate but easier to write.

For the simulation study, the data are created by Eq. (5) for the following initial concentration values

$$A_0 = 1, \quad B_0 = C_0 = 0$$

The 'true' rate constant values are chosen as

$$k_1 = 1, \quad k_2 = 0.5$$

and the sampling instances, i.e., the time points when the spectra are 'measured', are taken as

$$t = 0, 2, 4, 6, 8, 10.$$

Thus, number of the time points (n) is 6. Such small number of points is chosen with the intention to produce the maximum troubles in the estimation. It is the minimum number of measurements when esti-



Fig. 1. Simulated concentration profiles.

mation of parameters from single wavelength kinetics is possible. Corresponding kinetic curves are plotted in Fig. 1.

2.2. Pure component spectra matrix P

This matrix is simulated in an ordinary way using Gaussian overlapping spectral peaks. Each spectrum p is normalized as max(p) = 1. These spectral curves for all pure components of reactions (Eq. (3)) are plotted in Fig. 2. Here are 53 (m) wavelengths. The actual wavelength values x will never be used in the simulated example; therefore, only conventional wavelength values are denoted. These values are just the numbers that correspond to the actual wavelength values. They vary from 1 up to 53.

We consider that these spectra are rather complex. They are constructed for the purpose of demonstrating the possibilities of the proposed method even for complicated spectral data. In this simulated example,



Fig. 2. Simulated pure component spectra.



Fig. 3. Simulated true kinetic spectra (curves) and corresponding data with errors (points). t=0 (1, \blacklozenge), t=2 (2, \blacksquare), t=4 (3, \blacktriangle), t=6 (4, \blacklozenge), t=8 (5, \bigcirc), t=10 (6, \square).

all spectra values are nonnegative and we use this knowledge in processing.

2.3. Data matrix Y

This matrix is calculated by Eq. (2), where concentration matrix C is given by system Eq. (5). In addition, white noise with a relative error of 3% is added to the simulated data.

Corresponding values are plotted in Fig. 3, where curves represent errorless values and points present 'measured' data calculated with random errors.

3. Method. Successive Bayesian estimation

Our goal is to find unknown kinetic constants $k=(k_1, k_2)^t$ from spectral data *Y*, which is the (6×53) matrix. If pure component spectra vectors *p*, *q*, and *r*, which correspond to the components *A*, *B*, and *C* accordingly, are known, we obtain a rather simple OLS problem [11]—to minimize the sum of the squares—

$$\min_{\mathbf{k}} \sum_{i=1}^{m} \sum_{j=1}^{n} [Y_{ij} - p_i A(t_i, \mathbf{k}) - q_i B(t_i, \mathbf{k}) - r_i C(t_i, \mathbf{k})]^2,$$
(6)

where functions A, B, and C are presented by Eq. (5) and p_i , q_i , and r_i are given values. However, if one or more pure spectra vectors p, q, and r are unknown—which is the usual case—the situation changes dramatically. Practically, it is very difficult to find the

minimum of the sum (Eq. (6)) with respect to 161 unknown parameters (2 stands for kinetics, plus 53 spectrum values at each of the 3 pure components) owing to the problem with ill-posed matrices.

As model (2) is linear in spectrum parameters P, they may be excluded from sum (Eq. (6)), so the initial estimating problem reduces to

 $\min \|\boldsymbol{Y} - \boldsymbol{C}(\boldsymbol{C}^{\mathsf{t}}\boldsymbol{C})^{-1}\boldsymbol{C}^{\mathsf{t}}\boldsymbol{Y}\|^{2}.$

Further improvement of this approach [5,6] leads to the weighted curve resolution method (WCR), which is discussed below.

Applying the SBE method, we do not exclude spectrum parameters from problem (Eq. (6)), but reduce its dimension in the other way. The full description of this technique may be found in Refs. [8,9] and also in Appendix A, where some essential formulae are presented. Here, just a brief summary is given. The main concept of SBE is to split the whole data set into several parts. Afterwards, estimation of parameters is performed successively-fraction by fraction-with maximum likelihood method (MLM). It is important that results obtained on the previous step are used as a priori values (in the Bayesian form) for the next part. The first (initial) fraction is processed by the OLS method without any a priori information. During this procedure, the sequence of the parameter estimates is produced and its last term is the ultimate estimate. For linear regression, it was shown [8] that SBE gives the same estimates as the traditional OLS approach, and these values do not depend on the order of fractions (see theorem in Appendix A). In nonlinear case, the situation is more difficult but all these properties are asymptotically the same.

With respect to the problem in question, the following algorithm presents the SBE method for kinetic parameters k_1 and k_2 estimation.

Step 0 (Initiate stage). Several wavelengths x_1 , x_2 ,... (usually 3–6) are selected and corresponding spectral data sets Y_1 , Y_2 ,... are processed simultaneously by the OLS method (see Eqs. (9)–(12)).

Step 1. The results—kinetic parameter estimates, covariance matrix, etc.—are transformed into a priori information in accordance with Eqs. (28)-(31).

Step 2. A new single wavelength x_i is selected and corresponding spectral data set Y_i is processed separately by MLM with respect to the a priori information, obtained at Step 1 (see Table 1).

Last Step. Steps 1 and 2 are repeated until all wavelengths are treated, and we obtain the ultimate results.

According to Appendix A, here we use a priori information of *first type* as we know that the error variance is the same at each wavelength. In addition, parameters k_1 and k_2 are treated as common for all wavelengths, while parameters p_i , q_i , and r_i are considered as partial.

4. Results. Simulated data

4.1. Problem of first step

The SBE algorithm needs an initiate stage, where the kinetic data are processed without a priori information. Sometimes this stage may bring difficulties. In our example, fitting of single wavelength data set is troublesome because those data have only six measurements versus five unknown parameters. Only some wavelengths are applicable, e.g., conventional wavelength 16 (see Fig. 4).

However, we can get together several wavelengths to obtain the initial estimates. In the investigated example, it appears that the kinetic data sets of any four wavelengths can solve the problem of the first step. These data contain 24 measurements that are fitted by 14 unknown parameters. Initial wavelengths may be chosen arbitrarily. They could be the first, or the last, or any random four wavelengths.

In Fig. 4, an example of the initial data is presented. These kinetic curves are used in the random



Fig. 4. Kinetic data used at the initiate stage with a random order of conventional wavelengths: x=5 (1, \diamond), x=8 (2, \Box), x=16 (3, \triangle), x=25 (4, \bigcirc).

successive procedure described in Section 4.2. The initial a priori information is designed using the results of this stage, and thereupon the SBE procedure begins.

4.2. Kinetic rate constants estimation

In general, it is known that the order of series in the successive algorithm affects the results of estimation of nonlinear model. However, this effect is not essential for practical purposes. To demonstrate this, we performed the successive estimation for different orders of conventional wavelengths. These are the direct order (i.e., $1, 2, 3, 4, 5, \ldots, 53$), the inverse order (i.e., $16, 5, 29, 8, 41, \ldots$). The leading four numbers in these sequences present the wavelengths that are used

at the initiate stage. The results are shown in Fig. 5, plots (a-c).

Here, the solid curves represent the successive estimates of the kinetic rate constants, and the shaded areas represent the standard deviations added (sub-tracted) to them to show the uncertainties. Each plot contains both estimates; the upper is k_1 estimate and the lower is k_2 estimate. The dashed lines show the 'true' values of the kinetic constants. All the data are plotted versus numbers (the X-axis), which represent conventional wavelengths in the order (from left to right) as they are used in the successive procedure. The first four points in each plot show the results of the estimation at the initial stage.

The plots demonstrate that while different orders of wavelengths produce different intermediate values of estimates, the ultimate results are rather close. The



Fig. 5. Results of the successive estimation of kinetic rate constants depending on wavelength order: (a-c) and the 0.95 confidence ellipses of the final estimates (d). In plots (a-c), solid curves present estimates of k_1 (1) and k_2 (2); thin curves (1a, 1b, 2a, 2b) show borders of standard deviation regions; dashed lines (1c, 2c) mark the 'true' values of constants. In plot (d), ellipses and points present: direct $(1, \blacksquare)$, inverse $(2, \diamondsuit)$, and random (3, O) order of wavelengths; point $(4, \blacktriangle)$ shows the 'true' values of constants.

last plot (d) in Fig. 5 illustrates this idea. Here are the 0.95 confidence ellipses for all the ultimate estimates. Each ellipse and the mark in its center present the results of SBE with a corresponding order of wavelengths. Comparing the trajectories for the different arranges, one can see that the intermediate estimates, as well as their uncertainties, really depend on the wavelength order. It seems that the best is the direct order, where both estimates and deviations vary rather smoothly and slowly without large jumps. Another interesting result can be seen in plot (b), where estimation with the inverse order of wavelengths is presented. The initial stage yields very bad estimate of k_1 parameter. Afterwards, uncertainty falls down. However, these values are far from the concluding results yet. Just the last wavelengths 2 and 1 make the estimate and the deviation closer to the common ultimate values.

4.3. Spectral parameters estimation

As soon as the common kinetic rate constants k_1 and k_2 have been estimated, it is natural to find partial spectral parameters p, q, and r. Of course, it can be done rather easily. If the rate constants are fixed on their estimated values, the set of spectral parameters p_i, q_i , and r_i can be obtained for each wavelength i by the OLS method. However, the uncertainties of these estimates will be calculated wrong because the OLS approach cannot take into account the uncertainties of fixed kinetic constants.

We suggest applying the SBE approach again. At the last step of the kinetic constants estimation, the resulting a posteriori information can be created and then transformed into the truncated a priori information (see Eq. (28)), which concerns the kinetic constants only. This should be used as a priori information of the second type (see Appendix A) for each wavelength *i* when the set of spectral parameters p_i , q_i , and r_i is estimated. Of course estimating should be performed under the additional constrains that kinetic parameters have already been found and therefore they are not estimated again. In practice, we assign the zero values to the elements of matrix V (Eq. (10)), which correspond to kinetic parameters.

The results are presented in Fig. 6. Each plot shows the 'true' simulated pure component spectra (solid curve 1, left Y-axis). The results of the estimation are

11 21 31 41 conventional wavelength 1.2 0.6 c) Spectrum C 1 0.4 0.8 Iccurac 2a 0.2 0.6 0.4 0 and 0.2 0 -02 1 11 21 31 41 51 conventional wavelength Fig. 6. Simulated pure component spectra (solid curves, 1, left Y-

axis). Difference between spectrum estimate and 'true' spectra (point 2, right Y-axis). Tripled standard deviation regions (shaded areas 2a, 2b, right Y-axis).

presented as the difference between the estimated and the 'true' spectra (point 2). Also, the tripled standard deviation regions (shaded areas 2a and 2b) are shown. These data are plotted to the right Y-axis. It can be seen that obtained estimates are rather accurate, particularly for components A and C.

iccuracy

accuracy



5. Discussion. Simulated data

The proposed method provides a new approach to the identification of chemical kinetic models. To demonstrate that this method is reliable, we perform the following 'validation test'. The simulated 'experimental' data are processed by the OLS method, where 161 parameters are estimated by minimizing the sum of squares (Eq. (6)). The results of the test are presented in Fig. 7. The points in the centers of ellipsis show the estimates obtained with the help of the OLS and the SBE methods. They are very close as well as the corresponding confidence regions, which are represented by the ellipses. Thus, the test confirms the conclusions of the theorem (Appendix A); the SBE estimates are asymptotically the OLS estimates. However, we are not sure in the confidence regions. It is known that covariance analysis of nonlinear models is a hard problem. The linearization is valid only if the model is 'similar' to the linear model. A measure of this similarity is a coefficient of nonlinearity that was proposed in Ref. [9].

The coefficient for the linear model is equal to 1. The greater the coefficient, the more nonlinear the model is. This coefficient is calculated for the given example and it appears to be 1. This result shows that this model is actually 'linear-like' so one can trust in the constructed ellipses. The statistical simulation technique [12] sustains this claim.

The initial stage is the main trouble of the proposed method. In Fig. 5, one can see that the successful



Fig. 7. Results of estimation presented by the 0.95 confidence ellipses SBE $(1,\blacksquare)$ and OLS $(2,\bullet)$ methods; true value $(3,\blacktriangle)$.



Fig. 8. Results of successive estimation of kinetic rate constants for the 'optimal' order of wavelengths. Solid curves present estimates of k_1 (1) and k_2 (2); thin curves (1a, 1b, 2a, 2b) show borders of doubled standard deviation regions; dotted lines (1c, 2c) mark the 'true' values of constants.

choice of initial wavelengths can improve the estimation (plot a) as well as a bad choice could decline it (plot b). To make the choice of series order automatic, we suggest a simple instrument.

It is clear that the kinetic curve with the largest change is the most 'informative'. In addition, the nonmonotonic curves with extremes are good for fitting. Summarizing these ideas, we introduce an empirical criterion for evaluating the relative 'informativeness' of the kinetic curves. This is the following expression

$$L = \sum_{j=1}^{n} \sqrt{\left(t_j - t_{j-1}\right)^2 + \left(Y_j - Y_{j-1}\right)^2} - \left(t_m - t_1\right)$$

where t_j are the time (predictor) values, Y_j are the kinetic (response) values, and *n* is the number of measurements. The main term in this expression is the length of the curve. The larger the value of *L*, the more 'informative' the kinetics is. The straight line that is parallel to the *t*-axis has the 'informativeness' L=0. Curves presented in Fig. 4 have the following values: L(16)=0.0476, L(5)=0.0286, L(29)=0.0174, L(8)=0.0007. If we range all wavelengths in accordance with this criterion, we can expect the 'optimal' successive estimation. Fig. 8 sustains this idea.

6. Real-world example

6.1. Data

This example originates from Refs. [2] and [13]. The data set consists of SW-NIR spectra of 2,5-di-*tert*butyl-1,4-benzoquinone. The reaction of two-step epoxidation was studied as an example of real experimental process. The reagents, experimental setup, and other details have been described in Ref. [2]. Raw experimental data may be found at Ref. [15]. There are 240 spectra with wavelength range 800– 1100 nm and interval of 1.0 nm. Total reaction time is 1200 s. The eight repeated individual batch processes were recorded.

The raw spectral data are preprocessed in accordance with a procedure described in Ref. [2]. Bellow, a short summary is given. The fourth spectrum is used as blank and is subtracted from all other spectra. The second derivative spectra are calculated applying a Savitzky–Golay filter with 15 data points window. The small wavelength range 860–880 nm is used for data processing.

6.2. Results

We have performed the processing of first data batch with the direct and optimal successive procedure. In this example, 'spectral' parameters p, q, and rcould be negative, thus no constrains are applied. The results of direct procedure are presented in Fig. 9, where the successive estimates of the rate constants are



Fig. 9. Results for real-world example obtained with the direct SBE algorithm. Solid curves present estimates of k_1 (1) and k_2 (2); shaded areas (1a, 1b, 2a, 2b) show borders of standard deviation regions.



Fig. 10. Results for real-world example obtained with different methods. Each method is presented by the 0.95 confidence ellipse: CCR (1), WCR (2), GRAM (4), LM-PAR (3), and SBE (5).

plotted versus wavelengths. The setup of this plot is similar to Fig. 5. The final values of estimates, obtained on the last wavelength 880 nm, are $k_1 =$ 0.267 ± 0.015 (min⁻¹) and $k_2 = 0.095 \pm 0.010$ (min⁻¹). Here, standard deviations are also given. Correlation coefficient is r = -0.18. The optimal SBE procedure gives the following values: $k_1 =$ 0.238 ± 0.015 (min⁻¹), $k_2 = 0.102 \pm 0.010$ (min⁻¹), and r = -0.22.

In order to compare the successive method with other known approaches to kinetic constants estimation, we have constructed Fig. 10. There, the results of processing of this example with different methods are shown. Each method is shown by its 0.95 confidence ellipse that was constructed using data presented in Ref. [13]. All these methods are completely described in Ref. [13], so only a brief summary is given here.

Classical curve resolution (CCR—ellipse 1) is the successive two-step algorithm that utilizes the linearity of spectral parameters. This technique also gives the OLS-solution as no constraints are used. Weighted curve resolution (WCR—2) is a combination of a 'soft' method using singular value decomposition of data matrix Y (Eq. (2)) and a 'hard' presentation of concentration matrix C. The generalized rank annihilation method (GRAM—3) is a 'soft' approach that utilizes a simple equation

$$\frac{\mathrm{e}^{-kt}}{\mathrm{e}^{-k(t+s)}} = \mathrm{e}^{kt}$$

showing that the reaction rate constants can be extracted from the ratio of the nonshifted and shifted exponentially decaying function. The Levenberg–Marquardt algorithm and alternating least squares steps of the PARAFAC model (LM-PAR—4) are a 'soft' method using iterative procedure that enhances GRAM estimates. Both GRAM and LM-PAR method can only be used for modelling pseudo first-order kinetics.

The data for ellipse 5 (SBE), which corresponds to the optimal SBE solution, have been given above. One can compare the size and the shape of the ellipses. From Fig. 10, it can be seen that ellipses 1 and 5 have the smallest size and roundest shape. This means that the SBE method yields minimum deviations and correlation among the observed methods, and its results are again close to the OLS solution. The optimality of the OLS estimators is well known [11], although this method could be unreliable for the large data sets. The properties of the SBE method (see Appendix A) are similar to OLS but it is more stable.

7. Conclusions

In this paper, the successive Bayesian estimation method has been presented in order to estimate reaction rate constants from spectral measurements of a reaction system when individual component spectra are unknown. This method is of general nature and it can be used for any kind of kinetic models. Its feasibility was illustrated by the simulated and the real-world example of two-step kinetics. Simulations show that this algorithm can deal with a strong spectral overlap and with an extremely small number of time points. It was also demonstrated that SBE agrees with OLS. From the real-world example, it can be concluded that the successive method leads to lower deviations and correlations of reaction rate constants estimates in comparison with some known methods. Moreover, it gives additional information that is the trajectories of estimates in dependence on wavelength, which can be used for a supplementary study of data.

The SBE method is rather fast, but it is slower than OLS. The ratio of computational times $t_{\text{SBE}}/t_{\text{OLS}}$ may be roughly estimated [14] by the formula $[1+(p/l)]^3$, where *p* is the number of kinetic constants and *l* is the

number of components. For a two-step reaction, the ratio is equal to 4.6; however, in our examples, the actual values were about 2.

It is of interest that the SBE method is Bayesian only in its form but not in its concept. This means that no subjective a priori data are actually used. Each a priori information element yields from the experimental data processed at the previous step, and only the way of its application is dictated by the Bayes theorem. No extra assumptions (number of PCs, time-shift, pseudo first-order) are needed for this method implementation.

8. Software implementation

Fitter software [10] was used for modelling. Derivation of spectra was done in the Unsrambler version 6.11 [16].

Acknowledgements

Oxana Rodionova (Institute of Chemical Physics, Moscow) is thanked for stimulating discussions during the preparation of this paper. The author would also like to thank Sabina Bijlsma (Nutrition and Food Research Institute, Zeist) for her assistance in the realworld example.

Appendix A. The SBE Theory

The usual regression setup is considered

$$y_i = f(\mathbf{x}_i, \mathbf{a}) + \varepsilon_i, \qquad i = 1, \dots, N, \ \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}),$$

where $a=(a_1,...,a_p)^t$ is the vector of unknown parameters and σ^2 is the unknown error variance. The proper likelihood function (argument y is omitted) is given by

$$L_0(\boldsymbol{a},\sigma^2) = (2\pi)^{-N/2} \sigma^{-N} \exp\left(-\frac{S(\boldsymbol{a})}{2\sigma^2}\right), \qquad (7$$

where

$$S(a) = \sum_{i=1}^{N} (y_i - f_i)^2$$
(8)

is the sum of squares. Applying maximum likelihood principle to the function Eq. (7), one can find the following values.

Parameters a estimates-

$$\hat{\boldsymbol{a}} = \arg\min S(\boldsymbol{a}) \tag{9}$$

and the Fisher matrix A that characterizes the accuracy of estimation—

$$A = V^{t}V, \quad \text{where } V_{\alpha i} = \frac{\partial f(x_{i}, \hat{a})}{\partial a_{\alpha}},$$

$$\alpha = 1, \dots, p; \ i = 1, \dots, N$$
(10)

This $(p \times p)$ matrix A is the Hessian of S(a) in the Gauss–Newton approximation [9], and if it is invertible, then $cov(\hat{a}, \hat{a}) = \sigma^2 A^{-1}$.

Variance σ^2 estimate is

$$s^2 = \frac{S(\hat{a})}{N_{\rm f}} \tag{11}$$

where $N_{\rm f}$ is the number of degrees of freedom (NDF) for estimator Eq. (11)—

$$N_{\rm f} = N - p \tag{12}$$

Near the point of maximum, the likelihood function can be approximated by

$$L_0(\boldsymbol{a}, \sigma^2) \approx (2\pi)^{-N/2} \sigma^{-N} \times \exp\left[-\frac{s^2}{2\sigma^2} \left((\boldsymbol{a} - \hat{\boldsymbol{a}})^{\mathrm{t}} \boldsymbol{A}(\boldsymbol{a} - \hat{\boldsymbol{a}}) + N\right)\right],$$
(13)

Now, let us consider the case when there is a priori information presented by some distribution $h(a,\sigma^2)$. Then the likelihood function is

$$L(\boldsymbol{a},\sigma^2) = h(\boldsymbol{a},\sigma^2)L_0(\boldsymbol{a},\sigma^2)$$
(14)

Our aim is to construct the distribution h basing on the available information that in fact consists of the following items, which correspond to the values in Eqs. (9)–(12).

(1) A priori parameter values

$$\boldsymbol{b} = (b_1, \dots, b_p)^{\mathsf{t}} \tag{15}$$

(2) A priori information matrix

$$\boldsymbol{H} = \{h_{\alpha\beta}\}, \qquad \alpha, \beta = 1, \dots, p \tag{16}$$

(3) A priori variance value

$$s_0^2$$
 (17)

(4) A priori value of NDF

$$N_0$$
 (18)

A priori information that includes all four items is referenced as of type 1. Sometimes, the two last items, concerning variance and NDF, are absent. In this case, information is called type 2.

In the Bayesian approach, *a* can be considered (see Eq. (13)) as a normal random vector with the expectation *b* and the accuracy matrix γH

$$\boldsymbol{a} \sim N(\boldsymbol{b}, \gamma \boldsymbol{H}) = \frac{\sqrt{\gamma^{p}}}{\sqrt{(2\pi)^{p}}} \sqrt{\det \boldsymbol{H}} \exp\left[-\frac{\gamma}{2}R(\boldsymbol{a})\right]$$
(19)

where R(a) is a quadratic form

$$R(\boldsymbol{a}) = (\boldsymbol{a} - \boldsymbol{b})^{\mathrm{t}} \boldsymbol{H}(\boldsymbol{a} - \boldsymbol{b}), \qquad (20)$$

Factor γ depends on the type of information, and it is $\gamma = s_0^2/\sigma^2$ for type 1 and $\gamma = 1$ for type 2.

A priori distribution of variance σ^2 can be written using the χ^2 approach and values in Eqs. (17) and (18)

$$\sigma^{2} \sim \frac{(2N_{0}s_{0}^{2})^{\frac{N_{0}}{2}}}{\Gamma\left(\frac{N_{0}}{2}\right)} \sigma^{-N_{0}-2} \exp\left(-N_{0}\frac{s_{0}^{2}}{2\sigma^{2}}\right).$$
(21)

Combining Eqs. (19) and (21) with Eq. (14), the likelihood function with a priori information of type 1 can be presented as

$$L(\boldsymbol{a}, \sigma^{2}) = C_{1} \sigma^{-N-N_{0}-2} \\ \times \exp\left[-\frac{1}{2\sigma^{2}} \left(S(\boldsymbol{a}) + s_{0}^{2} R(\boldsymbol{a}) + s_{0}^{2} N_{0}\right)\right],$$
(22)

Table 1

and with a priori information of type 2 by

$$L(\boldsymbol{a},\sigma^2) = C_2 \sigma^{-N} \exp\left[-\frac{1}{2}\left(\frac{S(\boldsymbol{a})}{\sigma^2} + R(\boldsymbol{a})\right)\right].$$
(23)

Factors C_1 and C_2 are out of interest because they do not depend on a and σ^2 .

The MLM estimate is the point where likelihood function has a maximum. It is easy to see that this can be found as a minimum of objective function Q(a). Differentiating Eqs. (22) and (23), one can obtain that for information of type 1, the objective function is

$$Q(\boldsymbol{a}) = S(\boldsymbol{a}) + B(\boldsymbol{a}) \tag{24}$$

and for type 2, it is

$$Q(\boldsymbol{a}) = S(\boldsymbol{a})B(\boldsymbol{a}) \tag{25}$$

Here, B(a) is the Bayesian term of the form

$$B(a) = s_0^2 [N_0 + R(a)]$$
(26)

for information of type 1, and of the form

$$B(\boldsymbol{a}) = \exp\left[\frac{R(\boldsymbol{a})}{N}\right]$$
(27)

for information of type 2. In Table 1, some essential statistics of MLM with a priori information are presented.

In SBE, the whole data set is split into the parts, which are processed successively. At each step (except the first one), MLM is applied with a priori information that originates from the results of the previous step. Now we show how that information is calculated.

Let us consider the results, which are obtained after processing on the *i*th step (see Table 1). Obviously they can be referenced as a posteriori information. This information is similar to the items presented in Eqs. (15)-(18), and it could naturally be used as a priori information for the next (i + 1)th step. However, there are some difficulties that require the additional consideration.

Results of MLM estimation with a priori information of type 1 and type 2

J I -		
Statistics	A priori information of type 1	A priori information of type 2
Parameter estimators The Fisher matrix	$\hat{a} = \arg \min Q(a),$ Q(a) = S(a) + B(a) $A = V^{t}V + s_{0}^{2}H$	$\hat{a} = \arg \min Q(a),$ Q(a) = S(a)B(a) $A = \exp\left(\frac{R(\hat{a})}{N}\right)$ $\times \left(V^{t}V + \frac{S(\hat{a})}{N}H\right)$
Variance estimator	$s^2 = rac{Q(\hat{a})}{N_{ m f}}$	$s^2 = \frac{S(\hat{a})}{N_{\rm f}}$
NDF	$N_{\rm f} = N + N_0$	$N_{\rm f} = N$

It often occurs that each part of split data set is fitted by its own regression function $f_i(\mathbf{x}, \mathbf{a}_i)$, which depends both on *common* parameters and on *partial* parameters

$$\boldsymbol{a}_i = (a_1, \ldots, a_r, a_{r+1}, \ldots, a_{p_i})^{\mathsf{t}}$$

Parameter subset a_1, \ldots, a_r is called common because each regression function f_i , $i = 1, \ldots, M$, depends on it and the parameter subsets a_{r+1}, \ldots, a_{p_i} are called partial because each subset presents just in the single model f_i .

When a priori information is constructed from a posteriori one, it is essential to separate data relating to the common and to the partial parameters. Common information should be kept for further usage, while partial one should be removed since it does not coincide to the next portion of the data. The a posteriori Fisher matrix A (step index i is omitted for simplicity) can be represented by a block matrix

$$\mathbf{A} = egin{bmatrix} A_{00} & A_{0I} \ & & \ A_{0I} & A_{II} \end{bmatrix},$$

where A_{00} is the $(r \times r)$ square matrix corresponding to the common parameters, A_{11} is the $(p_i - r) \times (p_i - r)$ square matrix corresponding to the partial parameters, and A_{01} is the $r \times (p_i - r)$ matrix. The a priori information matrix H is recalculated from matrix A by the formula

$$\boldsymbol{H} = \frac{1}{s^2} \begin{bmatrix} \boldsymbol{A}_{00} - \boldsymbol{A}_{01} \boldsymbol{A}_{11}^{-1} \boldsymbol{A}_{01}^{\dagger} & \boldsymbol{0} \\ & & \\ \boldsymbol{0} & & \boldsymbol{0} \end{bmatrix}$$
(28)

where s^2 is the a posteriori value of the error variance. Matrix dimension should correspond to the number of parameters in the next portion of data, i.e., $(p_{i+1}) \times (p_{i+1})$, so the matrix is completed with zero values. A priori parameter values are transformed in parallel

$$b_{\alpha} = \begin{cases} \hat{a}_{\alpha}, & 0 < \alpha \le r \\ 0, & r < \alpha \le p_{i+1} \end{cases}$$
(29)

For information of type 1, a priori variance value is equal to a posteriori value

$$s_0^2 = s^2,$$
 (30)

but NDF should be recalculated by

$$N_{\rm f} = N_i - N_0 - p_i + r, \tag{31}$$

where N_i is the number of data and N_0 is a posterior NDF value in the *i*th portion of data. Now, Eqs. (28)–(31) present the a priori information that is applied on the next step of the SBE procedure.

Let us compare SBE and OLS methods. In OLS, the objective function

$$S(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_M)=S_1(\boldsymbol{a}_1)+\ldots+S_M(\boldsymbol{a}_M) \tag{32}$$

is used. Here

$$S_j(\boldsymbol{a}_j) = \sum_{i=1}^{N_j} (y_{ji} - f_j(\boldsymbol{x}_{ji}, \boldsymbol{a}_j))^2, \quad j = 1, \dots, M$$
 (33)

is the partial sum of squares regarding to the *j*th portion of data.

The OLS estimates of parameters a are

$$\hat{\boldsymbol{a}} = (\hat{\boldsymbol{a}}_1, \dots, \hat{\boldsymbol{a}}_M) = \arg\min S(\boldsymbol{a}_1, \dots, \boldsymbol{a}_M),$$
 (34)

and the estimate of error variance is

$$s^2 = \frac{S(\hat{a}_1, \dots, \hat{a}_M)}{N - p},$$
 (35)

Here, $N = N_1 + \ldots + N_M$ is the full number of all measurements, and $p = r + p_1 + \ldots + p_M$ is the full number of all parameters *a*.

The SBE estimates are presented by the following algorithm.

- (1) The whole data set is split into M subsets (parts).
- (2) The first part is processed by OLS method (see Eqs. (8)-(10)).
- (3) A posteriori information is transformed to a priori one (see Eqs. (28)–(31)).
- (4) The next part is processed by MLM with a priori information (see Table 1).
- (5) Steps 3 and 4 are repeated until the last part of data.
- (6) The ultimate results are the SBE estimates.

In general, the SBE estimates depend on the subsets order, which is applied in processing. However, in the linear case, the following properties of the SBE method [8] can be proved.

Theorem. Let functions $f_j(\mathbf{x}, \mathbf{a}_j)$, j = 1, ..., M be linear in parameters \mathbf{a}_j , and errors be homoscedastic, i.e., $\sigma_1^2 = \ldots = \sigma_M^2 = \sigma^2$, then the following OLS and SBE outcomes coincide for any order of processing in the SBE method:

- (1) estimators of common parameters;
- (2) covariance matrices of common parameters;
- (3) estimators of error variance;
- (4) numbers of degrees of freedom.

It can also be proved that in the nonlinear case, these properties are achieved asymptotically.

References

- H. Haario, V.-M. Taavitsainen, Chemometr. Intell. Lab. Syst. 44 (1998) 77–98.
- [2] S. Bijlsma, D.J. Louwerse, W. Windig, A.K. Smilde, Anal. Chim. Acta 376 (1998) 335–339.
- [3] S. Bijlsma, D.J. Louwerse, A.K. Smilde, J. Chemom. 13 (1999) 311–329.
- [4] V.-M. Taavitsainen, H. Haario, J. Chemom. 15 (2001) 215–239.
- [5] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Chemometr. Intell. Lab. Syst. 54 (2000) 123–141.

- [6] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Anal. Chim. Acta 442 (2001) 337–350.
- [7] E. Bezemer, S.C. Rutan, Chemometr. Intell. Lab. Syst. 59 (2001) 19-31.
- [8] G.A. Maksimova, A.L. Pomerantsev, Ind. Lab. 61 (1995) 432–435.
- [9] E.V. Bystritskaya, A.L. Pomerantsev, O.Ye. Rodionova, J. Chemom. 14 (2000) 667–692.
- [10] Fitter AddIn [Online]. http://polycert.chph.ras.ru/fitter.htm (1 June 2000).
- [11] Y. Bard, Nonlinear Parameter Estimation, Academic Press, New York, 1974.
- [12] A.L. Pomerantsev, Chemometr. Intell. Lab. Syst. 49 (1999) 41-48.
- [13] S. Bijlsma, A.K. Smilde, J. Chemom. 14 (2000) 541-560.
- [14] R. Rajko, Anal. Lett. 27 (1) (1994) 215-228.
- [15] Datasets [Online]. http://www-its.chem.uva.nl/research/pac (1 December 2000).
- [16] Unscramber [Online]. http://www.camo.no (1 December 2000).