Kinetics and Catalysis, Vol. 45, No. 4, 2004, pp. 455–466. Translated from Kinetika i Kataliz, Vol. 45, No. 4, 2004, pp. 485–497.

On One Method of Parameter Estimation in Chemical Kinetics Using Spectra with Unknown Spectral Components

O. E. Rodionova and A. L. Pomerantsev

Semenov Institute of Chemical Physics, Russian Academy of Sciences, Moscow, Russia

Received January 20, 2003

Abstract—The method of successive estimation of regression parameters, which is widely used in nonlinear regression analysis, is applied to obtain kinetic information from spectral data for the case when the spectra of individual components are unknown. Using a model example with a two-step successive reaction, the reliability of the algorithm is demonstrated. To compare the proposed method with other known methods for estimating kinetic parameters literature data are used. All simulations were done using a new software for nonlinear regression analysis: FITTER. The proposed approach is especially useful when the spectra of reaction components are unknown and when formal calibration methods do not provide desirable accuracy.

A mechanistic study of a complex chemical process is one of the most important tasks in chemical kinetics. The necessity of estimating kinetic parameters from experimental data (which is called the inverse kinetic problem) was realized in the 1970–1980s. Both mathematicians [1, 2] and kinetics researchers [3, 4] were involved in the development of the methods for parameter estimation. To solve these problems, the results of measurements of reactant concentrations at different times of the process (that is, kinetic curves) were traditionally used. However, the concentrations of many intermediate species are so low that they cannot be measured, and therefore experimental data are not complete, and the problem of kinetic parameter estimation will not have a unique solution. This may lead to unidentifiability of the system [2], which is probably the main difficulty in parameter estimation. However, this aspect is beyond the scope of this work. At the same time, parameter estimation is topical for both scientific research and solving practical problems (e.g., for online control of a chemical process). Spectral measurements are one of the fastest and most readily available methods for obtaining online information on the process, although the results of such measurements require additional mathematical processing. In this paper, we consider the problem of extracting kinetic information from spectral data. This task is complicated by the fact that the spectra of all or some reactants are assumed to be unknown.

We can distinguish two main radically different approaches to solving the problem of parameter estimation. The first is the so-called soft approach [5–7] according to which the problem is solved in two stages. Initially a separate soft model is constructed that performs "soft calibration" of spectra against reactant concentrations. Then, concentrations predicted by this model are used in "hard" parameter fitting (i.e., kinetic parameter estimation for a chemical process). This approach often does not provide the required accuracy. According to the other approach, hard physicochemical modeling is applied [4, 8] based on the main kinetic principles, which allows one to obtain very accurate parameter estimates. However, hard methods for solving the inverse problems can be applied when pure component spectra are known. Otherwise a researcher meets the problem of simultaneous nonlinear estimation of many unknown parameters. The processing of several characteristic spectral lines simplifies the problem, but a considerable portion of information is ignored, which again leads to diminished accuracy of the results.

For solving the inverse problems, we propose that the method of successive Bayesian estimation (SBE) [9, 10] be used which performs conceptual physicochemical modeling and makes it possible to avoid difficulties associated with simultaneous estimation of many unknown parameters. The SBE method takes into account the fact that not all of the unknown parameters are equivalent: there are kinetic parameters $\boldsymbol{k}$ that are common for the whole model and there are spectral parameters $\boldsymbol{P}$ applicable to the proper wavelength x. The Bayesian approach makes it possible to use the whole body of experimental data by consecutively analyzing one wavelength after another. To preserve kinetic information obtained at the preceding stage, it is transformed into the Bayesian a priori information and is taken into account at the next stage. This makes it possible to divide the whole task of parameter estimation into a chain of smaller tasks so that having the loss in the "pathway" we have the gain in the "strength."

This method is illustrated by the two examples of the successive two-step reaction

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{C}. \qquad (1)$$

0023-1584/04/4504-0455 © 2004 MAIK "Nauka/Interperiodica"

The first example is a model where experimental data were simulated to test the efficiency of the approach. The second example is real IR spectra obtained in the epoxidation of 2,5-di-*tert*-butyl-1,4-benzoquinone. Simulated data make it possible to compare estimates with the true values of parameters and thus to check the applicability of the SBE method. Real data make it possible to compare the SBE method with other methods for estimating the rate constants using spectral data.

SIMULATED DATA

A model that describes the kinetics of changes in spectral data in the course of a chemical reaction can be described as a function of time t, the wavelength x, and unknown kinetic parameters $\boldsymbol{k}$:

$$y(t, x, \boldsymbol{k}) = \sum_{i=1}^{l} C_i(t, \boldsymbol{k}) p_i(x). \quad (2)$$

Here y is the spectral signal, C_i is the concentration of components, p_i is the pure component spectra, and l is the number of reactants. In the discrete case when the spectra are separated into m wavelengths and the time is represented by n points, one can use the matrix notation:

$$\mathbf{Y} = \mathbf{CP} + \mathbf{E}. \quad (3)$$

Here $\mathbf{Y}$ is the $(n \times m)$ matrix of spectral data; $\mathbf{C}$ is the $(n \times l)$ matrix of concentrations, which depends on the unknown kinetic parameters, and $\mathbf{P}$ is the unknown $(l \times m)$ matrix of pure component spectra. Furthermore, model (3) contains the $(n \times m)$ matrix of errors $\mathbf{E}$. In such formal consideration, the nature of measured spectra (IR, UV, NMR, etc.) is irrelevant. It is only important that the measured signal $\mathbf{Y}(t, x, \boldsymbol{k})$ is an explicit function of time t and the "wavelength" x and an implicit function of the rate constants $\boldsymbol{k}$ through the concentrations $\mathbf{C}$. Of course, Eq. (3) is only true in the ideal case. In practice, data contain a shift, baseline errors, interactions between spectra, etc. These errors can sometimes be corrected by special transformations of raw data. However, the SBE method does not assume the linearity of Eq. (3), and can be applied to complex nonideal models.

The matrix of concentrations $\mathbf{C}$ can be obtained from kinetic model (1), which has an analytical solution:

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

$$B = \frac{K_1 A_0}{k_1 - k_2}[\exp(-k_2 t) - \exp(-k_1 t)] + B_0 \exp(-k_2 t),$$

$$C = A_0 + B_0 + C_0 \quad (4)$$

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

Here the same notation for the reaction components (A, B, and C) and for their concentrations [A] = A, [B] = B, [C] = C is used, which, of course, is not correct but simplifies formulas.

Data for the simulated example were calculated using Eqs. (4) for the following values of initial concentrations: $A_0 = 1$, $B_0 = C_0 = 0$. The "true" values of the constants were chosen as follows: $k_1 = 1$ and $k_2 = 0.5$; the observation points where the spectra were measured were $t = 0, 2, 4, 6, 8$, and 10.

Thus, the number n of observations was 6. Few observation points were taken on purpose to create more difficulties in parameter estimation. Six points is the minimum number for parameter estimation from one kinetic curve corresponding to any wavelength to be possible. Indeed, the number of unknown parameters in such a model is 5: two kinetic and three spectral parameters.

The matrix of spectra of individual components $\mathbf{P}$ was constructed in the usual manner using overlapping Gaussian spectral peaks. Each spectrum $\boldsymbol{p}$ was normalized so that $\max(\boldsymbol{p}) = 1$. They were separated into 53 (m) wavelengths. The physical nature of wavelengths x is completely unimportant for the simulated example, and we will use some arbitrary wavelengths. These values are just numbers ranging from 1 to 53.

The matrix of experimental data $\mathbf{Y}$ was calculated using Eq. (3), where the matrix of concentrations $\mathbf{C}$ is given by system (4). Furthermore, white noise was added to these data with an error of 3%. These values are shown in Fig. 1, where curves represent exact values and points refer to experimental data containing random errors. Such spectra are rather difficult. They were constructed to demonstrate that the proposed approach can successfully be applied to complex spectral data. In the simulated example, all spectral values are nonnegative, and we will use this information in data processing. Simulated data were published in [11].

THE METHOD OF SUCCESSIVE BAYESIAN ESTIMATION

Our goal is to find the unknown kinetic parameters $\mathbf{k} = (k_1, k_2)^t$ using spectral data $\mathbf{Y}$ described in the form of a (6×53) matrix. If the vectors $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{r}$ of the concentrations of reactants A, B, and C are known, the inverse kinetic problem [12] is rather simple: one has to find the minimum of the sum of squares:

$$\min_{k} \sum_{i=1}^{m} \sum_{j=1}^{n} [Y_{ij} - p_i A(t_i, \boldsymbol{k}) - q_i B(t_i, \boldsymbol{k}) - r_i C(t_i, \boldsymbol{k})]^2, \quad (5)$$

where functions A, B, and C are described by Eqs. (4), and p_i, q_i, and r_i are the known preset values. However, if at least one spectrum $\mathbf{p}$, $\mathbf{q}$, or $\mathbf{r}$ is unknown (which is quite usual), the situation radically changes. It is in practice very difficult to find the minimum of sum (5) with respect to 161 unknown parameters (2 kinetic

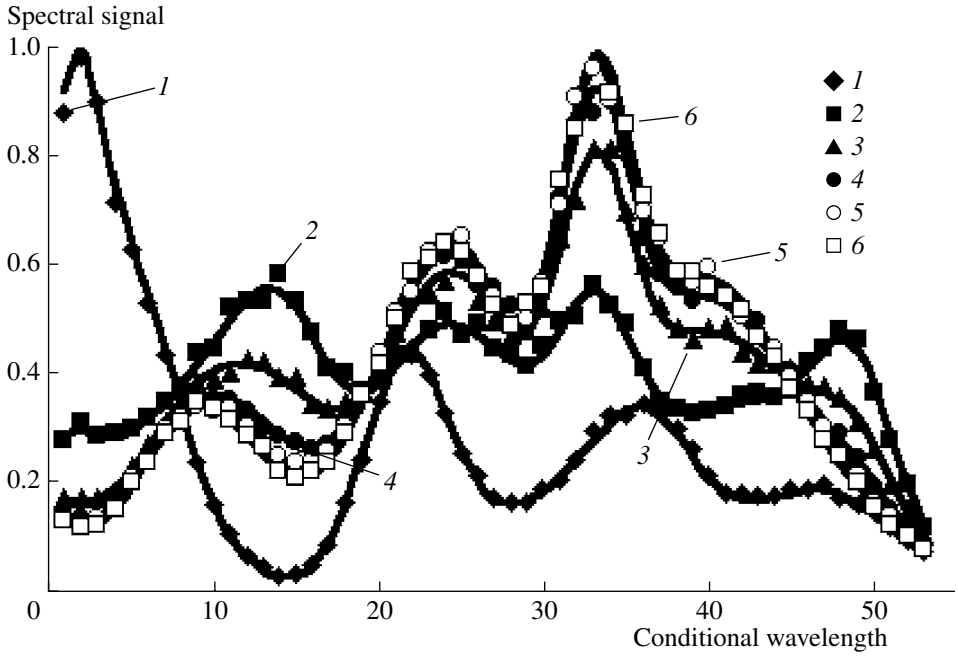

Fig. 1. Simulated data: "true" spectra (curves) and "experimental" values (points) at t = (*1*) 0, (*2*) 2, (*3*) 4, (*4*) 6, (*5*) 8, and (*6*) 10.

parameters and 53 spectral values for each of the three reactants) due to the problem of ill-conditioned matrices.

The SBE method may help in solving the problem of parameter estimation. This approach was described in [9, 13] where all necessary theorems were proven; the appendix contains the main formulas of the method. The main idea of the SBE method consists in separating the initial set of data into several parts (series). Parameters are estimated successively, one series after another, using the maximum likelihood method (MLM) [12]. Results obtained at the preceding stage are used as a priori values in the Bayesian form at a next step. The first initial portion of data is processed using the ordinary least-squares method (LSM) without a priori information. In the course of the SBE procedure, a sequence of parameter estimates is constructed. The last member in this series is the final estimate in the SBE method. It was shown in [9] that in the linear case, the method gives the same estimates as the conventional LSM. Therefore, the result of the SBE method is independent of the order in which data series are processed. In the nonlinear case, the situation is more complex, but these properties are fulfilled asymptotically.

For the task of estimating kinetic parameters k_1 and k_2, the SBE method can be described as the following algorithm:

Step 0 (Initial step). Several wavelengths $x_1, x_2\ldots$ are chosen (usually 3–6) and the corresponding spectral data $\mathbf{Y}_1$, $\mathbf{Y}_2$, … are processed together using the standard LSM method (9)–(12) (hereafter we refer to the formulas in the appendix).

Step 1. Results obtained at step 0 (estimated kinetic parameters, the F matrix, and other a posteriori information) are recalculated into the a priori information according to formulas (28)–(31).

Step 2. A new wavelength x_i is chosen, and the corresponding spectral data $\mathbf{Y}_i$ are processed separately using the MLM and taking into account a priori information obtained in step 1 (see the table in the Appendix)

Last step. Steps 1 and 2 are repeated until all wavelengths have been taken into account and all estimates obtained.

As shown in the Appendix, here one should use a priori information of the first type, because for all wavelengths, the error variance is the same. Parameters k_1 and k_2 are considered as common for all wavelengths, whereas p_i, q_i, and r_i are partial parameters.

PROCESSING OF SIMULATED DATA

Initial Procedure (Step Zero)

The SBE method requires the initial step where kinetic data are processed without a priori information (Step 0). Sometimes this step may cause a problem. In the example under consideration, the processing of kinetics for one wavelength is difficult because there are only 6 measurements for estimating 5 unknown parameters. Only some wavelengths allow such treatment, for instance conditional wavelength 16 does (see Fig. 2), but most do not.

However, it is possible to collect several wavelengths and treat the corresponding kinetic data together. In the example under consideration, it was

found that kinetic data for any four wavelengths may solve the problem of the zero step. The data contain 24 measurements for estimating 14 unknown parameters. These can be the first, last, or any four wavelengths chosen at random. Using the results of this step, one may construct a priori information and begin the SBE procedure.

Figure 2 shows the example of such initial data. These kinetic curves are used in a random SBE procedure described below.

Estimating Kinetic (Common) Parameters

It is known that in the general case the order in which data are processed by the SBE method affects the results of estimating a nonlinear model. However, from the practical standpoint, this effect is insignificant. To illustrate this idea we performed successive estimation using various orders of conventional wavelengths. The following sequences were considered: the direct order (i.e., 1, 2, 3, 4, 5, …, 53), the reverse order (i.e., 53, 52, 51, 50, 49, …, 1), and a random order (i.e., 16, 5, 29, 8, 41, …). The first four numbers in these sequences correspond to wavelengths used in the initial step. The results are presented in Figs. 3a–3c.

Curves *1* and *2* show how the estimates of kinetic parameters change in the course of the successive estimation. Shadowed regions (*1'*, *1"*, *2'*, *2"*) around curves demonstrate uncertainties of the estimates. They are formed by standard deviations added to (subtracted from) estimated parameters. Both the upper (k_1) and lower (k_2) estimates are shown on each plot. Dashed lines reflect the "true" values of kinetic constants. All data are shown depending on the numbers that are conditional wavelengths in the order (from left to right) they are used in the SBE procedure (x-axis). The first four points on each plot show the estimation of results at the initial step.

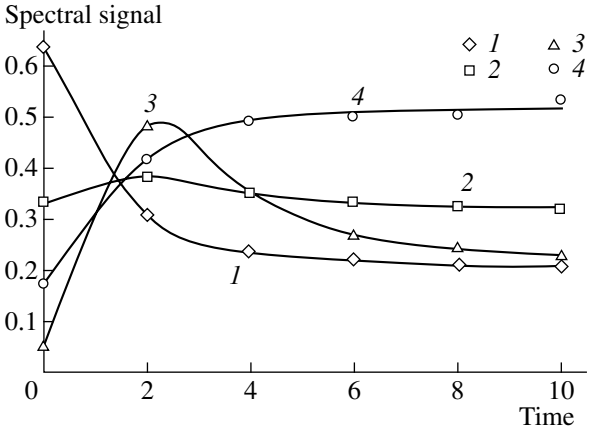

Fig. 2. Kinetic data used at the initial step of the SBE procedure with a random order of wavelengths x = (*1*) 5, (*2*) 8, (*3*) 16, and (*4*) 25.

It is seen that, although for different orders of processing the intermediate estimates are different, final estimates are very close. Figure 3d illustrates this conclusion and shows confidence ellipses (for a probability of 0.95) for all final results of the SBE procedure. Each ellipse and the label in its center represent the SBE with the corresponding order of spectrum processing. If we compare the trajectories for various methods of processing, it is noticeable that intermediate estimates and their uncertainties indeed depend on the order of the wavelengths. It is probable that the direct order is the best (Fig. 3a), because both the estimates and the deviations are changed smoothly and slowly without steps. Another interesting result can be seen in Fig. 3b, which shows the successive estimation with wavelengths in reverse order. The initial step provides a poor estimate of the parameter k_1. In the subsequent steps, uncertainty diminishes. However, these values are still far from the final results and only the two last wavelengths *2* and *1* drastically change the estimate and its deviation and increase the estimates to the final general level.

Of course, all these conclusions are correct for the example under consideration in which the spectra are as in Fig. 1. For other cases, the best order may be completely different. Let us consider this issue in more detail and begin with what causes the most problems with SBE from step zero.

It is seen from Fig. 3 that a successful choice of the initial wavelengths may considerably improve the estimation procedure (plot a), whereas a bad choice can make it worse (plot b). To automate the choice of the order of calculations, we propose use of the following simple expedient. It is clear that kinetic curves with more pronounced changes are more informative. Moreover, some nonmonotonic curves with extremums can be processed well. By combining these ideas, we obtain the following empirical criterion for estimating the comparative information content of kinetic curves:

$$L = \sum_{j=1}^{n} \sqrt{(t_j - t_{j-1})^2 + (Y_j - Y_{j-1})^2} - (t_m - t_1). \quad (6)$$

Here, t_j is the value of time (predictor), Y_j is the kinetic parameter (response), and n is the number of measurements. It can be seen that the main term in this expression is the length of a curve. The greater the value of the criterion L, the more informative the kinetic curve. The straight line parallel to the axis t is the least informative; $L = 0$. For instance, curves shown in Fig. 4 have the following characteristics: $L(16) = 0.0476$, $L(5) = 0.0286$, $L(29) = 0.0174$, and $L(8) = 0.0007$. If one arranges all wavelengths in order of decreasing information content of the corresponding kinetic data, one may expect the "optimal" result of SBE. Figure 4 confirms the correctness of this approach.

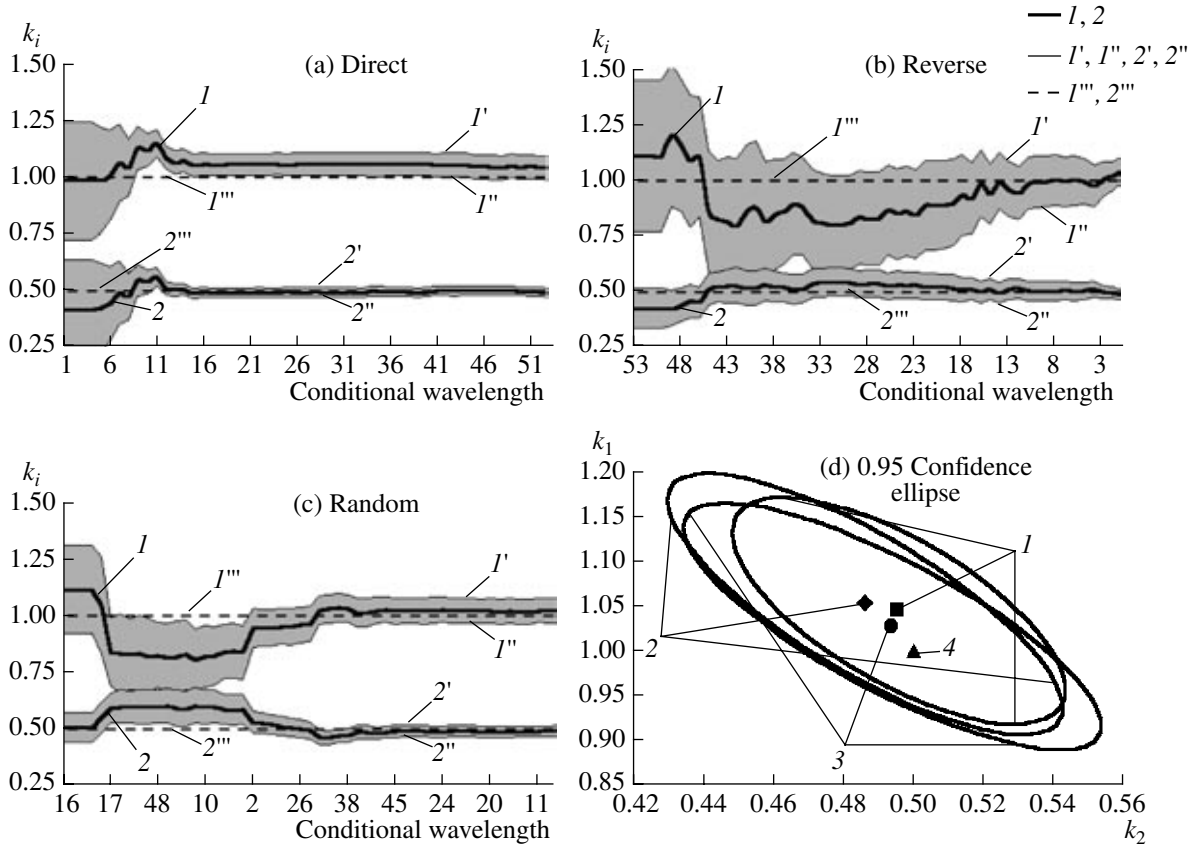

Fig. 3. Intermediate results of SBE rate constants depending on the order of wavelengths (a–c) and 0.95-confidence ellipses for the final estimates (d). On plots a–c: (*1*) k_1 estimates, (*2*) k_2 estimates; (*1'*, *1"*, *2'*, *2"*) boundaries of the regions of standard deviations; (*1"'*, *2"'*) "true" values of parameters. On plot *d* ellipses and points show (*1*) direct, (*2*) reverse, and (*3*) random orders of wavelengths; point *4* corresponds to the "true" value of the rate constant.

Estimates of Spectral (Specific) Parameters

When common kinetic parameters k_1 and k_2 are estimated, it is reasonable to determine the specific spectral parameters **p**, **q**, and **r**. Of course, this can easily be done. If the rate constants of the reactions are fixed at the level of their estimates, the set of spectral parameters p_i, q_i, and r_i can be obtained for each wavelength i using the ordinary LSM. However, the errors of such estimates will be obtained incorrectly, because the ordinary LSM cannot take into account uncertainties in the fixed kinetic parameters.

Here we can use the Bayesian approach again. With this goal, at the last step of the SBE procedure, the final a posteriori information is created, which is then transformed into the final a priori information using formula (28), which only contains data on the kinetic parameters.

This information is used as a priori information of the second type (see the table in the Appendix) for each wavelength i when the set of spectral parameters p_i, q_i, and r_i is estimated. Function (25) should be minimized taking into account additional constraints:

$$\frac{\partial S_i}{\partial k_a} = 0, \quad a = 1, 2,$$

which show that kinetic parameters should not be re-estimated. Here S_i is the sum of squares determined in (8).

Figure 5 shows the "true" model spectra (curves *1*). The results of estimates are shown as the difference between the estimated and "true" spectra (points *2*). The regions of three standard deviations are also shown (*2'*, *2"*). It is seen from these plots that the estimates are very accurate, especially for the components A and C.

SBE Method Validation Using Simulated Data

The method proposes a new approach to determining kinetic constants using spectral data. It can be used in situations when a large amount of partial parameters prevent common parameters from being estimated accurately as in the task under consideration.

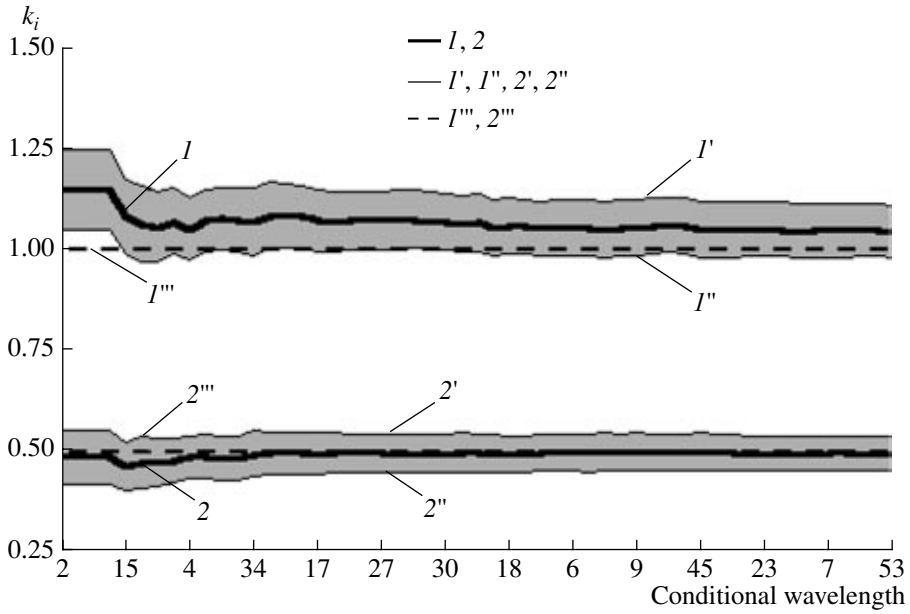

Fig. 4. Intermediate results for the "optimal" order of wavelengths: (*1*) k_1 estimates, (*2*) k_2 estimates; (*1*', *1*", *2*', *2*") boundaries of the regions of standard deviations; (*1*"', *2*"') "true" values of parameters.

To demonstrate that this method is reliable, we performed the following "validation test". We processed all model simulated data using the LSM. In the search, we simultaneously estimated all 161 parameters in sum (5). The results of the test are illustrated in Fig. 6. Points 1 and 2 in the centers of ellipses show the values estimated using the SBE method and LSM, respectively. They and the corresponding elliptical confidence regions are very close. This calculation supports the results of the theorem (see Appendix) consisting in that the results of the SBE procedure asymptotically approach the LSM estimates. It is interesting to determine whether we can trust these confidence regions. It is known that the covariation analysis is very difficult in a nonlinear case, whereas the use of the linearization method is only justified when the model is close to linear. The measure of this proximity can be estimated by the coefficient of nonlinearity, which is calculated using the Monte Carlo method [13].

For a linear (or close to linear) model, the coefficient should be equal to unity. The higher this coefficient, the more nonlinear is the model under study. In our case, the coefficient of nonlinearity was calculated and its value proved to be equal to unity; that is, our model is close to linear and the calculated ellipse correspond to the correct confidence regions. This result was supported by the method of statistical modeling in [14].

REAL EXAMPLE

To check the correctness of the approach we took data from [6, 15] where a set of IR spectra for 2,5-di-*tert*-butyl-1,4-benzoquinone epoxidation was considered. The real experimental process is described by a two-step reaction (1). The reactants, measurement procedures and other details of the experiment are described in detail in [6]. Raw experimental data can be found in [16]. In that paper, 240 spectra were reported that were recorded in the range 800–1100 nm with a 1.0-nm step. The reaction time was 1200 s, and the measurements were carried out each 5 s.

Raw spectral data were preprocessed according to the procedure described in [6]. It can briefly be described as follows. The fourth spectrum (at $t = 20$ s) is used as base one and subtracted from all other spectra. Then, for these spectra, the second derivatives are calculated using the Savitzky–Golay filter [17] with a 15-point window. Finally, we used a narrow range of wavelengths in calculations 860–880 nm. Figure 7 shows these data.

Data were processed using the SBE procedure with "direct" and "optimal" orders of wavelengths. In this example spectral parameters **p**, **q**, and **r** may be negative; therefore, no constraints were imposed. The results of the "direct" procedure are shown in Fig. 8, which illustrates how estimates and their accuracy change in the course of the successive procedure. The plot is constructed as in the case of Fig. 3. The final estimates of rate constants, achieved at the last step for wavelength 880 nm were $k_1 = 0.267 \pm 0.015$ min^{-1} and $k_2 = 0.095 \pm 0.010$ min^{-1}. Here, the standard deviations are also shown. The correlation coefficient is $r = -0.18$. The "optimal" SBE procedure gave the following values: $k_1 = 0.238 \pm 0.015$ (min^{-1}), $k_2 = 0.102 \pm 0.010$ min^{-1}, and $r = -0.22$.

It is interesting to compare these results with analogous estimates obtained using other, conventional

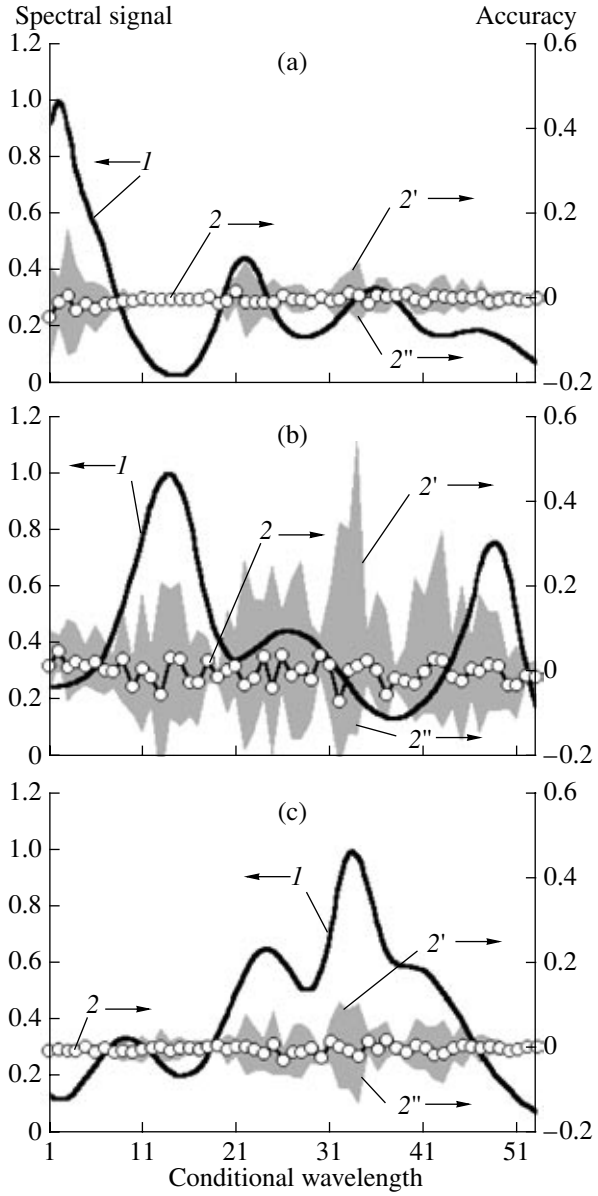

Fig. 5. Model spectra of reagents (a) A, (b) B, and (c) C (curves *1*). The difference between the estimate and the "true" spectrum (points *2*). Regions *2'*, *2"* represent three standard deviations.

methods using the same set of experimental data. These methods were described in detail in [15]; therefore we only provide here a brief reminder.

The classical curve resolution (CCR) method is an iterative procedure that exploits the linearity of model (3) with respect to spectral parameters. The application of this method usually assumes that the spectra of initial species A and product C are known and the spectrum of the intermediate species B and the rate constants are unknown. Upon choosing some initial values of the rate constants, one may estimate the spectral parameters of the intermediate species using the standard LSM and

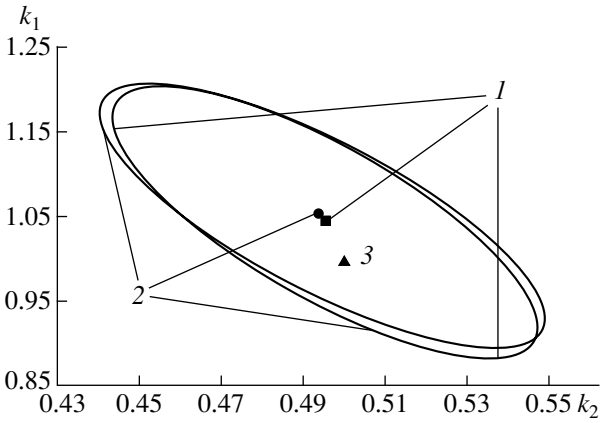

Fig. 6. Estimates represented by 0.95-confidence ellipses: (*1*) SBE, (*2*) LSM, and (*3*) "true" values.

construct an approximate matrix $\hat{\mathbf{P}}$ of individual spectra (constrained by the fact that the spectrum of the intermediate species is nonnegative). Then, using this matrix, the rate constants are made more exact using the Levenberg–Marquardt method [18, 19]. The algorithm is repeated unless the process converges.

The weighted curve resolution (WCR) method [20] is also iterative and unites a formal approach that uses the singular decomposition of the data matrix **Y** and a kinetic approach for calculating the matrix of concentrations **C** as in the CCR method. In contrast to CCR, the use of the WCR method does not require the knowledge of individual spectra of species A and C. However, there are no strict conditions that guarantee the convergence of the CCR and WCR procedures.

The above methods belong to the so-called 2-way methods. They are in contrast to 3-way methods, which are used more widely. In the 3-way methods, each component of model (3) is replaced by a 3D matrix (tensor). To form a tensor, the initial data array is divided into two subarrays using a time shift. Then, the superposition of these subarrays forms a 3-way model. The simplest of these methods is the generalized rank annihilation method (GRAM) [21]. This one is based on a formal approach in which a simple equation is used:

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

which shows that the rate constant of the first-order reaction can be determined from the ratio of the initial and shifted exponents. The GRAM is a very fast iterative method, and this is its main advantage. If the level of noise in spectral data is high, the accuracy is low. In that case GRAM estimates can be used as the initial approximation for the iterative formal LM–PAR method [6]. It improves GRAM estimates using the Levenberg–Marquardt algorithm [18, 19] and the procedure PARAFAC [22]. Note that both GRAM and

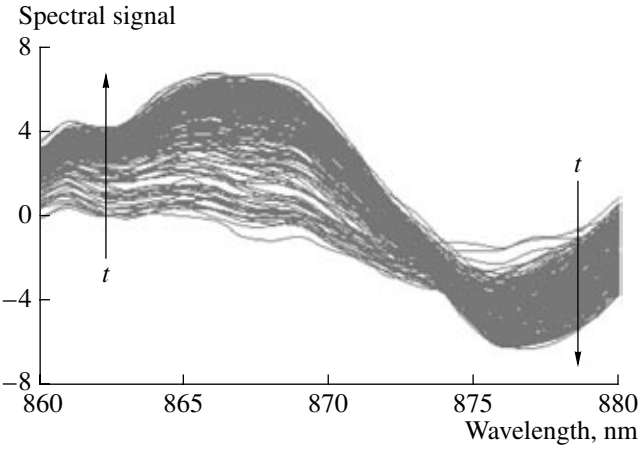

Fig. 7. Spectra treated in the real example.

LM–PAR can only be applied to pseudo first-order kinetics.

Estimates obtained by different methods are shown in Fig. 9 where each method is represented by a confidence ellipse (for a probability of 0.95) constructed using the data from [15]. Data for ellipse *5*, which corresponds to the optimal SBE method, are described above. Comparison of the size and forms of ellipses allows for comparison of the methods. Thus, it can be seen that ellipses *1* and *5* have the smallest size and the most round shape. This means that the SBE method produces the smallest variances and correlations of all the methods considered here and that its results are close to the results of the CCR/LSM procedure. It is well known that the LSM is optimal, although this method is unreliable in the case of highly dimensional data. The properties of the SBE method are similar to the properties of the LSM, but SBE is more stable.

CONCLUSIONS

The SBE method can be applied in determining the rate constants of chemical reactions using spectral data for the case when individual spectra of the reactants are unknown. This is a rather general approach and it can be used for any kinetic model. The reliability of the method was demonstrated using a simulated and a real example. The simulated example showed that the proposed method is efficient for both the case of a small number of observations and for the case of strongly overlapping spectra. It was also shown that the SBE method gives results close to those obtained using the conventional LSM. The results using SBE for the real example suggest that this method provides the smallest errors compared to other conventional procedures. It is important that the method is Bayesian in form but not in essence. This means that it can be used without assuming any subjective a priori information. Each element of the a priori information is constructed using the result of the preceding step of the procedure and only the form of application of this information is dictated by the Bayes theorem. No additional assumption (like the number of principal components, time shift, or the first order of the reaction) should be made for the use of this method.

The SBE method is a rather fast method for obtaining estimates but slower than the LSM. The ratio of calculation time t_{SBE}/t_{LSM} can be estimated [23] as

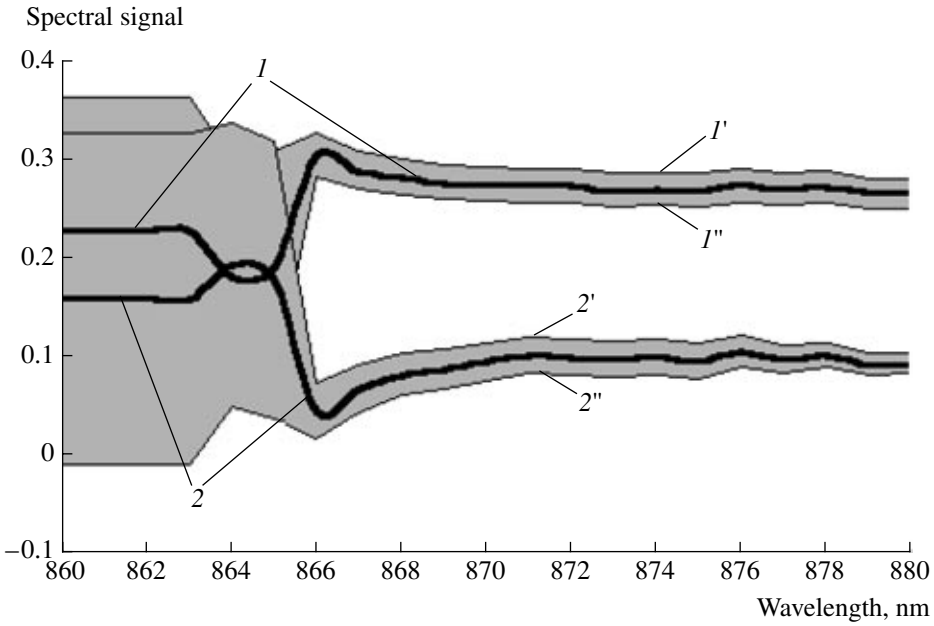

Fig. 8. Intermediate results of SBE kinetic constants in the real example: (*1*) k_1 estimates, (*2*) k_2 estimates; (*1*', *1*", *2*', *2*") boundaries of the regions of standard deviations.

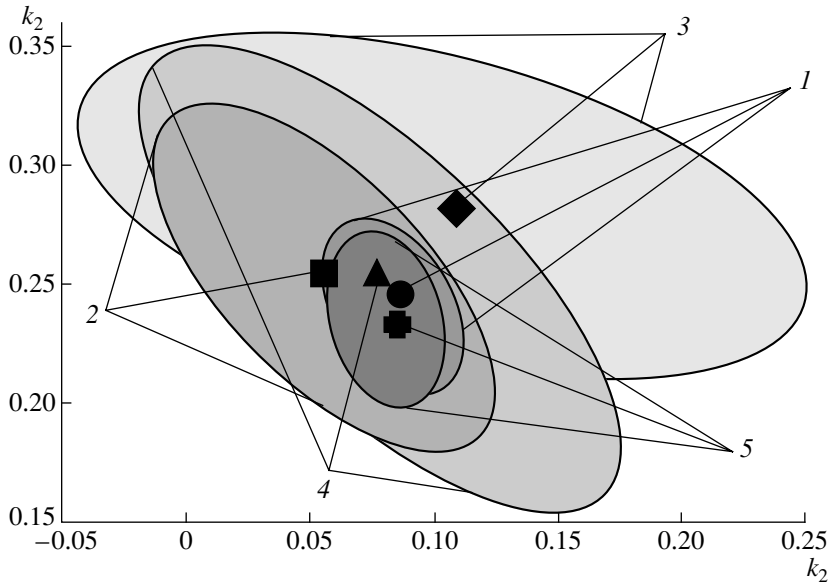

Fig. 9. Estimates obtained for the real example by different methods: (*1*) CCR, (*2*) WCR, (*3*) LM-PAR, (*4*) GRAM, and (*5*) SBE.

$\left(1+\frac{p}{l}\right)^3$, where p is the number of rate constants (steps) and l is the number of reactants. For the two-step reaction this ratio is 4.6, although in our examples it was close to 2.

SOFTWARE

All calculations were carried out using the FITTER software [24, 25] for MS Excel.

APPENDIX

The Method of Successive Bayesian Estimation (SBE)

Let us consider the conventional regression task:

$$y_i = f(\mathbf{x}_i, \mathbf{a}) + \varepsilon_i, \quad i = 1, \ldots N, \quad e \sim N(\mathbf{0}, \sigma^2\mathbf{I}),$$

where $\mathbf{a} = (a_1, \ldots, a_p)^t$ is the vector of the unknown parameters and σ^2 is the unknown variance of error. The corresponding likelihood function (the argument $\mathbf{y}$ is omitted) has the following form:

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

where

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

is the sum of squares. Using the MLM for function (7), we obtain the following values:

Estimates of parameters $\mathbf{a}$,

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

The Fisher matrix $\mathbf{A}$, which characterizes the accuracy of the estimate

$$\mathbf{A} = \mathbf{V}^t\mathbf{V}, \text{ where } V_{\alpha i} = \frac{\partial f(\mathbf{x}_i, \hat{\mathbf{a}})}{\partial a_\alpha}, \quad \alpha = 1, \ldots, p; \quad i = 1, \ldots, N. \tag{10}$$

This $(p \times p)$ matrix is the Hessian of the function $S(\mathbf{a})$ in the Gauss–Newton approximation [12], and if it is invertible, then $\mathrm{cov}(\hat{\mathbf{a}}, \hat{\mathbf{a}}) = \sigma^2\mathbf{A}^{-1}$.

The variance estimate σ^2

$$s^2 = \frac{S(\hat{\mathbf{a}})}{N_f}, \tag{11}$$

where N_f is the number of degrees of freedom (NDF) for estimate (11):

$$N_f = N - p. \tag{12}$$

Near the maximum, the likelihood function can be approximated by the expression

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

Let us consider the case when the a priori information is available and described by some distribution $h(\mathbf{a}, \sigma^2)$. Then, the likelihood function takes the form

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

Let us construct the distribution function h using information consisting of the following values corresponding to the values in expressions (9)–(12):

Statistics obtained by the maximum likelihood method with a priori information

Statistics	A priori information of type 1	A priori information of type 2
Parameter estimates	$\hat{\boldsymbol{a}} = \arg\min Q(\boldsymbol{a})$ $Q(\boldsymbol{a}) = S(\boldsymbol{a}) + B(\boldsymbol{a})$	$\hat{\boldsymbol{a}} = \arg\min Q(\boldsymbol{a})$ $Q(\boldsymbol{a}) = S(\boldsymbol{a}) + B(\boldsymbol{a})$
Fisher matrix	$\boldsymbol{A} = \boldsymbol{V}^t\boldsymbol{V} + s_0^2\boldsymbol{H}$	$\boldsymbol{A} = \exp\left(\frac{R(\hat{\boldsymbol{a}})}{N}\right)\left(\boldsymbol{V}^t\boldsymbol{V} + \frac{S(\hat{\boldsymbol{a}})}{N}\boldsymbol{H}\right)$
Variance estimates	$s^2 = \frac{Q(\hat{\boldsymbol{a}})}{N_f}$	$s^2 = \frac{S(\hat{\boldsymbol{a}})}{N_f}$
NDF	$N_f = N + N_0$	$N_f = N$

(1) a priori values of parameters

$$\mathbf{b} = (b_1, \ldots, b_p)^t, \tag{15}$$

(2) a priori information matrix

$$\mathbf{H} = \{h_{\alpha\beta}\}, \quad \alpha, \beta = 1, \ldots, p, \tag{16}$$

(3) a priori variance value

$$s_0^2, \tag{17}$$

(4) and a priori NDF

$$N_0. \tag{18}$$

A priori information involving all these four values is called information of *the first type*. Sometimes variance (17) and NDF (18) are not available. In this case information is called information of *the second type*.

In the Bayesian approach, the value **a** can be considered (see (13)) as a normal random vector with expectation **b** and the accuracy matrix $\gamma\mathbf{H}$

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

where $R(\mathbf{a})$ is the quadratic form and

$$R(\mathbf{a}) = (\mathbf{a} - \mathbf{b})^t\mathbf{H}(\mathbf{a} - \mathbf{b}). \tag{20}$$

The multiplier γ depends on the type of information and is equal to $\gamma = \frac{s_0^2}{\sigma^2}$ for type 1 and $\gamma = 1$ for type 2.

The a priori distribution of variance σ^2 can be expressed in terms of function χ^2 and the values (17) and (18)

$$\sigma^2 \sim \frac{(2N_0s_0^2)^{\frac{N_0}{2}}}{\Gamma\left(\frac{N_0}{2}\right)}\sigma^{-N_0-2}\exp\left(-N_0\frac{s_0^2}{2\sigma^2}\right). \tag{21}$$

By combining (19) and (21) with equation (14), one can describe the likelihood function as

$$L_0(\mathbf{a}, \sigma^2) = C_1\sigma^{-N-N_0-2}\exp\left[-\frac{1}{2\sigma^2}(S(\mathbf{a}) + s_0^2R(\mathbf{a}) + s_0^2N_0)\right], \tag{22}$$

taking into account a priori information of type 1 and as

$$L_0(\mathbf{a}, \sigma^2) = C_2\sigma^{-N}\exp\left[-\frac{1}{2}\left(\frac{S(\mathbf{a})}{\sigma^2} + R(\mathbf{a})\right)\right], \tag{23}$$

taking into account a priori information of type 2. Multipliers C_1 and C_2 are not important, because they are independent of both **a** and σ^2.

The MLM estimates are a point in which the likelihood function has a maximum. It can easily be shown that it corresponds to the minimum of some objective function $Q(\mathbf{a})$. By differentiating (22) and (23), we obtain that the objective function takes the form

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

for the information of type 1 and

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

for information of type 2. Here $B(\mathbf{a})$ is the Bayesian term, which is

$$B(\mathbf{a}) = s_0^2[N_0 + R(\mathbf{a})] \tag{26}$$

for information of type 1 and

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

for information of type 2. Some essential statistics of the MLM with a priori information are described in the table.

In the SBE method, the initial array of data is separated into parts (series) that are processed in a successive manner. At each step (excluding the first one) the MLM is applied with a priori information constructed from the results of the preceding step. Let us illustrate this.

Suppose we have statistics (see table) obtained after estimation at the *i*th step. They can be called a posteriori information. This information corresponds to the

values in expressions (15)–(18) and, of course, can be used as a posteriori information at the next $(i + 1)$th step. However, it is necessary to make some important details more exact.

Often, each series of data is described by its own regression function $f_i(\mathbf{x}, \mathbf{a}_i)$, which depends on both common and partial parameters.

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

The subset of parameters $a_1, \ldots, a_r$ is called common because each regression function f_i, $i = 1, \ldots, M$ depends on all these parameters. On the other hand, the subset $a_{r+1}, \ldots, a_{p_i}$ is called specific because each such subset of parameters is only present in the single function f_i.

In constructing a priori information, it is necessary to separate the portion of a priori information that corresponds to common parameters. It should be preserved for further calculations, while the portion corresponding to partial parameters should be deleted. The a posteriori Fisher matrix $\mathbf{A}$ (the subscript denoting a step i is omitted for simplicity) can be presented as the block matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{00} & \mathbf{A}_{01} \\ \mathbf{A}_{01}^t & \mathbf{A}_{11} \end{bmatrix},$$

where $\mathbf{A}_{00}$ is a square $(r \times r)$ matrix corresponding to the common parameters; $\mathbf{A}_{11}$ is a square $(p_i - r) \times (p_i - r)$ matrix corresponding to the partial parameters; and $\mathbf{A}_{01}$ is a rectangular $r \times (p_i - r)$ matrix. Then, the a priori information matrix $\mathbf{H}$ is calculated from the matrix $\mathbf{A}$ according to equation

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

where s^2 is the a posteriori value of error variance. The dimensionality of this matrix should correspond to the number of parameters in the next series of data (that is, $(p_{i+1}) \times (p_{i+1})$). Therefore, it should be supplemented with zeros. The a priori values of the parameters are recalculated in a similar manner:

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

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

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

but the a priori number of degrees of freedom should be recalculated as

$$N_f = N_i - N_0 - p_i + r, \quad (31)$$

where N_i is the number of data and N_0 is the a posteriori value of NDF in the ith series. Thus, Eqs. (28)–(31) determine the a priori information which is applied at the next step of the SBE procedure.

Now let us compare the SBE procedure with the conventional LSM. In the LSM, the objective function takes the following form:

$$S(\mathbf{a}_1, \ldots, \mathbf{a}_M) = S_1(\mathbf{a}_1) + \ldots + S_M(\mathbf{a}_M). \quad (32)$$

In the SBE method, this is a sum of squares that involves only the jth series of data:

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

In the LSM, the estimate of parameters $\mathbf{a}$ has the form

$$\hat{\mathbf{a}} = (\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_M) = \arg\min S(\mathbf{a}_1, \ldots, \mathbf{a}_M), \quad (34)$$

and the estimate of the variance of error is

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

Here $N = N_1 + \ldots + N_M$ is the overall number of measurements and $p = r + p_1 + \ldots + p_M$ is the total number of parameters $\mathbf{a}$.

The SBE procedure can be described in the form of the following algorithm:

(1) The whole set of data is separated into M series;

(2) The first series is processed by the standard LSM (8)–(10);

(3) A posteriori information is transformed into a priori information (28)–(31);

(4) The next series is processed using the MLM and a priori information (see table);

(5) Steps 3 and 4 are repeated unless the last series is processed;

(6) The results are the SBE estimates.

Generally speaking, the SBE estimates depend on the order the series are processed. However, in the linear case, it is possible to show that the SBE estimates possess the following property [9].

Theorem. Let the regression functions $f_j(\mathbf{x}, \mathbf{a}_j)$, $j = 1, \ldots, M$ be linear with respect to parameters a_j and the error be homoscedastic (that is, $\sigma_1^2 = \ldots = \sigma_M^2 = \sigma^2$), then the following estimates obtained by LSM and SBE methods coincide:

(1) The estimates of common parameters;

(2) The covariance matrices of common parameters;

(3) The estimates of error variance; and

(4) The number of degrees of freedom.

Therefore, the SBE estimates are independent of the order in which the data series are processed and these properties are fulfilled asymptotically in a nonlinear case.

REFERENCES

1. Lavrent'ev, M.M., Kraeva, A.G., and Bukhgeim, A.V., *Obratnaya zadacha khimicheskoi kinetiki* (Parameter Estimation in Chemical Kinetics), Novosibirsk: Computer Center, Siberian Division, USSR Acad. Sci., 1980.
2. Spivak, S.I. and Gorskii, V.G., *Dokl. Akad. Nauk SSSR*, 1981, vol. 257, no. 2, p. 412.
3. *Primenenie vychislitel'noi matematiki v khimicheskoi i fizicheskoi kinetike* (Applications of Computational Techniques in Chemical and Physical Kinetics) Polak, L.S., Ed., Moscow: Nauka, 1969.
4. Pavlov, B.V. and Brin, E.F., *Khim. Fiz.*, 1984, vol. 3, no. 3, p. 393.
5. Haario, H. and Taavitsainen, V.-M., *Chemometrics Intell. Lab. Syst.*, 1998, vol. 44, p. 77.
6. Bijlsma, S., Louwerse, D.J., Windig, W., and Smilde, A.K., *Anal. Chim. Acta.*, 1998, vol. 376, p. 339.
7. Bijlsma, S., Louwerse, D.J., and Smilde, A.K., *J. Chemometrics*, 1999, vol. 13, p. 311.
8. Brin, E.F. and Pomerantsev, A.L., *Khim. Fiz.*, 1986, vol. 5, no. 12, p. 1674.
9. Maksimova, G.A. and Pomerantsev, A.L., *Zavod. Lab.*, 1995, vol. 61, p. 432.
10. Pomerantsev, A.L., *Chemometrics Intell. Lab. Syst.*, 2003, vol. 66, no. 2, p. 127.
11. Fitter Solutions, http://polycert.chph.ras.ru/solution.htm.
12. Bard, Y., *Nonlinear Parameter Estimation*, New York: Academic.
13. Bystritskaya, E.V., Pomerantsev, A.L., and Rodionova, O.Ye., *J. Chemometrics*, 2000, vol. 14, p. 667.
14. Pomerantsev, A.L., *Chemometrics Intell. Lab. Syst.*, 1999, vol. 49, p. 41.
15. Bijlsma, S. and Smilde, A.K., *J. Chemometrics*, 2000, vol. 14, p. 541.
16. Datasets, http://www.its.chem.uva.nl/research/pac.
17. Savitzky, A. and Golay, M.J.E., *Anal. Chem.*, 1964, vol. 36, p. 1627.
18. Marquardt, D.W., *SIAM J.*, 1963, vol. 11, p. 431.
19. Levenberg, K., *Quart. Appl. Math.*, 1944, vol. 2, p. 164.
20. Lawton, W.H. and Sylvestre, E.A., *Technometrics*, 1971, vol. 13, p. 617.
21. Wilson, B., Sanchez, E., and Kowalski, B.R., *J. Chemometrics*, 1989, vol. 3, p. 493.
22. Harshman, R.A. and Lundy, M.E., *Comp. Stat. Data Anal.*, 1994, vol. 18, p. 39.
23. Rajko, R., *Anal. Lett.*, 1994, vol. 27, no. 1, p. 27.
24. Pomerantsev, A.L. and Rodionova, O.E., *Nadstroika FITTER* (FITTER Add-in), reg. no. 2002611562, 2002.
25. Fitter Add-in, http://polycert.chph.ras.ru/fitter.htm.