ISSN 1062-8738, Bulletin of the Russian Academy of Sciences: Physics, 2009, Vol. 73, No. 2, pp. 245–248. © Allerton Press, Inc., 2009.
Original Russian Text © I.D. Gorlachev, B.B. Knyazev, A. Kuketayev, F.M. Pen'kov, 2009, published in Izvestiya Rossiiskoi Akademii Nauk. Seriya Fizicheskaya, 2009, Vol. 73, No. 2, pp. 257–260.

Properties of Integral Least Squares Method

I. D. Gorlachev[a], B. B. Knyazev[a], A. Kuketayev[b], F. M. Pen'kov[a]

[a] *Institute of Nuclear Physics, Almaty, Kazakhstan.*

[b] *Karaganda State University, Karaganda, Kazakhstan.*

e-mail: penkov@inp.kz

Abstract—A new modification of the least squares method (LSM) is proposed. The main idea is to consider the fitting parameters β_i as independent random variables with a certain distribution density $F(\beta_1, \beta_2, \dots, \beta_k; \varphi_1, \dots, \varphi_m)$, which depends on a set of m experimental points φ_j. Within this approach, the estimates of the parameters $\hat{\beta}_i$ minimize squared deviations and are equivalent to means of the probability distribution $\hat{\beta}_i = \bar{\beta}_i = \int \beta_i F(\beta_1, \beta_2, \dots, \beta_k; \varphi_1, \dots, \varphi_m)\, d\beta_1 d\beta_2 \dots d\beta_k$.

DOI: 10.3103/S1062873809020269

1. INTRODUCTION

Various methods for finding the best possible fit within a certain family of (model) functions for a certain sample of experimental points are based on the criterion of closeness (measures) for such samples. The following measure is especially popular in experimental physics:

$$R = \sum_{i=1}^{N} \frac{(\varphi_i - \varphi_i^{\exp})^2}{\sigma_i^2}, \tag{1}$$

which is understood as a weighted squared deviation that can be used to apply the χ^2 criterion [1, 2]. The notation $\varphi_i = \varphi_i^{\exp} \pm \sigma_i$ for an experimental value means that φ_i is distributed around $\varphi_i^{\exp}$ with the standard deviation σ_i. Most frequently, it is assumed that this distribution is normal (Gaussian) [3]. Therefore, the probability distribution density for finding φ_i

$$F(\varphi_i) \propto \exp\left\{-\frac{(\varphi_i - \varphi_i^{\exp})^2}{2\sigma_i^2}\right\} \tag{2}$$

gives an evident expression for the probability density for measuring all values φ_i $(1 \le i \le m)$:

$$F(\varphi_1, \varphi_2, \dots, \varphi_m) = \prod_{i=1}^{m} F(\varphi_i) \propto \exp\left\{-\sum_{i=1}^{m} \frac{(\varphi_i - \varphi_i^{\exp})^2}{2\sigma_i^2}\right\}. \tag{3}$$

The minimum of the exponent in (3) (i.e., minimum in (1)) determines the maximum probability of the experimental values φ_i.

Assuming that random variables form some unambiguous model

$$\varphi_i = \varphi_i^{\mathrm{mod}}(\beta_1, \beta_2, \dots, \beta_k):\ k < m, \tag{4}$$

one can formulate the problem of finding unknown parameters β_i as a problem for finding $\hat{\beta}_i$ ensuring a maximum of probability (3) (or a minimum of (1), i.e., minimum $R = R(\boldsymbol{\beta})$). Thus, the standard problem of minimizing the value of χ^2 statistics or the least (in the sense of (1)) squares method (LSM), corresponds to maximization of the joint density (3). Note that the theoretical background of LSM has been developed for only linear implementation of model (4):

$$\varphi_i^{\mathrm{mod}}(\beta_1, \beta_2, \dots, \beta_k) = \sum_{j=1}^{k} \beta_j f_i^j, \tag{5}$$

where the functions f_i^j constitute the model and their explicit form is determined by the "physics" of measured quantities. The models with more complicated (nonlinear) dependences on β_i are often linearized near the global minimum of (1), after which one can use the LSM theory for linear models to draw certain conclusions on the quality of the theoretical fit (for example, find confidence limits for obtained parameters).

LSM has a reputation of a fairly reliable approach for solving data smoothing problems, finding unknown model parameters, and even finding confidence limits for those parameters. Nevertheless, there are problems that require much more sophisticated LSM versions, for example, finding LSM parameters when there are certain constraints on the values of the model parameters or determining some function of not the values of $\hat{\beta}_i$ corresponding to a minimum of (1) but of some β_i distributions.

Below, we propose another approach to finding the LSM parameters. We call it the integral MLS (IMLS). In our opinion, this method admits extension of the standard LSM to a wider class of problems. Of course, for standard problems, the parameters $\hat{\beta}_i$ found within ILSM and LSM coincide, and IMLS is only of methodological importance for such problems.

2. INTEGRAL LSM

In this section, we develop an approach for determining the parameters $\hat{\beta}_i$, which is based on the following simple observation: the means of the random variables $\bar{\beta}_i$

$$\bar{\beta}_i = \int \beta_i F(\boldsymbol{\beta}|\varphi^{\exp}) d\beta_1 \ldots d\beta_k, \tag{6}$$

calculated from

$$F(\boldsymbol{\beta}|\varphi^{\exp}) = \frac{e^{-R(\boldsymbol{\beta})}}{\int (e^{-\mathbf{R}(\beta)} \cdot d\beta_1 d\beta_2 \ldots d\beta_k)}, \tag{7}$$

coincide with the solution to the system of linear LSM equations:

$$\frac{\partial R(\beta_1, \beta_2, \ldots, \beta_k)}{\partial \beta_n} = 2\sum_{i=1}^{m}\left(\sum_{j=1}^{k} \beta_j f_i^j - \varphi_i^{\exp}\right) f_i^n = 0. \tag{8}$$

Moreover, the standard deviations $\Delta\beta_i = \sqrt{\overline{\beta_i^2} - \bar{\beta}_i^2}$, calculated with the distribution density $F(\boldsymbol{\beta}|\varphi^{\exp})$, coincide with the errors of the linear MLS parameters. We do not present here a proof of the above statements because it is simple but tedious. Instead, we will consider below an example of finding the MLS parameters using two approaches: standard MLS and IMLS.

Of course, the fact the two approaches yield identical results is not a mere technical coincidence. Indeed, if we impose k constraints given by Eq. (5), the joint distribution density of the experimentally observed values (3) transforms into the distribution density $F(\boldsymbol{\beta}|\varphi^{\exp})$. Because of that, the parameters β_i are random variables with the joint distribution density (7). Note that the denominator of Eq. (7) is just a normalizing factor.

One can consider the distribution density (3) as a maximum likelihood function [2, 4]. Here, we deal with a particular case where the maximum likelihood function coincides with the distribution density of random variables β_i. Generally, such a statement is incorrect (see, e.g., [2]), and ILSM might yield biased estimators. To stress the connection of the distribution density (7) with the maximum likelihood function, we will use the notation $F(\boldsymbol{\beta}|\varphi^{\exp})$, which is usually reserved for likelihood functions.

In the general (nonlinear) case of the LSM, the proposed method ILSM can produce biased estimators, which differ from the estimators corresponding to the global minimum of Eq. (1). Note however that there are no arguments justifying the optimality of the estimators produced by the global minimum of Eq. (1) in the case of nonlinear LSM. For instance, there is no guarantee that the global minimum of Eq. (1) gives least-variance estimators. On the contrary, it may happen that the estimators produced by ILSM are not only unbiased but also have substantially smaller variances than those of the conventional LSM estimators.

We summarize the above discussion with the following statement: there is a joint distribution function of the parameters β_i of non-linear LSM models, which allows one to calculate any means of these parameters, including their functions, under any restrictions on the domain of existence of the parameters β_j. These restrictions simply determine the domain of integration for the distribution densities (7), including the normalization range. In the case of nonlinear LSM models, the ILSM data require a separate analysis.

Below, we consider a simple example where the estimators for the model parameters within the standard LSM and ILSM coincide. We will also consider an example where the standard deviation of the parameter-estimation errors is comparable with the expected values In this case, the standard deviations found according to the well-known recipe is incorrect because it is computed by averaging random variables that can take any values from the interval (–∞, ∞), while the model parameters may have physical constraints ($\beta_{\min}$, $\beta_{\max}$). For instance, the peak amplitudes in spectroscopy must be only nonnegative. In such cases, it is not difficult to obtain better confidence intervals within ILMS by finding the bounds (a, b) satisfying the following equations (for the 95% confidence level)

$$\begin{cases} 2.5\% = \iiint d\beta_1 \ldots d\beta_k \int\limits_{\beta_{min}}^{a} F(\boldsymbol{\beta}|\varphi^{\exp}) d\beta_j, \\ 97.5\% = \iiint d\beta_1 \ldots d\beta_k \int\limits_{b}^{\beta_{max}} F(\boldsymbol{\beta}|\varphi^{\exp}) d\beta_j. \end{cases} \tag{9}$$

3. NUMERICAL RESULTS

In this section, we consider two numerical examples of applying ILSM. The first example demonstrates the coincidence of the ILSM and LSM data for models linear in parameters. The second example shows the capabilities of ILSM in the presence of constraints on the model parameters.

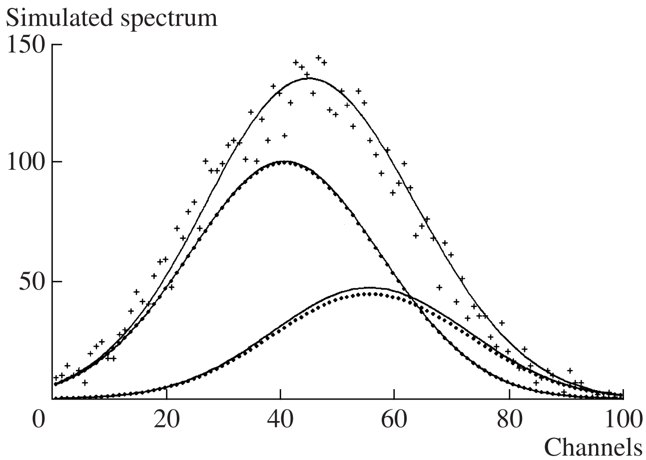

Fig. 1. Example 1: (+) noisy experimental data, (solid line) true curve, and the curve components—small and large peaks; the dotted lines are the peaks recovered by IMLS.

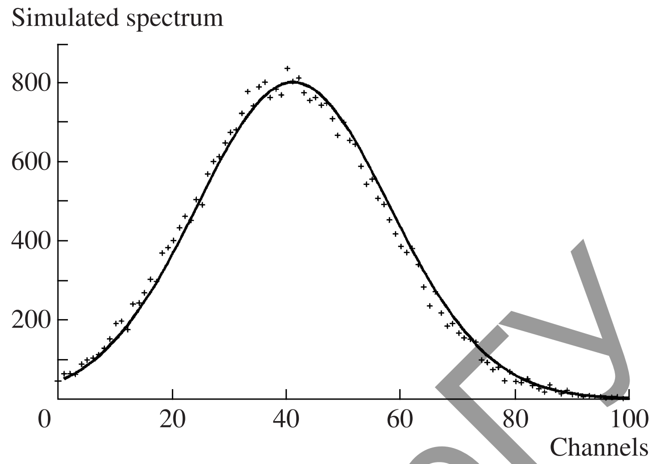

Fig. 2. Example 2: (+) noisy experimental data and (solid line) the true spectrum.

For the first example, we chose two Gaussians:

$$\varphi_{0,i} = \beta_0 \exp\left\{-\frac{(i-i_0)^2}{2\sigma^2}\right\},$$
$$\varphi_{1,j} = \beta_1 \exp\left\{-\frac{(i-i_1)^2}{2\sigma^2}\right\}, \quad (10)$$

where $\varphi_{0,i}$ and $\varphi_{1,i}$ are, respectively, the "true" large and small Gaussian peaks; β_0 and β_1 are their amplitudes;, i is the channel number; i_0 and i_1 are the positions of the large and small peaks; and σ is the peak width.

To obtain $\varphi_i^{\exp}$, we distorted the sum of the two peaks by Poisson noise.

Figure 1 shows the data obtained with the following values of the parameters: $i_0 = 40$, $i_1 = 55$, $\beta_0 = 100$, and $\beta_1 = 47$. The value of σ correspond to the half-width at half-maximum (20). The above values were chosen to ensure a significant overlap of the two peaks. For comparison, we also show the peaks recovered with ILSM. As was expected, LSM [2] and ILSM yield identical estimates: $\hat{\beta}_0 = \overline{\beta_0} = 99.33$, $\hat{\beta}_1 = \overline{\beta_1} = 44.13$. The standard deviations of the estimation errors given by ILMS and MLS coincide: $\Delta\beta_0 = 1.8$ and $\Delta\beta_1 = 1.6$; i.e., the estimated values are comfortably within the two standard deviations (≈95% confidence level) from their true values.

In the second example, we use the same random variable as in the first one but with the different values of the parameters. We have changed the following quantities: $i_1 = 60$, $\beta_0 = 800$, $\beta_1 = 5$. Figure 2 shows the obtained fit. We do not present separate peaks because the second one is outside of the plot scale.

Using LSM we obtained the following 68% error bars for the peak amplitudes: $\hat{\beta}_0 = 796.6 \pm 5.15$ and $\hat{\beta}_1 = 1.131676 \pm 2.77$. Therefore, even at the 68% confidence level, the difference between the second peak amplitude and zero is not statistically significant. At the same confidence level, ILSM yields the following error bars: $\hat{\beta}_0 = 795.7 \pm 3.4$ and $\hat{\beta}_1 = 2.05 \pm 1.44$. At a confidence level of 95%, ILSM yields $\hat{\beta}_0 = 795.7 \pm 6.6$ and $\hat{\beta}_1 = 2.05^{+3.2}_{-1.95}$. Thus, a change in the model parameters may lead to a situation where the deviation of the estimated parameter $\hat{\beta}_1$ from 0 is not statistically significant. In such cases, ILSM give a reliable upper bound for this parameter. In our example, it suggests that the second peak amplitude does not exceeds a particular value. Both examples considered in this section indicate that ILSM, even in standard problems of finding parameter values from experimental data, is not worse than conventional LSM.

4. CONCLUSIONS

The numerical computations within the proposed ILSM are much easier to perform for low-quality experimental data, i.e., when the measurement errors are large. It is also possible to combine ILSM with the conventional LSM. In this case, one can estimate some of the easier-to-deal-with model parameters using LSM, while the remaining parameters (where it is necessary to exclude unphysical values of the estimated parameters) are estimated using ILSM.

Another advantage of the proposed method of parameter estimation is the possibility of calculating complicated functions of experimental parameters $\hat{\beta}_i$.

REFERENCES

1. Borovkov, A.A., *Matematicheskaya statistika, otsenka parametrov, proverka gipotez* (Mathematical Statistics, Parameter Estimation, and Hypothesis Testing), Moscow: Nauka, 1984.
2. Varden, B.L., *Matematicheskaya statistika* (Mathematical Statistics), Moscow: Izd-vo Inostr. Lit, 1960.
3. Gnedenko, B.V. and Kolmogorov, A.N., *Predel'nye raspredeleniya dlya summ nezavisimykh sluchainykh velichin* (Limiting Distributions for Sums of Independent Random Variables), Moscow: Izd-vo Tekhn.-Teoret. Lit, 1949.
4. Fisher, R.A., *Philos. Trans. R. Soc. London, Ser. A*, 1922, vol. 222, p. 309.