DOI 10.34185/1562-9945-5-142-2022-04
UDC 004.02:519.254

S.M. Vovk

TWO-STAGE METHOD FOR PARAMETER ESTIMATING IN UNKNOWN NOISE ENVIRONMENT

Abstract. This paper addresses the parameter estimation problem in the case of an unknown noise environment. A two-stage method consisting of the tuning and estimating stages is proposed for solving this problem. At the tuning stage, the estimator is tuned to the noise environment by minimizing the estimation root-mean-square error for a known data fragment in the space of the three free parameters; these free parameters allow modifying the behavior of the minimization problem objective function. At the estimating stage, estimation is carried out by solving the corresponding minimization problem using already tuned free parameters. The features of the proposed method, including algorithms for the tuning and estimating stages, are demonstrated for estimating a Gaussian pulse that slowly moves in the unknown noise environment and locates on a known constant background. The numerical simulations confirm the high performance of the proposed method.

Keywords: parameter estimation, cost function, Gaussian pulse.

Introduction. Classical estimation methods use an assumption about the completeness of data statistics. These methods solve the estimation problem in a noise environment formed by additive noise with a known distribution. A theoretical basis of these methods is the maximum-likelihood criterion, where the likelihood function is a common probability density function to be maximized [1] – [2]. The well-known results of this theory include the least-squares method and the least-absolute deviation method. These methods obey the assumption that data samples are independent and identically Gaussian or Laplacian distributed, respectively. The maximum-likelihood criterion has remarkable achievements related to the myriad and meridian filterings, which are generalized in [3]. These methods assume that the data samples are independent and identically distributed according to the Cauchy and Meridian statistics. They produce the so-called myriad (or Cauchy [4]) estimator and meridian estimator, respectively. Because of the impulsive nature of the Cauchy and Meridian noise, these estimators belong to the set of robust estimators.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Robust estimation methods use the incompleteness of data statistics [5] – [9]. They solve the estimation problem in the noise environment formed by additive noise and anomalous values. A theoretical basis of these methods is the generalized maximum-likelihood criterion that does not have a strictly probabilistic sense [7]. For that reason, the corresponding estimators usually are referred to as the "maximum-likelihood type" estimators or the M-estimators [6]. In contrast to the classical approach, the robust one requires choosing a cost (loss) function that should keep the data (inliers) and reject the anomalous values (outliers). The appropriate cost functions used for these purposes have been proposed by Tukey [6], Hampel [7], Andrew [8], Meshalkin [9], Geman and McClure [10], Demidenko [11], as well as by other authors [12] – [13]. Assembling various cost functions into a common set, which we call the "superset of cost functions," eliminates the redundancy of well-known cost functions and provides a wide range of possible solutions, including the mean, median, myriad, and meridian [14]. The modified version of this superset is given in [15]. An advantage of using the superset of cost functions is the possibility of tuning the M-estimator to the current noise environment.

The purpose of this paper is to present the two-stage method for parameter estimating in the unknown noise environment under the general assumption about the symmetry of noise distribution. This method uses the generalized maximum-likelihood criterion and the superset of cost functions. Through this paper, we show the features and performance of the proposed method by solving the problem of estimating a slowly moving Gaussian pulse located on the known constant background, where a mixture of Cauchy noise and positive outliers simulates the unknown noise environment.

Problem formulation. Let for a given observation interval X the observed data $g(x)$ are the sum of a given data model $s(x;\theta)$, which specified up to a vector $\theta=(\theta_1,\ldots,\theta_M)$ of the unknown parameters $\theta_1,\ldots,\theta_M$, and a random distortion $\xi(x)$, which describes the unknown noise environment, i.e.,

$$g(x)=s(x;\theta)+\xi(x), \tag{1}$$

where $x \in X$, and M is given. The estimation problem is to obtain the estimates of $\theta_1,\ldots,\theta_M$. Using (1), this problem can be formulated as the optimization problem:

$$\widehat{\theta}=\arg\min_{\theta} F[g(x)-s(x;\theta)]; \quad x \in X, \tag{2}$$

where F denotes an error functional that defines the estimation quality, and $\widehat{\theta}=(\widehat{\theta}_1,\ldots,\widehat{\theta}_M)$ is a vector of the estimates $\widehat{\theta}_1,\ldots,\widehat{\theta}_M$ for $\theta_1,\ldots,\theta_M$. For the discrete case, the problem (2) has the form:

$$\widehat{\theta} = \arg\min_{\theta} \sum_{n=1}^{N} \rho[g_n - s_n(\theta)]\,; \quad x \in X\,, \tag{3}$$

where $g_n = g(x_n)$, $s_n(\theta) = s(x_n;\theta)$, $x_n = n \cdot \Delta x$, Δx is a sampling step size, and N is a number of data samples on the observation interval X. From (3), it follows that the value of F is the sum of the values of cost function $\rho(\ldots)$ for each n-th data sample, and these values depend on θ. Since the cost function forms an objective function to be minimized, the problem of parameter estimation in the unknown noise environment consists of the problem of choosing the best cost function for the noise environment and the problem of parameter estimation after the choice of the best cost function.

Analysis of recent research and publications. A classical approach to the parameter estimation problem uses two main assumptions. Firstly, it assumes that the statistical nature of random distortions is known completely. Secondly, it assumes that each parameter to be estimated takes a single value. Within the framework of estimation theory, these assumptions dictate the search for a global minimum (or a global maximum) of an estimation quality indicator. Usually, the unconditional average risk is taken as this indicator. For a given loss (cost) function, the minimum of this risk corresponds to a Bayesian estimate. Using the "0-1" loss function and assuming that the probability density of the unknown parameter is a constant, its Bayesian estimate transfers to the maximum likelihood estimate [16]. The latter is widely used to obtain optimal or suboptimal solutions. For example, under the assumption of Gaussian data distribution, the corresponding maximum likelihood criterion becomes a quadratic one, leading to the least-squares problem [17]. An advantage of this problem is that, for the case of the data model described by a constant function, the least-squares problem leads to the analytical solution, yielding the arithmetic mean of data samples as the estimate. The assumption about the Laplace distribution of data samples leads to the least-absolute deviations problem, which gives a median value of data samples as the estimate for the same data model described by the constant function [18]. However, the quadratic or other convex functions are unsatisfactory for processing the data containing anomalous values [19]. In this case, robust approach based on the generalized maximum likelihood criterion is used. For the vector θ of unknown parameters of data model, it has the form of the following minimization problem:

$$\min_{\theta} \sum_{n=1}^{N} \rho[r_n;\theta]\,, \tag{4}$$

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

where r_n is a residual between the observed data and data model for a given n, and ρ is the cost function [6]-[7]. Usually, the robust approach (4) is implemented by the cost function with horizontal asymptotes [9].

To obtain a wide range of cost functions, a "superset of cost functions" has been proposed [14]. This "superset" is defined by the cost function:

$$\rho_S(x) = k_S[(1+|x/\alpha|^q)^{\beta/q} - 1], \tag{5}$$

where $\alpha > 0$ is a smoothing parameter with the same unit as x; q is a smoothing degree parameter ($0 < q < \infty$); β is a shape parameter ($-\infty < \beta \le 1$; $\beta < q$); $k_S(x) = 1/[(1+|x_0/\alpha|^q)^{\beta/q} - 1]$ is a constant, which is used to transform the cost function within the "superset" framework; x_0 is a normalization point, where $\rho_S(x_0) = 1$. Let $x_0 = 1$. The following limits are true:

$$\lim_{\alpha\to\infty} \rho_S(x) = |x|^q; \qquad \lim_{\alpha\to 0} \rho_S(x) = \begin{cases} |x|^\beta; & 0 < \beta \le 1 \\ \varphi(x); & -\infty < \beta \le 0 \end{cases};$$

$$\lim_{\beta\to\pm 0} \rho_S(x) = k_1 \cdot \ln(1+|x|^q/\alpha^q); \qquad \lim_{\beta\to-\infty} \rho_S(x) = \varphi(x);$$

$$\lim_{\beta\to 1} \rho_S(x) = k_2 \cdot [(1+|x|^q/\alpha^q)^{1/q} - 1]; \qquad \lim_{\substack{\beta\to-2 \\ q\to 2}} \rho_S(x) = k_3 \cdot \frac{x^2}{x^2+\alpha^2}, \tag{6}$$

where $\varphi(x) = \operatorname{sgn}^2(x)$; $k_1 = 1/\ln(1+1/\alpha^q)$; $k_2 = 1/[(1+1/\alpha^q)^{1/q} - 1]$; $k_3 = (1+\alpha^2)$.

The modified version of the "superset" is given in [15]. It has been obtained by equalizing behavior of cost function at zero. This modification expands the "superset" by a generalized Meshalkin's cost function [11]:

$$\rho_M(x) = [1-\exp(-|x/\alpha|^q)]/[1-\exp(-|x_0/\alpha|^q)], \tag{10}$$

where the value of α must be recalculated concerning the value of α for the original "superset." For the modified superset, it holds: $\lim_{\beta\to-\infty} \rho_{Sm}(x) = \rho_M(x)$, where $\rho_{Sm}(x)$ denotes the cost function of the modified superset. From (10), it follows that $\lim_{\alpha\to 0} \rho_M(x) = \varphi(x)$, $\lim_{\alpha\to\infty} \rho_M(x) = |x|^q$, and $\lim_{q\to 0} \rho_M(x) = \varphi(x)$, where $x_0 = 1$. Besides this, if $0 < \alpha < x_0$, then $\lim_{q\to\infty} \rho_M(x) = \rho_{pit}(x)$, where $\rho_{pit}(x)$ has the shape of a "rectangular pit" of 2α wide, i.e.,

$$\lim_{q\to\infty} \rho_M(x)\Big|_{|x|<\alpha_M} = 0; \ \lim_{q\to\infty} \rho_M(x)\Big|_{|x|=\alpha_M} = 1 - 1/e \approx 0{,}63; \ \lim_{q\to\infty} \rho_M(x)\Big|_{|x|>\alpha_M} = 1, \tag{11}$$

where e is Euler's number. The use of (11) changes the maximum generalized likelihood method into the maximum histogram method, where the histogram bin width is equal to 2α. Since the "superset" covers many symmetric cost functions, it can be used to tune the estimation to the noise environment with the symmetric noise distribution and with the outliers to be rejected.

Theory of the two-stage estimation method. The proposed two-stage estimation method uses the generalized maximum likelihood criterion and the superset of cost functions. There are two ways to apply this method. For the first one, it needs to set the three free parameters α, β, and q to a priori known values. For the second one, it needs to tune these three free parameters to the current noise environment. The latter we discuss for a Gaussian pulse that slowly moves in the unknown noise environment and locates on the known constant background. Here, we assume that the observed data contain a background fragment with unknown location in the measuring space, and a sufficient number of data realizations is available. Then, by estimating the background at the tuning stage, we choose the best values of free parameters, thereby realizing some training procedure. At the estimating stage, we estimate the location, amplitude, and half-width of the Gaussian pulse by the iterative algorithm proposed.

Let the data model is:

$$s(x; A_0, A, m, \sigma_p) = A_0 + A \cdot \exp[-(x-m)^2/(2\sigma_p^2)], \tag{12}$$

where A_0 is a known constant background, and A, m, σ_p are the Gaussian pulse parameters to be estimated. Let m is slowly changed in time comparing to the data sampling rate, and the length of the observation interval X is larger than the half-width σ_p of the Gaussian pulse in several times. Also suppose that $\sigma_p >> \Delta x$, $m \in X$, and A is within a known range. In such a case, we have a Gaussian pulse, which has a new location for each data measuring; we call this pulse a "moving Gaussian pulse." By using (12) for the discrete case, we state the following two estimation problems. The first one is:

$$\widehat{A}_0 = \arg\min_{A_0} \sum_{n=1}^{N} \rho_S \{g_n - A_0 - A\exp[-(x_n - m)^2/(2\sigma_p^2)]\}, \tag{13}$$

where ρ_S denotes the superset cost function, which depends on the three free parameters α, β, q [14]. Assuming $A = 0$ and excluding the Gaussian pulse model from (13), we have the first problem in the following form:

$$\widehat{A}_0 = \arg\min_{A_0} \sum_{n=1}^{N} \rho_S [g_n - A_0]. \tag{14}$$

The problem (14) describes the case of a "pure" noise environment when observed data do not contain a Gaussian pulse to be estimated. However, below we ap-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

ply (14) both in this case and in the case of the presence of the Gaussian pulse. Solving (14) gives the estimate $\widehat{A}_0$ of the background parameter A_0 for the given values of free parameters α, β, and q. Since A_0 has a known value, the best values of free parameters should minimize the estimation error. Note that subtracting the known constant background from observed data gives a similar problem for zero background estimating.

The second problem is the following:

$$(\widehat{A}, \widehat{m}, \widehat{\sigma}_p) = \arg\min_{A,m,\sigma_p} \sum_{n=1}^{N} \rho_S \{g_n - A_0 - A\exp[-(x_n - m)^2/(2\sigma_p^2)]\}, \tag{15}$$

This is the Gaussian pulse estimation problem under the assumption that values are known for A_0, α, β and q.

Let us consider the tuning stage. Substituting (5) into (14) yields:

$$\widehat{A}_0 = \arg\min_{A_0} \left\{ k_S \cdot \sum_{n=1}^{N} [(1 + |g_n - A_0|^q / \alpha^q)^{\beta/q} - 1] \right\}, \tag{16}$$

where $\widehat{A}_0$ is the estimate of A_0. For most cases, the objective function in (16) is a non-convex. Therefore, its minimization needs using the zero-order optimization methods. Naturally assuming $\widehat{A}_0 \in [g_{\min}, g_{\max}]$, where $g_{\min} = \min_{1 \le n \le N}(g_n)$ and $g_{\max} = \max_{1 \le n \le N}(g_n)$, we offer to solve the minimization problem (16) by searching for the best value of A_0 either among data samples or given values. The former is more attractive for the following reason. Because the value of each data sample turns into zero at least one term of the objective function in (16), such a value can be considered a "quasi-optimal" one. There are no more than N quasi-optimal values for the problem (16). Besides this, if $q < 1$, the quasi-optimal value coincides with the optimal value that minimizes the objective function in (16) (although, for the given values of α, β, q this optimal value of A_0 may not coincide with the true value of A_0).

The tuning stage is implemented by training performed over a given number of random data realizations and aims to obtain the best values for α, β, q. For training, we use the Mean-Square-Error (MSE) criterion. Let J be a number of data realizations used for training. The MSE, which we denote by the ε symbol, is the three-dimensional function for a fixed value of J, i.e.,

$$\varepsilon(\alpha, \beta, q) = \left[\frac{1}{J} \sum_{j=1}^{J} |\widehat{A}_0(j; \alpha, \beta, q) - A_0|^2 \right]^{1/2}, \tag{20}$$

where $\widehat{A}_0(j; \alpha, \beta, q)$ is the estimate of A_0 for the j-th realization and given α, β, q.

The MSE minimum points to the best values $\widehat{\alpha},\widehat{\beta},\widehat{q}$, i.e.,

$$(\widehat{\alpha},\widehat{\beta},\widehat{q}) = \arg\min_{\alpha,\beta,q} \varepsilon(\alpha,\beta,q), \tag{21}$$

For tuning, it needs to choose the sequences of free parameters values. Assuming that the scale of inlier distortions is less than 1, we propose the following ones. For α, the sequence is $(10^{-4},10^{-3},10^{-2},10^{-1},1,10)$. For β, the sequence is $(-\infty,-16,-4,-2,-1,-1/2,-1/4,-1/8,0,1/8,1/4,1/2,1)$. For q, the sequence is $(1/2,1,3/2,2,5/2,3,4,5,10)$.

Let us consider the estimating stage. Substituting (5) into (15) yields:

$$(\widehat{A},\widehat{m},\widehat{\sigma}_p) = \arg\min_{A,m,\sigma_p} k_S \sum_{n=1}^{N} (\{1+|g_n - A_0 - A\exp[-(x_n-m)^2/(2\sigma_p^2)]|^q/\alpha^q\}^{\beta/q} - 1), \tag{22}$$

where α,β,q are already tuned parameters and A_0 is known. To solve (22), we propose to use a zero-order minimization method that consists in generating the initial estimates and their refining.

Generating the initial estimates is to obtain the approximate values of A,m,σ_p. To generate the initial estimates, we assume that each of the data samples $g_n; n=1,\ldots,N$ can specify the location and the amplitude of the Gaussian pulse. In that case, the problem reduces to estimate the half-width σ_p. To solve this problem, we construct a set of trial values for σ_p and select such a trial value that minimizes the objective function presented in (22). We include in this set the following values. The first value is $\sigma_p = \infty$, indicating the transformation of the Gaussian pulse into a constant background. Although it contradicts the assumption about the presence of the Gaussian pulse, it can be used to calculate a maximum value of the objective function for the given data sample. Other trial values are generated by the formula:

$$\widehat{\sigma}_p = \frac{|j-i|\cdot\Delta x}{\sqrt{2\ln[(g_i - A_0)/(g_j - A_0)]}} \tag{23}$$

under the condition that $(g_i - A_0)/(g_j - A_0) > 1$, where i is the index of the data sample g_i considered as a center of the Gaussian pulse, and j is the index of some other data sample g_j, for which the inequality mentioned above holds. If, for the given g_i, it is impossible to find $g_j; j \neq i$ such that $(g_i - A_0)/(g_j - A_0) > 1$, then g_i is excluded from contenders to be a center of the Gaussian pulse, and the trial value is not generated for σ_p. However, if for the given data sample g_i there is $g_j; j \neq i$ such

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

that the inequality: $(g_i - A_0)/(g_j - A_0) > 1$ holds, the corresponding trial value is generated for σ_p. After obtaining the set of trial values, we select among them such a value that minimizes the objective function in (22). Thus, for each data sample, we obtain the initial estimates of A, m, σ_p and the corresponding value of the objective function. Then, passing through all the data samples, we select the best initial estimates that give a minimum value of the objective function.

Refining of the estimates consists in iteratively refining the estimates of A, m, σ_p. To implement refining, we use the trial values calculated as:

$$\widehat{m} = i \cdot \Delta x - \widehat{\sigma}_p \sqrt{2\ln[\widehat{A}/(g_i - A_0)]}\ ;\ \text{if}\quad \widehat{A}/(g_i - A_0) > 1, \tag{24}$$

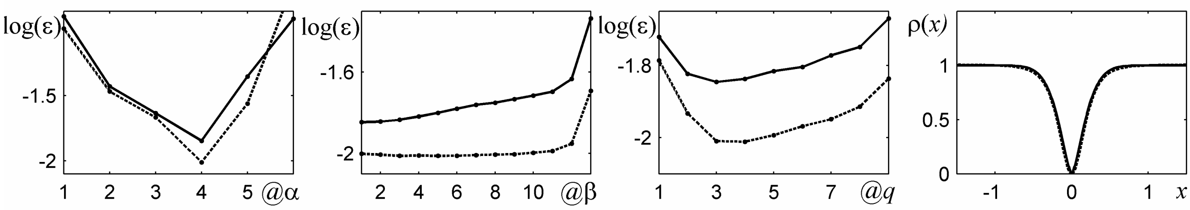
$$\widehat{A} = (g_i - A_0) \cdot \exp[(i \cdot \Delta x - \widehat{m})^2 / (2\widehat{\sigma}_p^2)], \tag{25}$$

$$\widehat{\sigma}_p = \frac{|\, i \cdot \Delta x - \widehat{m}\,|}{\sqrt{2\ln[\widehat{A}/(g_i - A_0)]}}\ ;\ \text{if}\quad \widehat{A}/(g_i - A_0) > 1, \tag{26}$$

where the current estimate is at the left side of the equal sign, and the previous estimates are at the right side of the equal sign. We refine the Gaussian pulse parameters estimates by the following algorithm. (1) Store the parameter estimate and the objective function value. (2) Create an array of trial values of the estimated parameter. (3) Check the performance of each trial value by calculating the objective function value and comparing it with the stored value of the objective function. If some trial value of the parameter minimizes the objective function, it is the estimate of this parameter at the current iteration. We use this algorithm iteratively through the following sequence of Gaussian pulse parameters: the location, the amplitude, and the half-width, and finish it when the estimates achieve the given accuracy.

Simulations. For simulations, we used (16) - (26), as well as the models of positive outliers and Cauchy noise. We set $N = 256$, $A_0 = 1$, $A = 1$, $m = l \cdot \Delta x$, $\sigma_p = 8 \cdot \Delta x$, $\Delta x = 1$, where l has an integer value from 60 to 160. We used the Cauchy noise distribution $p(\xi) = (\nu/\pi)\{1/[(\xi - \vartheta)^2 + \nu^2]\}$ with $\vartheta = 0$ and $\nu = 0.1$. For positive outliers, we set the probability of their occurrence to 0.1, and the uniform distribution of their amplitudes in the range from 0 to 5. For each new data realization containing a Gaussian pulse, changing l created the effect of Gaussian pulse moving. Therefore, averaging the data realizations will not be effective for estimating. Moreover, averaging the Cauchy noise is ineffective.

At the tuning stage, we simulated 1000 random realizations of the noise environment, formed by the Cauchy noise and positive outliers in both cases of the absence and the presence of the Gaussian pulse. For the latter case, we used different values of l obtained first by its decrementing by 1 from 160 to 60, and then by its incrementing by 1 from 60 to 160, etc. By estimating the constant background for various values of α, β, and q, we calculated the three-dimensional function (20) for each of the indicated cases. Fig. 1 presents these simulations, where Fig. 1a, Fig. 1b, and Fig. 1c show cross-sections of the decimal logarithm of the three-dimensional function (20) by the planes $\beta=\widehat{\beta}$ and $q=\widehat{q}$, by the planes $\alpha=\widehat{\alpha}$ and $q=\widehat{q}$, as well as by the planes $\alpha=\widehat{\alpha}$ and $\beta=\widehat{\beta}$, respectively. In these figures, the symbol @ denotes the ordinal number of the element in the given sequences for α, β, q, which starts from one and changes by one. Thus, if $@\alpha=1$, then $\alpha=10^{-6}$; if $@\alpha=2$, then $\alpha=10^{-5}$, etc. In a like manner, if $@\beta=1$, then $\beta=-\infty$, etc., as well as if $@q=1$, then $q=0.5$, etc. Fig. 1d shows the cost functions for values $\widehat{\alpha}=0.1$, $\widehat{\beta}=-4$, $\widehat{q}=2$ and $\widehat{\alpha}=0.1$, $\widehat{\beta}=-\infty$, $\widehat{q}=1.5$, which minimized the MSE for the indicated cases, when $J=1000$. Here, we see that these cost functions visually almost coincide. The obtained dependences of $\widehat{\alpha},\widehat{\beta},\widehat{q}$ and MSE minimum on the number J of random realizations show that tuning of α was the fastest, while tunings of β and q were slower in both cases.



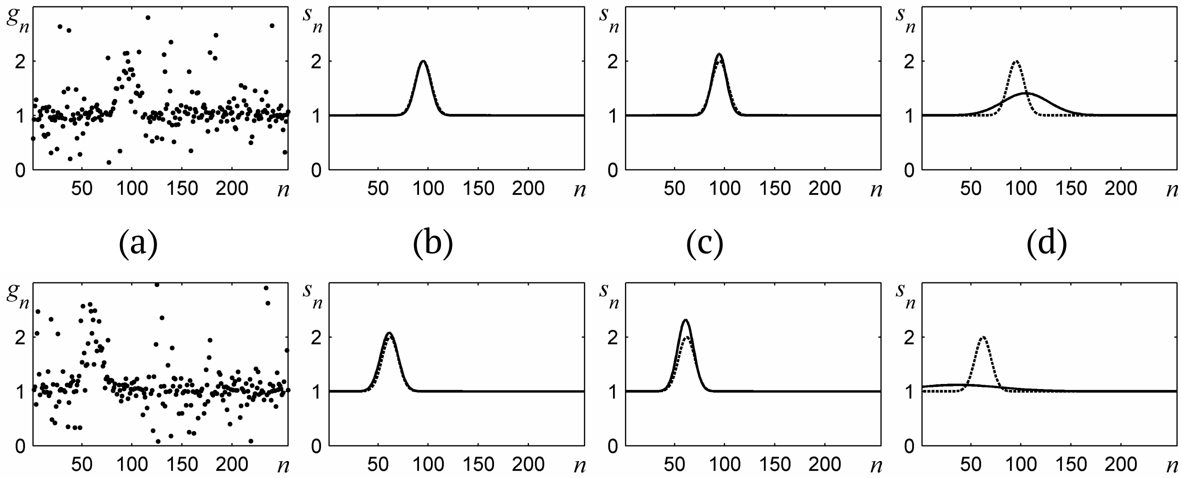
(a) (b) (c) (d)

Figure 1 – Tuning to the noise environment formed by the Cauchy noise and positive outliers: the log of MSE along α-axis (a), β-axis (b), q-axis (c), and the best cost functions (d); dotted and solid lines correspond to the absence and the presence of Gaussian pulse, respectively

The second stage of the proposed method is the estimating stage. Fig. 2 shows two examples of this stage for estimating the Gaussian pulse. Here, Fig. 2a and Fig. 2e show two different data realizations. Fig. 2b and Fig. 2f show the estimates obtained for "best" tuning (when $\alpha=\nu=0.1$, $\beta=-4$, $q=2$), where plots of the true Gaussian pulse and its estimates visually coincide. Fig. 2c and Fig. 2g present the estimates obtained for "good" tuning (when $\alpha=\nu=0.1$, $\beta=-\infty$, $q=1.5$). Fig. 2d and

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Fig. 2h show the estimates obtained for "bad" tuning (when $\alpha = \nu/100 = 0.001$, $\beta = -\infty$, $q = 1.5$). A full demonstration of this stage is available by the reference https://drive.google.com/file/d/1RVz9AKShqhWAUDpzGPffJC5INeCMtiY1.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 2 – Gaussian pulse estimation with known constant background: input data (a,e); the estimate (solid line) for “best” tuning (b, f), "good" tuning (c, g), and "bad" tuning (d, h); dotted line shows the true Gaussian pulse

Table 1 presents the average and maximal MSE values of the Gaussian pulse estimation obtained over 1000 realizations by using the mean, median, myriad, and meridian estimators, as well as by using the proposed estimator with the best, good, and two bad tunings, respectively. The plots in the table cells show the cost functions used for estimation. A feature of these cost functions is their normalization at $x_0 = 1$, i.e. $\rho_S(x_0) = 1$. Here, we see that the shape of the cost function and the scale of its argument are important at once.

Discussion. For considered example, the Gaussian pulse estimation should be performed for each data realization obtained after measuring. Unfortunately, the presence of the Gaussian pulse complicates tuning to the noise environment since the Gaussian pulse also contaminates it. However, Fig. 1 shows if the ratio of the observation interval length to the half-width of the Gaussian pulse is large enough, then the tuning gives the same results both in the absence and in the presence of the Gaussian pulse. Simulations also show that the tuning time can be reduced essentially by using the following algorithm. (1) Solve the three-dimensional tuning problem for α, β, q with a small number (e.g., 100) of realizations and take the best value of α. (2) Fixing this value of α, solve the two-dimensional tuning problem for

q and β with a moderate number (e.g., 200-500) of realizations and take the best value of q. (3) Fixing this value of q, solve the one-dimensional tuning problem for β with a large number (e.g., 500-1000) of realizations and take the best value of β.

Fig. 2 shows that the quality of estimation depends on the tuning quality. In this context, Table 1 shows that the best estimate corresponds to the "best" tuning (both in terms of the average value and maximal value of MSE). But for the "good" tuning, the estimate is slightly worse. Table 1 also shows that a small average value does not guarantee a small maximal value. In total, Fig. 2 and Table 1 confirm the effectiveness of the proposed approach for Gaussian pulse estimation for the case when the free parameters are properly tuned to the noise environment.

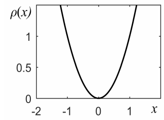
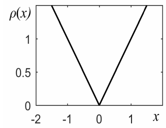
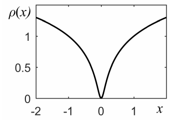
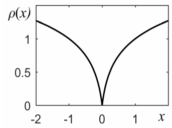
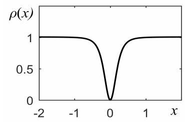
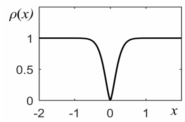
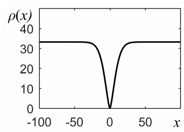
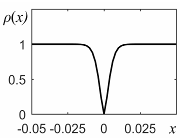
Problem (2) was formulated for a parametric data model without using the constraints on the values of data model parameters. Using such information and including it in the formulation of the estimation problem, we can improve the estimation quality. Besides this, if the unknown parameter takes several values, it is advisable to look for several local minima of the objective function.

Conclusion. The proposed two-stage method is an efficient method for parameter estimating of a given data model in an unknown noise environment. It can be used under the condition that it is possible to tune the parameter estimation procedure to the noise environment. It is shown that tuning can be performed on a fragment of known data. Further, subtracting the estimated data from the obtained data after the estimating stage allows refining tuning. The proposed method can be used to develop an intelligent system that can tune itself to the unknown noise environment and perform the best estimation of the parameters of a given data model.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Table 1

MSE of Gaussian pulse estimation

Estimator	Cost function	Average value of MSE	Maximal value of MSE
Mean		59.76	1502.44
Median		0.36	2.24
Myriad (α=ν, β=0, q=2)		0.50	3.00
Meridian (α=ν, β=0, q=1)		0.40	1.52
proposed: “best” tuning (α=ν, β=-4, q=2)		0.34	1.47
proposed: “good” tuning (α=ν, β=-∞, q=1.5)		0.37	2.39
proposed: “bad 1” tuning (α=ν*100, β=-∞, q=1.5)		1.19	22.08
proposed: “bad 2” tuning (α=ν/100, β=-∞, q=1.5)		3.39	20.17

REFERENCES

1. Smith B.W. Parameter estimation in the presence of measurement noise. International Journal of Control. 1966. Vol. 3, N. 4. P. 297-312.
2. Tong X., Gao F., Chen K., Cai D., Sun J. Maximum likelihood estimation in transformed linear regression with nonnormal errors. The Annals of Statistics. 2019. Vol. 47, N. 4. P. 1864-1892.
3. Carrillo R.E., Aysal T.C., Barner K. E. Generalized Cauchy distribution based robust estimation. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, 31 March – 4 April 2008. P. 3389–3392.
4. Balmer D.W., Boulton M., Sack R. A. Optimal solutions in parameter estimation problems for the Cauchy distribution. Journal of the American Statistical Association. 1974. Vol. 69, N. 345. P. 238-242.
5. Rousseeuw P.J., Hubert M. Anomaly detection by robust statistics. WIREs Data Mining Knowledge Discovery. 2018. Vol. 8, N. 2. P. 1-14.
6. Huber P. Robust statistics. New York: Wiley. 1981. 312 p.
7. Hampel F.R., Ronchetti E.M., Rousseeuw P. J., Stahel W. A. Robust statistics: the approach based on influence functions. Hoboken: Wiley, 2011. 502 p.
8. Andrews D. F. Bickel P. J., Hampel F. R., Huber P. J., Rogers W. H., Tukey J. W. Robust estimates of location. Survey and advances. Princeton : Princeton University Press, 1972. 372 p.
9. Shevlyakov G. L., Vil'chevski N. O. Robustness in data analysis: criteria and methods. Utrecht : VSP, 2002. 310 p.
10. Geman S., McClure D. Bayesian image analysis. An application to single photon emission tomography. Proceedings of the American Statistical Association, Statistical Computing Section. 1985. P. 12-18.
11. Vovk S., Borulko V. Family of generalized Demidenko functionals for robust estimation. Proceedings of the 17th Int. Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory, Tbilisi, Georgia, September 2012. P. 151–154.
12. Guo W., Lou Y., Qin J. et al. A Novel Regularization Based on the Error Function for Sparse Recovery. Journal of Scientific Computing. 2021. Vol. 87. 31.
13. Zhao Z., Wang S., Wong D., Sun C., Yan R., Chen X. Robust enhanced trend filtering with unknown noise. Signal Processing. 2021. Vol. 180, 107889.
14. Borulko V. F., Vovk S. M. Minimum-duration filtering. Radio Electronics, Computer Science, Control. 2016. N. 1. P. 7–14.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

15. Vovk S.M. General approach to building the methods of filtering based on the minimum duration principle. Radioelectronics and Communications Systems. 2016. Vol. 59, N. 7. P. 281–292.
16. Millar R.B. Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB. New York : Wiley, 2011. 376 p.
17. Wolberg J. Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments. Berlin : Springer-Verlag, 2005. 250 p.
18. Elgmati E.A., Gredni N.B. Quartile Method Estimation of Two-Parameter Exponential Distribution Data with Outliers. International Journal of Statistics and Probability. 2016. Vol. 5, N. 5. P. 12-15.
19. Chandola V. Banerjee A., Kumar V. Anomaly detection: A survey. ACM Computing Surveys. 2009. Vol. 41, N. 3. P. 15-58.

Received 20.06.2022.
Accepted 24.06.2022.

Двоетапний метод оцінювання параметрів у невідомому шумовому середовищі

Запропоновано двоетапний метод вирішення задачі оцінювання параметрів заданої моделі даних для випадку невідомого шумового середовища, який складається з етапів налаштування та оцінювання. На етапі налаштування оцінювач налаштовується на шумове середовище шляхом мінімізації середньоквадратичної помилки оцінювання для відомого фрагмента даних у просторі трьох вільних параметрів. На етапі оцінювання оцінювач розв'язує відповідну задачу мінімізації, використовуючи вже налаштовані вільні параметри. Наведено приклад оцінювання гауссівського імпульсу, який повільно рухається у невідомому шумовому середовищі.

Вовк Сергій Михайлович - к.ф.-м.н., доцент, Дніпровський національний університет, доцент кафедри комп'ютерних наук та інформаційних технологій.

Vovk Serhii Mikhailovich - Ph.D, Associate Professor, Oles Honchar Dnipro National University, Associate Professor of Department of Computer Science and Information Technology.