DOI 10.34185/1562-9945-6-125-2019-02
UDC 004.02

S.M. Vovk

PARAMETER ESTIMATION FOR COMPLICATED NOISE ENVIRONMENT

Abstract. For a complicated noise environment the use of M-estimator faces a problem of choosing a cost function yielding the best solution. To solve this problem it is proposed to use a superset of cost functions. The superset capabilities provide constructing a parameter estimation method for complicated noise environment. It consists in tuning the generalized maximum likelihood estimation to the current noise environment by setting values of three free superset parameters related to the scale, the tail heaviness and the form of noise distribution, as well as to the anomaly values that presence in data. In general case, this method requires to solve the optimization problem with a non-unimodal objective function, and it can be mostly implemented by using the zero-order optimization methods. However, if the noise environment has known statistics, the proposed method leads to the optimal estimation. If the noise environment is complicated or does not have a complete statistics, the proposed method leads to the more effective estimates comparing to those of mean, median, myriad and meridian estimators. Numerical simulations confirmed the method performance.
Key words: generalized maximum likelihood, cost function, robust estimation.

Introduction. The relevance of the parameter estimation problem for complicated noise environment is determined by the diversity of data, noise and anomaly models [1]. In such a case the traditional approach to parameters estimation is based on the M-estimation method [2]. However, within the M-estimation framework there is a problem of choosing a cost function that should provide the best solution for the current noise environment. Unfortunately, solving this problem is complicated by necessity to choose among a large number of known and pretty similar cost functions. For example, there are the Tukey's cost function [2], the Hampel's cost function [3], the Andrew's cost function [4], the Meshalkin's cost function [5], etc. To eliminate this shortcoming it is proposed to use a superset of cost functions [6]-[7], which generates various cost functions by setting values of three free parameters. In this case, the use of the generalized maximum likelihood method leads to a new method of parameter estimation. This method can provide the optimal data processing for the noise with a known statistical nature as well as it can provide the tuning to best results for the noise with unknown statistics. In the latter case, the tuning can be done by training.

Problem formulation. For simplicity of consideration, let a data model be a constant and let this constant is observed in additive contaminated noise. Then the problem to determine a constant value is a problem to estimate the location parameter θ for the noisy data elements $x_n; \quad n = 1,...,N$. Using the idea of M-estimation [2], this problem is:

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{N} \rho(x_i - \theta), \qquad (1)$$

where $\rho(x)$ is an arbitrary function [2], which is referred to as either the cost function [8], the loss function or the contrast function [5]. This function forms the objective function for a minimization problem. The problem formulation is to choose such a function $\rho(x)$ which can be tuned to the current noise environment. The goal of this research is to present the parameter estimation method for complicated noise environment.

Analysis of recent research and publications. In **[6]-[7]** the superset of cost functions is proposed. It is defined as:

$$\rho_S^{(\alpha,\beta,q)}(x) = k_S^{(\alpha,\beta,q)}[(1+|x/\alpha|^q)^{\beta/q} - 1], \qquad (2)$$

where α is a smoothing parameter ($\alpha > 0$); q is a smoothing degree parameter ($0 < q < \infty$); β is a parameter of the form of cost function ($-\infty < \beta \le 1$; $\beta < q$); $k_S^{(\alpha,\beta,q)}(x) = 1/[(1+|x_0/\alpha|^q)^{\beta/q} - 1]$ is a constant, which is necessary to perform the transformations of cost functions one to another within the superset framework; x_0 is a normalization point, at which the equality: $\psi_S^{(\alpha,\beta,q)}(x_0) = 1$ is ensured ($x_0 = 1$ is supposed usually). The superset has the following properties. Assuming $x_0 = 1$ and using (2), one can obtain that

$$\lim_{\alpha\to\infty} \rho_S^{(\alpha,\beta,q)}(x) = \lim_{\alpha\to\infty} \frac{(1+|x/\alpha|^q)^{\beta/q} - 1}{(1+|x_0/\alpha|^q)^{\beta/q} - 1} = \frac{|x|^q}{|x_0|^q} = |x|^q \qquad (3)$$

and

$$\lim_{\alpha\to 0} \rho_S^{(\alpha,\beta,q)}(x) = \lim_{\alpha\to\infty} \frac{(\alpha^q + |x|^q)^{\beta/q} - \alpha^\beta}{(\alpha^q + |x_0|^q)^{\beta/q} - \alpha^\beta} = \begin{cases} |x|^\beta; & 0 < \beta \le 1 \\ \chi(x); & -\infty < \beta \le 0 \end{cases}, \qquad (4)$$

where $\chi(x)=\begin{cases}0, & x=0\\ 1, & x\neq 0\end{cases}$ is the "zero-one" cost function called also the Titchmarsh's cost function [6]-[7]. Equation (3) indicates that if $\alpha\to\infty$, the limiting case of superset is the cost function with the form either of quasi-norm (for $0<q<1$) or norm (for $1\le q<\infty$) of the mathematical space. Equation (4) indicates that if $\alpha\to 0$, the limiting case of superset is the cost function with the form either of quasi-norm of L_β-space (for $0<\beta<1$) or norm of L_1-space (for $\beta=1$), and with the form of the 1-norm of L_0-space (for $-\infty<\beta\le 0$). Thus, the superset encompasses the whole set of possible norms and quasi-norms of L_p-space, where $0\le p<\infty$. In addition to this, from (2) one can obtain that

$$\lim_{\beta\to\pm 0}\rho_S^{(\alpha,\beta,q)}(x)=\frac{\ln(1+|x/\alpha|^q)}{\ln(1+|x_0/\alpha|^q)} \quad (5)$$

and

$$\lim_{\beta\to-\infty}\rho_S^{(\alpha,\beta,q)}(x)=\chi(x). \quad (6)$$

The cost function (5) with $q=1$ and $q=2$ becomes the cost function of Gonzalez-Arce [9] and Aysal-Barner [8], respectively. Thus, according to the value of β, the superset can be divided into the following main sets. They are [7]: 1) the set of the "q-smoothed root" cost functions (for $0<\beta\le 1$ and $\beta<q$), including the pseudo-Huber's cost function (for $\beta=1$ and $q=2$); 2) the set of the "q-smoothed logarithmic" cost functions (for $\beta=0$), including the Gonzalez-Arce's cost function and the Aysal-Barner's cost function; 3) the set of the "Demidenko's q-smoothed" cost functions [10] including the Geman-McClure's cost function [11] (for $\beta=-2$ and $q=2$); 4) the set that contains just the Titchmarsh's cost function.

The superset can be modified by equalizing the behavior of its cost functions at the zero. This expands the superset by including in it the generalized Meshalkin's cost function [7]

$$\rho_M^{(\alpha,q)}(x)=\frac{1-\exp(-|x/\alpha|^q)}{1-\exp(-|x_0/\alpha|^q)}. \quad (7)$$

where the value of α must be recalculated with respect to the original value of α for the superset. From (7) it follows if $0<q<\infty$ and $\alpha\to 0$, the cost function (7) tends to the cost function (6). But if $0<q<\infty$ and $\alpha\to\infty$, then (7) tends to $|x|^q$. On the other hand, if $\alpha>0$ and $q\to 0$, then (7) tends to (6) also. But if $0<\alpha<x_0<\infty$ and $q\to\infty$, then (7) tends to the "inverse" rectangular cost function, which has the shape of a rectangular pit of 2α wide, i.e.

$$\rho_{pit}(x)=\begin{cases}0, & |x|<\alpha \\ (e-1)/e\approx 0.63, & |x|=\alpha \\ 1, & |x|>\alpha\end{cases}. \tag{8}$$

The use of (8) transforms the maximum likelihood method into the histogram method, where the histogram bin width is equal to $\alpha/2$.

The modified superset of cost functions is obtained by modification of α value. This value is modified in such a way that the second derivative at zero for any cost function, which belongs to the superset, should be equal to the second derivative at zero of the "*q*-smoothed logarithmic" cost function. Using $x_0=1$, this is achieved by solving the non-linear equation [7]:

$$\alpha^q[(1+1/\alpha^q)^{\beta/q}-1]=(\beta/q)\alpha_{\log}^q\ln[1+1/\alpha_{\log}^q], \tag{9}$$

where $\alpha_{\log}$ is the given smoothing parameter for "*q*-smoothed logarithmic" cost function, and α is the smoothing parameter to be calculated from (9) for given values of $\alpha_{\log}$, β and q. Such calculations performed by the Newton's method usually require about 3-5 iterations. In particular, for $\alpha_{\log}=0.1$, $q=2$ and β values: 1, 0.5, 0, -1, -2, -100, the following α values are obtained: 0.024, 0.062, 0.100, 0.166, 0.220, 1.519 . It is seen that the third value of α, which corresponds to $\beta=0$, is equal to $\alpha_{\log}=0.1$. But some values of α have become less than $\alpha_{\log}$ while others have become larger.

To recalculate the value of α in (7) it is necessary to solve the non-linear equation:

$$\alpha^q[1-\exp(-1/\alpha^q)]=\alpha_{\log}^q\ln[1+1/\alpha_{\log}^q], \tag{10}$$

where α is unknown, $\alpha_{\log}$ is given, and $x_0=1$. The solution of (10) can also be obtained by the Newton method. However, for modified superset the generalized cost function (2) tends not towards the Titchmarsh's cost function, but rather towards the generalized Meshalkin's cost function with recalculated value of α. On the other hand, as mentioned above, the Meshalkin's cost function tends to the Titchmarsh's cost function as $\alpha\to 0$. However, for large enough value of q (e.g., $q > 50$), superset modification loses any sense since in this case the cost functions of the modified superset are practically the same as the cost functions of the original superset.

Summarizing, it can be concluded that considered superset covers a very large number of cost functions. Therefore, this superset can be used to tune the maximum generalized likelihood method to the noise environment of various types.

Main results. Based on (1) and (2), the proposed estimation method consists in solving the following minimization problem:

$$\min_{\theta}\left\{k_S^{(\alpha,\beta,q)} \cdot \sum_{n=1}^{N}[(1+|x_n-\theta|^q/\alpha^q)^{\beta/q}-1]\right\}, \tag{11}$$

where values of α, β, and q should be adjusted to the current noise environment. The processing of any data sequence $x_i; i = 1,\ldots,I$ consists in using a digital window of length N that slides through the data sequence and produces the current output values by solving the minimization problem (11).

Since the objective function in (11) is non-unimodal, it is necessary to use the zero-order optimization methods to minimize it. Generally saying, this is a difficult problem. But its computational complexity can be reduced by using a quasi-optimal value of θ which coincides with the value of some data element. Such quasi-optimal value turns at least one of the terms of (11) into zero. In addition, if $q < 1$, the quasi-optimal value coincides with the optimal value. However, for this case the first derivative of objective function will not be zero at its minimum, since it will have a discontinuity at this minimum. Therefore, in general case it is impossible to apply the basic approach, which consists in replacing the optimization problem (11) by the problem of solving a nonlinear equation [3].

Using (5) as well as using the inequalities: $k_S^{(\alpha,\beta,q)} > 0$ for $0 < \beta < 1$ and $k_S^{(\alpha,\beta,q)} < 0$ for $-\infty < \beta < 0$, the problem (11) can be represented as an union of the following three problems:

$$\min_{\theta}\left\{\sum_{n=1}^{N}[(1+|x_n-\theta|^q/\alpha^q)^{\beta/q}]\right\};\quad 0 < \beta < 1, \tag{12}$$

$$\min_{\theta}\left\{\sum_{n=1}^{N}\ln(1+|x_n-\theta|^q/\alpha^q)\right\};\quad \beta \pm 0, \tag{13}$$

$$\max_{\theta}\left\{\sum_{n=1}^{N}\frac{1}{(1+|x_n-\theta|^q/\alpha^q)^{|\beta|/q}}\right\};\quad -\infty<\beta<0, \qquad (14)$$

which can be solved in a parallel way. For each of (12) – (14), one can always specify an admissable search area of global minimum.

Simulations. Figure 1 shows the plots for the sum of the constant that equals to 1 and the random realization of noise environment numbered from #1 to #6, where dots denote data element values. Noise environment number (#) has the following sense. #1 denotes the sum of Gaussian noise and random wide Gaussian pulse of unit amplitude, where the latter has the half-width, which is equal to the window length for 15 consecutive samples, and the uniform distributed location within the range from 30th to 60th sample. #2 denotes the sum of Gaussian noise and the interference in the form of a random sequence of narrow positive Gaussian pulses. The amplitude, location, and half-width of these pulses are uniformly distributed in [0, 2], [1, 101] and [0, 2] intervals, respectively. #3 denotes the sum of Gaussian noise and outliers with a probability of their occurrence p = 0.56 and with the amplitudes uniformly distributed in [2, 3] interval. #4, #5 and #6 denote the sum of the Cauchy noise with the same anomalies that correspond to the #1, #2 and #3, respectively. During simulations, the Gaussian noise: $p(\xi)=\frac{1}{\sqrt{2\pi}\sigma}\cdot\exp[-(\xi-m)^2/(2\sigma^2)]$ had a zero mean $m=0$ and a standard deviation $\sigma=0.1$; the Cauchy noise: $p(\xi)=\frac{\nu}{\pi}\cdot\frac{1}{(\xi-\vartheta)^2+\nu^2}$ had a zero location parameter $\vartheta=0$ and a scale parameter $\nu=0.1$.

Figure 2 shows the objective functions corresponding to the random realizations in Figure 1. They are constructed for the following values of free parameters. The objective function represented by curve 1 is constructed for $\alpha=10$, $\beta=-4$ and $q=2$. Since $\alpha>>\sigma$ and $\alpha>>\nu$, this objective function is quadratic in all plots in Figure 2. The objective function represented by curve 2 is constructed for $\alpha=10$, $\beta=-4$ and $q=1$. Since here $\alpha>>\sigma$ and $\alpha>>\nu$, this objective function is close to the sum of cost functions having the form $\rho(x)=|x|$. The objective function represented by curve 3 and constructed for $\alpha=0.1$, $\beta=0$, $q=2$ is the sum of cost functions having the form

$\rho(x)=\ln(x^2+\alpha^2)$. Finally, the objective function represented by curve 4 corresponds to the best values of free parameters. They are: $\alpha=0.1$, $\beta=-16$, $q=10$ (for #1 and #2); $\alpha=0.1$, $\beta=-16$, $q=2$ (for #3); $\alpha=0.01$, $\beta=-16$, $q=1$ (for #4); $\alpha=0.01$, $\beta=-16$, $q=1.5$ (for #5); $\alpha=0.1$, $\beta=-4$, $q=2$ (for #6). It can be seen that the global minimum of curve 4 indicates to the desired value of constant value almost exactly, while for the other curves the global minimum indicates to it in most cases inaccurately.

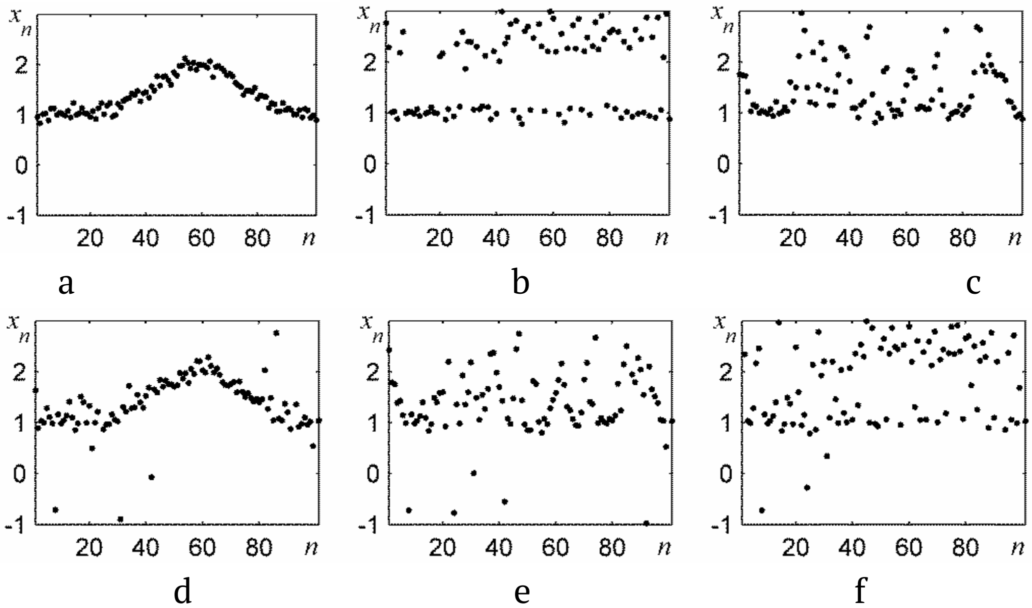


Figure 1 - Random realization of noise environment:
a – f are corresponded to #1 – #6

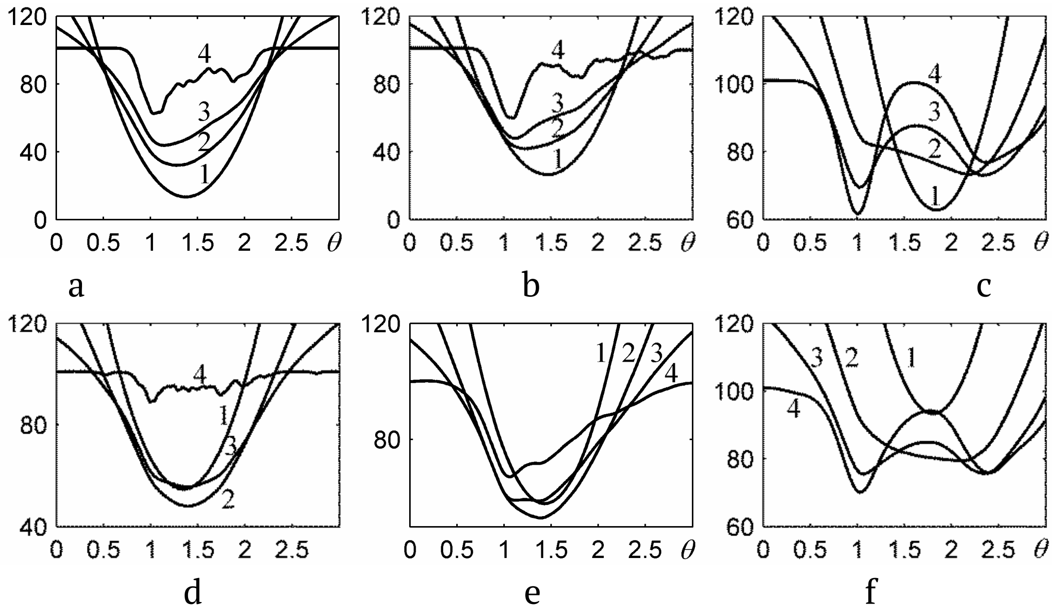
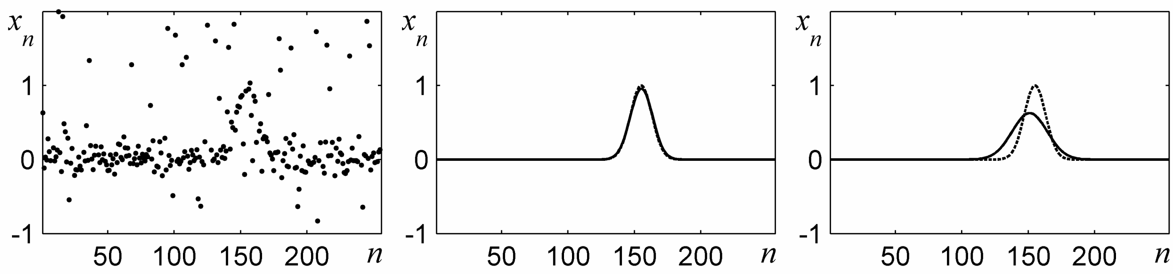


Figure 2 - Objective functions for noise environment:
a – f are corresponded to a – f in Figure 1

Figure 3 shows the Gaussian pulse estimation in a complicated noise environment caused by Cauchy noise ($\nu = 0.1$) and positive outliers (their probability is 0.1). Figure 3a shows the input data sequence; Figure 3b and Figure 3c show the estimation results obtained by the "good" tuning ($\alpha = \nu$, $\beta = 0$, $q = 2$) and by the "bad" tuning ($\alpha = 10\nu$, $\beta = 0$, $q = 2$), respectively.



a b c

Figure 3 - Gaussian pulse estimation: a – input data sequence; b, c – "good" and "bad" tuning (pulse estimation – solid, true pulse – dashed)

Discussion. The use of cost function $\rho(x) = |x|^2$ always results with a large estimation error (curve 1 in Figure 2). However, the use of $\rho(x) = |x|$ sometimes gives the larger error than that for $\rho(x) = |x|^2$ (curve 2 in Figure 2c and Figure 2f). The use of $\rho(x) = \ln(x^2 + \alpha^2)$ gives acceptable results (curve 3) for the noise environment of #1, #2 and #3. But for the noise environment of #4, #5 and #6 its application cannot be considered satisfactory. On the other hand, the use of cost functions, which are obtained by the proposed estimation method, led to the best results (curve 4). Gaussian pulse estimation by the 100 random trials always gave a consistently good result.

Thus, the results of numerical simulations confirmed the feasibility and effectiveness of the proposed estimation method. This is achieved by adjusting the values of its free parameters for a given noise environment. Therefore, when working with a specific noise environment instead of obtaining the statistics necessary to determine the noise distribution and then choosing the appropriate cost function, one can tune the values of free parameters and use the proposed estimation method.

Conclusions. For a complicated noise environment, it is suitable to use the estimation method based on the generalized maximum likelihood criterion with the superset of cost functions. Its efficiency is achieved by tuning the superset free parameters to the current noise environment. If the noise environment is simple and has known statistics, this method leads to the optimal estimation. If the noise environment is complicated and does not have a complete statistical description, this method leads to more effective estimates comparing to those of mean, median, myriad and meridian estimators.

ЛИТЕРАТУРА / ЛІТЕРАТУРА

1. Rousseeuw P.J. Anomaly detection by robust statistics / P.J. Rousseeuw, M. Hubert // WIREs Data Mining Knowledge Discovery. – 2018. – V. 8. – N. 2. – P. 1-14. DOI= https://doi.org/10.1002/widm.1236.
2. Huber P. Robust statistics / P. Huber, E. M. Ronchetti. – Hoboken: Wiley. – 2009. – 370 p. DOI: 10.1002/9780470434697.
3. Hampel F. R. Robust statistics: the approach based on influence functions / F.R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel. – Hoboken: Wiley. – 2011. – 536 p. DOI=10.1002/9781118186435.
4. Andrews D. F. Robust estimates of location: survey and advances / D. F. Andrews, P.J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, J.W. Tukey. – Princeton, NJ: Princeton University Press. – 1972. – 372 p.
5. Shevlyakov G. L. Robustness in data analysis: criteria and methods / G. L. Shevlyakov, N. O. Vilchevski. – Utrecht: VSP. – 2002. – 310 p.
6. Borulko V.F. Minimum-duration filtering / V. F. Borulko, S. M. Vovk // Radio Electronics, Computer Science, Control. — 2016. — N. 1. — C.7-14. DOI: 10.15588/1607-3274-2016-1-1.
7. Vovk S.M. General approach to building the methods of filtering based on the minimum duration principle / S. M. Vovk // Radioelectronics and Communications Systems. – 2016. – V. 59. – N. 7 – P. 281-292.
DOI: 10.3103/S0735272716070013.
8. Aysal T. C. Meridian filtering for robust signal processing / T. C. Aysal, K. E. Barner // IEEE Tr. on Signal Processing. – 2007. – V. 55. – N. 8. – P. 3949–3962.
9. Gonzalez J. G. Optimality of the myriad filter in practical impulsive-noise environments / J. G. Gonzalez, G. R. Arce // IEEE Tr. on Signal Processing. – 2001. – V. 49. – N. 2. – P. 438–441. DOI: 10.1109/78.902126.

10. Vovk S. Family of generalized Demidenko functionals for robust estimation / S. Vovk, V. Borulko // Proc. XVIIth Int. Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory (Tbilisi, Georgia, September 24-27, 2012). – IEEE. – 2012. – P. 151–154.
11. Geman S. Bayesian image analysis. An application to single photon emission tomography / S. Geman, D. McClure // Proc. of the American Statistical Association, Statistical Computing Section. – 1985. – P. 12-18.

REFERENCES

1. Rousseeuw P. J. Anomaly detection by robust statistics / P. J. Rousseeuw, M. Hubert // WIREs Data Mining Knowledge Discovery. – 2018. – V. 8. – N. 2. – P. 1-14. DOI= https://doi.org/10.1002/widm.1236.
2. Huber P. Robust statistics / P. Huber, E. M. Ronchetti. – Hoboken: Wiley. – 2009. – 370 p. DOI: 10.1002/9780470434697.
3. Hampel F. R. Robust statistics: the approach based on influence functions / F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel. – Hoboken: Wiley. – 2011. – 536 p. DOI=10.1002/9781118186435.
4. Andrews D. F. Robust estimates of location: survey and advances / D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, J.W. Tukey. – Princeton, NJ: Princeton University Press. – 1972. – 372 p.
5. Shevlyakov G. L. Robustness in data analysis: criteria and methods / G.L. Shevlyakov, N. O. Vilchevski. – Utrecht: VSP. – 2002. – 310 p.
6. Borulko V.F. Minimum-duration filtering / V. F. Borulko, S. M. Vovk // Radio Electronics, Computer Science, Control. — 2016. — N. 1. — C.7-14.
DOI: 10.15588/1607-3274-2016-1-1.
7. Vovk S.M. General approach to building the methods of filtering based on the minimum duration principle / S. M. Vovk // Radioelectronics and Communications Systems. – 2016. – V. 59. – N. 7 – P. 281-292.
DOI: 10.3103/S0735272716070013.
8. Aysal T. C. Meridian filtering for robust signal processing / T. C. Aysal, K. E. Barner // IEEE Tr. on Signal Processing. – 2007. – V. 55. – N. 8. – P. 3949–3962.

9. Gonzalez J. G. Optimality of the myriad filter in practical impulsive-noise environments / J. G. Gonzalez, G. R. Arce // IEEE Tr. on Signal Processing. – 2001. – V. 49. – N. 2. – P. 438–441. DOI: 10.1109/78.902126.
10. Vovk S. Family of generalized Demidenko functionals for robust estimation / S. Vovk, V. Borulko // Proc. XVIIth Int. Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory (Tbilisi, Georgia, September 24-27, 2012). – IEEE. – 2012. – P. 151–154.
11. Geman S. Bayesian image analysis. An application to single photon emission tomography / S. Geman, D. McClure // Proc. of the American Statistical Association, Statistical Computing Section. – 1985. – P. 12-18.
Received 02.12.2019.
Accepted 04.12.2019.

Оцінювання параметрів у складному шумовому оточенні

Розглядається проблема оцінювання значень параметрів моделі даних, отримуваних у складному шумовому оточенні. Вирішення цієї проблеми грунтується на методі узагальненої максимальної правдоподібності та на застосуванні супермножини вартісних функцій, яка надає можливість виконувати налаштування методу оцінювання на поточне шумове оточення. Основний розгляд проведено для моделі даних у вигляді константи, який подано постановками відповідних задач оптимізації та обговоренням методів їх розв'язання. Представлено приклади чисельного моделювання з оцінювання значення константи, яка спотворена адитивним шумом та різними аномаліями, а також з оцінювання гауссівського імпульсу в складному шумовому оточенні. Зазначені приклади демонструють ефективність використання запропонованого підходу для оцінювання невідомих параметрів.

Parameter estimation for complicated noise environment

The estimation parameter problem for data obtained in complicated noise environment is considered. The problem solving is based on the method of generalized maximum likelihood with the use of a superset of cost functions, which gives the opportunity to tune the estimation method to the current noise environment. The main consideration is given for data model, having form of a constant, and it is presented by corresponding optimization problems and by discussion of their solving methods. Simulation examples for estimating the constant value distorted by additive noise with various anomalies, as well as for estimating Gaussian pulse in complicated noise environment are presented. These examples demonstrate performance of proposed approach to parameter estimation.

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

Вовк Сергей Михайлович - к.ф.-м.н., доцент, Днепровский национальный унситет, доцент кафедры компьютерных наук и информационных технологий.

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