

Robust Adaptation to Background Noise in Multichannel C-OTDR Monitoring Systems

Andrey V. Timofeev, Viktor M. Denisov

Abstract—A robust sequential nonparametric method is proposed for adaptation to background noise parameters for real-time. The distribution of background noise was modelled like to Huber contamination mixture. The method is designed to operate as an adaptation-unit, which is included inside a detection subsystem of an integrated multichannel monitoring system. The proposed method guarantees the given size of a nonasymptotic confidence set for noise parameters. Properties of the suggested method are rigorously proved. The proposed algorithm has been successfully tested in real conditions of a functioning C-OTDR monitoring system, which was designed to monitor railways.

Keywords—Guaranteed estimation, multichannel monitoring systems, non-asymptotic confidence set, contamination mixture.

I. INTRODUCTION

THE detection of targeted noise-like signals observed on the noise-background is topical in multichannel large scale monitoring systems. An example of such a system is brand-new C-OTDR-systems [1]-[4] for monitoring of super-extended objects (oil & gas pipelines, national borders, railways, etc.). There is a prior uncertainty about statistical characteristics of noise and signals, and there are tens of thousands of C-OTDR-channels, whose data need to be processed in real-time. An important factor is the dynamics of noise statistical characteristics in different channels. For example, in C-OTDR monitoring systems (CMS) this dynamics depends of time of the day, the season, and various external factors (technological works on monitoring objects and in its vicinity, vibrations of highways, sounds of underground rivers etc.). So at different times and different places the conditions of the observation are dramatically different. These circumstances influence sensors systems very strongly. The influence implies an increase in Type I and Type II errors. While noise level may be dramatically different in various time intervals for one and the same channel, the industrial noise power and its spectral characteristics as a rule are stable for extended periods of time (no less than a few minutes or sometimes even hours). Unlike the industrial noises, targeted signals have high power and a short duration (no more than a few minutes). So, targeted signals have a shorter stability period with respect to the stability period of noises. The important feature of multichannel monitoring systems which are designed for monitoring of super-extended objects is an enormous quantity of channels. The quantity of

the channels may be up to 100 000 and more, and channel data have to be processed in real time. In this case, without real-time estimates of the background noise parameters (BN) it is not possible to build a target signal detector which would guarantee prescribed level for Type I and Type II errors. So, there is an objective necessity to design a real-time adaptation procedure (RTAP) which guarantees the quality of the BN-parameter estimation. In this paper it is proposed to find this problem solution from the standpoint of sequential analysis inside of the linear models class. The basic assumption, which has been used in suggested approach, is an existence of a long enough periods, which we call the period of the initial system adjustment or ISA-period. During the ISA-period we can observe only noise in each channel of the monitoring system, and we can use this period for an initial adaptation to noise in every channel. During this phase, the parameters of noise are calculated to be used in the subsystem of targeted signals detection. But the channel noise continues to change its characteristics, hence the results of the initial adaptation gradually becomes irrelevant. Because of that, the process of adaptation to noise has to be continued, and we will call this process a “regular adaptation” (RA). During the RA-process, channel observations are used, which do not contain the targeted signals. The time intervals which contain detected signals are cut out from the channel data stream; and the only remainders are used for adaptation. In CMS, the BN intensity is approximated well by a linear regression with unknown parameters. Those parameters are constant during certain time intervals, which we call as BN-intervals. The BN-intervals durations are enough to build a confidence set with a given size. The set of BN-intervals includes the ISA-period and all time intervals without signals. The main idea is to build the confidence sets for the BN-parameters from the standpoint of sequential analysis. Those sets will be having the given size, and they will be built for finite time. This approach was successfully used for detection of targeted signals in C-OTDR monitoring systems. Systems of that class are new monitoring systems and they are very effective in controlling the seismoacoustic field in the vicinity of the monitored objects. An ordinary fiberoptic cable buried close to the object is used as a sensor. The basis of this method is the use of a vibrosensitive infrared stream injected into a standard monomode fiber by means of a coherent semiconductor laser at the wavelength of 1550 nm. Probing is carried out in the pulsed mode, with the frequency of 8-15 kHz at the pulse length of 20-100 ns. The optical fiber (system sensor) is put into the ground, at the depth of 30-50 cm, at the distance of 5-10 m from the monitoring object. When a pulse is moving

A. V. Timofeev is with the LLP “EqualiZoom”, Astana, Kazakhstan (Phone: +7-911-191-42-67, e-mail: timofeev.andrey@gmail.com).

Viktor M. Denisov is with the Company “Flagman-Geo”, Saint-Petersburg, Russia, +7 911 982 39 09.

along the optical fiber, the Rayleigh elastic backscattering is realized on its natural irregularities, which due to high coherence of the used laser of 3B class leads to formation of the so-called stable interference structures of chaotic type, otherwise called speckles or speckle images. A sequence of speckles is received in the point of emanation using an ordinary welded coupler or a circulator. The central moment of the concept is the phenomenon that any seismic vibration arising on the surface of the optical fiber due to propagation of seismoacoustic waves from the sources of elastic oscillations, changes its local refractive index. Changes of the local refractive index are reflected in the time-and-frequency structure (TFS) of the respective speckle. Knowing the pulse duration and the velocity of wave propagation in the optical fiber, it is easy to determine the section where the TFS speckle deviation took place. Analysis of the sequence of speckle structures using wavelet conversion apparatuses (the phase of singling out of primary signs of target signals) and Lipschitz classifiers (the phase of target signals classification) makes it possible not only to reliably detect the target source of seismoacoustic radiation, but also to determine its type and area of occurrence. In particular, location of the target source of seismoacoustic radiation is determined with the accuracy of up to 5 m at the distance of up to 40 km from the laser location. Actually, as a result of logical processing, several thousands of the so-called C-OTDR channels are formed on the monitoring distance, each of which transfers information on seismoacoustic activity at the well-defined point of the space. It is obvious that the width of the typical C-OTDR channel is 5 m. The following problems are solved in the process of analysis of seismic activity in C-OTDR-systems: a) Target Seismic-Acoustic Events (TSAE's) detection, [5]; b) TSAE location assessment; and c) TSAE type classification, [6]. Proposed RTAP belongs to the detection subsystem (this subsystem must guarantee the upper bounds for the probability of errors of the first and the second kinds). The RTAP is guaranteed the BS-parameters estimate quality in non-asymptotic sense, hence, BS-parameters guaranteed estimates are a basis for building the TSAE-detection procedure with prescribed characteristics (the detection procedure must guarantee TSAE detection with given upper bounds for Type I and Type II errors). The proposed algorithm has been successfully tested in real conditions of a functioning of C-OTDR system, which was designed to monitor the ballast of railway tracks. Since the meaning of the RA-process is similar to meaning of the ISA procedure, in this article we will describe the RTAP for the initial phase only.

II. RESEARCH OBJECTIVE

Let us assume that we have a multichannel monitoring system. There are array of channels, which are used for getting targeted signals. Indexes of system channels in conjunction form a set:

$$\mathbf{Z} = \{1, 2, \dots, z\}$$

Observations are made at successive times, which form a set $N = \{1, 2, \dots\}$. Thus, the observations are form the following sets:

$$\mathbf{X} = \{X(n) | n \in N\}, \quad X(n) = (x_1(n), x_2(n), \dots, x_z(n)) \in R^z, \\ x_j(n) = \langle X(n) \rangle_j$$

are measurement of j-th channel at moment n.

For each channel $j \in \mathbf{Z}$ observations are described by:

$$\forall n < \tau_j : x_j(n) = a_j \theta_j^* + \xi_j(n). \quad (1)$$

Here, $\{\xi_j(n)\}$ are random variables; distributions of $\{\xi_j(n)\}$ are described by:

$$Q_{j_n}(x) = (1 - \varepsilon)F_{j_n}(x) + \varepsilon C_{j_n}(x), \quad j \in \mathbf{Z},$$

Here ε is unknown probability of outliers; $C_{j_n}(x)$ is unknown distribution of outliers;

$$\int x^2 dC_{j_n}(x) = \infty; \quad F_{j_n}(x)$$

is unknown distribution:

$$\mathbf{E}_F \xi_j(n) = 0, \quad \mathbf{E}_F \xi_j^2(n) \leq b_j(n), \quad \forall_{k \neq p} \mathbf{E}_F \xi_j(k) \xi_j(p) = 0$$

- noise parameters $\theta^* = (\theta_1, \theta_2, \dots, \theta_z)$ are priori unknown;
- parameters $\{a_j\}, \{d_n\}$ are given.

The research objective is to build the procedure of adaptation to background noise parameters in real time. This procedure has to guarantee the BS-parameters estimate quality in non-asymptotic sense: the prescribed estimation quality has to be achieved while using a finite number of observations only. The solution will be found in form of sequential plan (τ, δ) , where τ the stopping moment, $\delta = (\delta_1, \delta_2, \dots, \delta_z)$ is given z-vector (vector δ components are prescribed sizes of confidence set Ξ_τ), and besides:

- $\mathbf{P}(\tau < \infty) = 1,$
- $\|\Xi_\tau\| < \delta,$
- $\mathbf{P}(\theta^* \in \Xi_\tau) \geq P_c,$

for prescribed values $\delta_j > 0, P_c \in (0, 1)$.

III. SOLUTION METHOD

A. Case When Probability of Outliers ε Equals to Zero

Let us write (1) in following form:

$$X(n) = A(n)\theta^* + \xi(n), \quad (2)$$

where

- $\theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_z^*) \in R^z$
- $\xi(n) = (\xi_1(n), \xi_2(n), \dots, \xi_z(n)) \in R^z$, $E\xi(n) = 0$,
 $E\xi(n)\xi^T(n) \leq B(n) = \|b_y(n)\|$;
- $\{B(n)\}$ are given sequence of $(z \times z)$ matrixes;
- $\{A(n)\}$ are given sequence of $(z \times z)$ matrixes;
- $\eta(n) = \sum_{i=1}^n A^T(i)\xi(i)$.

Let us denote:

- $\Gamma(n) = \left(\sum_{i=1}^n A^T(i)A(i)\right)^{-1}$,
- $c(n) \in R^z, \forall i: \langle c(n) \rangle > 0$.

Let the matrices $\Gamma(n)$ exist for any n . Consider:

$$\begin{aligned} \theta^-(n+1) &= \Gamma(n+1) \left(\Gamma^{-1}(n)\theta^-(n) + c(n) + \right. \\ &\quad \left. A^T(n+1)X(n+1) - c(n+1) \right), \\ \theta^+(n+1) &= \Gamma(n+1) \left(\Gamma^{-1}(n)\theta^+(n) - c(n) + \right. \\ &\quad \left. A^T(n+1)X(n+1) + c(n+1) \right). \end{aligned}$$

Let us denote:

$$\begin{aligned} \theta_l(n) &: \left(\forall j: \langle \theta_l(n) \rangle_j = \min \left(\langle \theta^-(n) \rangle_j, \langle \theta^+(n) \rangle_j \right) \right), \\ \theta_h(n) &: \left(\forall j: \langle \theta_h(n) \rangle_j = \max \left(\langle \theta^-(n) \rangle_j, \langle \theta^+(n) \rangle_j \right) \right). \end{aligned}$$

Expressions $\theta_l(n)$ and $\theta_h(n)$ are a low bound and an upper bound of a rectangular parallelepiped $\Xi(n) \in R^z$. Let us consider matrixes:

$$\begin{aligned} D(n) &= [d_{ij}(n)] = E\eta(n)\eta^T(n) = \sum_{k=1}^n A^T(k)B(k)A(k), \\ D^+(n) &= \left[d_{ij}^+(n) \left(d_{ii}^+(n)d_{jj}^+(n) \right)^{-0.5} \right], n \geq 1. \end{aligned}$$

We define the sequence of vectors $\{c(n)\}$ as:

$$c(n) = \left(J(u_n^+) / (1 - P_c) \right)^{0.5} S(n),$$

where

$$\begin{aligned} S(n) &= (|d_{11}(n)|, |d_{22}(n)|, \dots, |d_{zz}(n)|), \\ J(x) &= z^{-2} \left(x^{0.5} + ((z^2 - x)(z - 1))^{0.5} \right)^2, \\ u_n^+ &= \varepsilon_z D^+ \varepsilon_z^T, \varepsilon_z = (1, 1, \dots, 1) \in R^z. \end{aligned}$$

Let us denote

$$\forall n \geq 1: \pi(n) = 2\Gamma(n)c(n) = \theta_h(n) - \theta_l(n).$$

The sequential plan for estimation of parameter θ^* we introduce in following form $(\pi(n), \tau)$, where τ is moment of observations stop:

$$\tau = \inf \left\{ n \geq 1 \mid \forall j: \langle \pi(n) \rangle_j \leq \delta_j \right\}.$$

The following theorem describes properties of sequential plan $(\pi(n), \tau)$.

Theorem 1. Let:

$$\lim_{n \rightarrow \infty} \Gamma(n)S^T(n) \rightarrow 0.$$

Then

1. $P_{\theta^*}(\tau < \infty) = 1$.
2. $P_{\theta^*}(\theta^* \in \Xi(\tau)) = 1, \Xi(\tau) = \{\theta \mid \theta_l \leq \theta \leq \theta_h\}$.

Proof of Theorem 1 is based on usage Theorem 4.1, p. 510 [7]. So, the confidence rectangular parallelepiped $\Xi(n)$ with given sizes will be built by the time moment τ .

The estimation of value $E\tau$ (the mean duration of observation in sequential plan $(\pi(n), \tau)$) is an important task for practice. We propose a new approach to obtain the upper bound of $E\tau$. Let us rewrite (1) in following form:

$$X(n) = A(n)\theta^* + G(n)\varepsilon(n),$$

$$\forall n \geq 1: \xi(n) = G(n)\varepsilon(n), B(n) = G^T(n)G(n),$$

$$E\varepsilon(n) = 0, E\varepsilon(n)\varepsilon^T(n) = E.$$

$G(n) - (m \times m)$ matrix, E - the identity $(m \times m)$ -matrix. Further, we rewrite $\pi(n)$ in following form:

$$\forall n \geq 1: \pi(n) = 2\mathbf{y}(n)(1 - P_c)^{-1}, \mathbf{y}(n) = \Gamma(n) \left(J(u_n^+) \right)^{0.5} S(n).$$

For some positive constants k and β we have:

$$\forall n \geq 1: P_{\theta^*} \left(\|\xi(n)\|^2 \geq k \|\varepsilon(n)\|^2 \right) = 1, \quad (3)$$

$$\forall n \geq 1, j: \langle \mathbf{y}(n) \rangle_j \leq \left(\beta \left(\sum_{i=1}^n \|X(i)\|^2 \right)^{0.5} \right)^{-1} \quad (4)$$

For some $H > 0$ let us denote:

$$\eta(H) = \inf \left\{ n \geq 1 \mid \sum_{i=1}^n \|\varepsilon(i)\|^2 \geq H \right\}$$

$\eta(H)$ is the random variable, which have the finite $E\eta(H)$.

Theorem 2. Let distribution density $g(z)$ of random variables $\{\varepsilon(n)\}$ is a centrally symmetric density. Thus, $g(z) = g(-z)$, the set $K(u) = \{x | g(x) \geq u\}$ is a convex set for any u , $0 < u < \infty$. In this case, when (3) and (4) are true for process $X(n)$ we have:

$$\mathbf{E} \tau \leq \mathbf{E} \eta(H) + 1,$$

where

$$H = 4 \left(\beta^2 \inf_{1 \leq i \leq z} \langle \delta \rangle_i^2 (1 - P_c) k \right)^{-1}$$

The proof is based on usage of the process $X(n)$ special decomposition. The question arise: how to estimate the $\mathbf{E} \eta(H)$? If $\{\varepsilon(n)\}$ are gauss variables we have:

$$\mathbf{E} \eta(H) = \begin{cases} H + 2, & \text{if } z=1; \\ H / 2^v + 1, & \text{if } z \geq 2. \end{cases}$$

$v = [\log_2 z]$, $[a]$ is the integer part of a . If $\{\varepsilon(n)\}$ are mutually independent, and they have uniform distribution, we have:

$$\mathbf{E} \eta(H) = \begin{cases} H + 3, & \text{if } z=1; \\ H / 2^v + 2, & \text{if } z \geq 2. \end{cases}$$

B. Case When Probability of Outliers ε Not Equal to Zero

Let us consider the following recurrent estimations:

$$\tilde{\theta}(n+1) = \Gamma(n+1) \left(\Gamma^{-1}(n) \tilde{\theta}(n) + \mathbf{A}^T(n+1) X(n+1) \right),$$

for some $k > 0$, $n > k$,

$$\theta_r^-(n+1) = \begin{cases} \theta^-(n), & \text{if } \tilde{\theta}(n+1) \notin \Xi(n) \\ \theta^-(n+1), & \text{if } \tilde{\theta}(n+1) \in \Xi(n), \end{cases}$$

$$\theta_r^+(n+1) = \begin{cases} \theta^+(n), & \text{if } \tilde{\theta}(n+1) \notin \Xi(n) \\ \theta^+(n+1), & \text{if } \tilde{\theta}(n+1) \in \Xi(n). \end{cases}$$

Further,

$$\theta_l^{(r)}(n) : \left(\forall j : \langle \theta_l^{(r)}(n) \rangle_j = \min \left(\langle \theta_r^-(n) \rangle_j, \langle \theta_r^+(n) \rangle_j \right) \right),$$

$$\theta_h^{(r)}(n) : \left(\forall j : \langle \theta_h^{(r)}(n) \rangle_j = \max \left(\langle \theta_r^-(n) \rangle_j, \langle \theta_r^+(n) \rangle_j \right) \right).$$

So, $\theta_l^{(r)}(n)$ and $\theta_h^{(r)}(n)$ are boundaries of the robust confidence rectangular parallelepiped $\Xi_r(n) \in R^z$. Let $\forall n \geq 1 : \pi_r(n) = \theta_h^{(r)}(n) - \theta_l^{(r)}(n)$. Consider the moment of observations stop as:

$$\tau_r = \inf \left\{ n \geq 1 \mid \forall j : \langle \pi_r(n) \rangle_j \leq \delta_j \right\}.$$

It is easy to see that

- $\mathbf{P}(\tau_r < \infty) = 1$,
- $\forall \theta_1, \theta_2 \in \Xi_r(\tau_r) : \forall j : |\langle \theta_1 - \theta_2 \rangle_j| \leq \delta_j$,
- $\mathbf{P}(\theta^* \in \Xi_r(\tau_r)) \geq P_c$,
- $\mathbf{E} \tau_r \leq (1 - \varepsilon)^{-1} \mathbf{E} \tau$,

for prescribed values $\delta_j > 0, P_c \in (0, 1)$, if $0 < \varepsilon < 1$ and $\forall n < k : Q_{jn}(x) = F_{jn}(x), j \in \mathbf{Z}$.

IV. RESULTS OF PRACTICAL USE

Table I contains results of usage the suggested approach to estimate the BN-parameters. In the experiment, the length of sensor of C-OTDR monitoring system was equal to 1200m. There were 1200 C-OTDR channels. The matrix $\mathbf{B}(n) = \mathbf{B}$ was prior estimated on base of serial BN-observations. The BN-parameters had been different for each channel. It was due to difference in seismoacoustic medium types along the sensor, and difference of the noise-sources distribution along the sensor. There were three basic types of BN: “underwater collector” (UP), “technological devices” (TD) (three types), “channel-noises”(CN). Combinations of those types could be in each channel. For practical reasons, the δ_i were choose as

$0.2(b_{ji})^{0.5}$. There are three types of TD:

- TD1: “perforator”;
- TD2: “diesel generator”;
- TD3: “traction substation”.

For each combination of BN-types the $\overline{\mathbf{E} \tau_r}$ (average of τ_r in serial tests) was estimated. The minimal value of $\overline{\mathbf{E} \tau_r}$ corresponds to CN-case. This was due to by low level of CN variance. On another hand, the maximal value of $\overline{\mathbf{E} \tau_r}$ corresponds to TD1&CN combination. In this case we had the maximal level of the observation process variance, and it implies the maximal value of $\overline{\mathbf{E} \tau_r}$ (for compensate of high variance we need use bigger sample size in compare of low variance case). The findings demonstrate acceptable characteristics of suggested adaptation method for practical usage.

TABLE I
 THE PRACTICAL DETECTION RESULTS

No	Types of BN	$\overline{\mathbf{E} \tau_r}$ (sec)
1	CN	7
2	UP&CN	12
3	TD1&CN	19
4	TD2&CN	21
5	TD3&CN	18

V. CONCLUSIONS

Proposed sequential method for robust adaptation to background noise parameters for real-time is nonparametric (not require to know probabilistic distribution of noises) and this method is non-asymptotic (the required estimation quality is achieved for finite time). The method is designed to operate as an adaptation-unit, which is included inside a detection subsystem of an integrated multichannel monitoring system. Proposed method guarantees the given size of a nonasymptotic confidence set for noises parameters. Properties of the suggested method are rigorously proved. The proposed algorithm has been successfully tested in real conditions of the functioning of C-OTDR monitoring system, which was designed to monitor the railways.

ACKNOWLEDGMENT

This study has been produced under the project "Development of a remote monitoring system to protect backbone communications infrastructure, oil and gas pipelines and other extended objects (project code name – OXY)", financed under the project "Technology Commercialization", supported by the World Bank and the Government of the Republic of Kazakhstan.

REFERENCES

- [1] K. N. Choi, J. C. Juarez, H. F. Taylor, "Distributed fiber optic pressure/seismic sensor for low-cost monitoring of long perimeters", Proc. SPIE 5090, Unattended Ground Sensor Technologies and Applications, 2003, pp. 134-141.
- [2] J. C. Juarez, E. W. Maier, K. N. Choi, and H. F. Taylor, "Distributed Fiber-Optic Intrusion Sensor System", *Journal of Lightwave Technology*, Vol. 23, Issue 6, 2005, pp. 2081-2087.
- [3] S. S. Mahmoud, Y. Visagathilagar, J. Katsifolis., "Real-time distributed fiber optic sensor for security systems: Performance, event classification and nuisance mitigation". *Photonic Sensors*, Vol.2, Issue 3, 2012, pp. 225-236.
- [4] V. Korotaev, V. M. Denisov, A. V. Timofeev, and M. G. Serikova, "Analysis of seismoacoustic activity based on using optical fiber classifier," in Latin America Optics and Photonics Conference, OSA Technical Digest (online) (Optical Society of America, 2014), paper LM4A.22.
- [5] Timofeev A.V. The guaranteed detection of the seismoacoustic emission source in the C- OTDR systems, *International Journal of Mathematical, Computational, Physical and Quantum Engineering*, Vol.8, Issue 10, 2014, pp. 1213-1216.
- [6] Timofeev A.V., Egorov D.V., Multichannel classification of target signals by means of an SVM ensemble in C-OTDR systems for remote monitoring of extended objects, MVML-2014 Conference Proceedings Prague, 2014, V.1.
- [7] S. Karlin, V. Studden, "Tchebycheff systems: with applications in analysis and statistics", Interscience, 1966 .

Timofeev Andrey V. was born in Chita (Russia). He received Dr. habil. Sc. ing. in Computer and Information Sciences from Tomsk State University of Control Systems and Radioelectronics, Russia, in 1994. A number of research publications in the International journals (JKSS (Elsevier), Statistical Methodology (Elsevier), Automation and Remote Control etc) and International/National conferences are at his credit. He is on the editorial board of several journals and conferences and a referee of several others. His research interests include non-asymptotic nonlinear methods of confidence estimation of multidimensional parameters of stochastic systems; machine learning, large margin classification in Banach spaces; confidence Lipschitz classifiers; technical diagnostics, C-OTDR systems; data mining; change-

point problem; alpha-stable laws; statistical classification in application to biometrics and seismics.

Viktor M. Denisov was born in Pskov (Russia). He received Dr. habil. sc. ing. in Computer and Information Sciences from ITMO University in 1994. A number of research publications in the International journals and International/National conferences are at his credit. He is main research interests include information and computer technology, instrumentation, measuring devices and systems, geotechnical monitoring, mobile and cloud computing, sensors and sensors, intelligent sensor networks, mobile medicine. Viktor M. Denisov is the CEO of FlagmanGeo, Ltd. He is an expert of the Ministry of Education and Science of the Russian Federation in the field of information and computer technologies. Also, he is a member of the editorial board «International Journal of E-Health and Medical Communications» (IJEHMC). He is also the chief designer of the family of the newest field of geophysical sensor devices. Under his leadership made a lot of major projects for the monitoring of complex engineering structures and dangerous natural objects. He developed a new method for geotechnical monitoring based on the use of arrays of micromechanical sensors based on flexible chassis