Volume 84, Number 7
July 2023

ISSN 0005-1179
CODEN: AURCAT

AUTOMATION AND REMOTE CONTROL

http://ait.mtas.ru

Automation and Remote Control

Vol. 84, No. 7, July 2023

Automation and Remote Control

ISSN 0005-1179

SCOPE

Automation and Remote Control is one of the first journals on control theory. The scope of the journal is control theory problems and applications. The journal publishes reviews, original articles, and short communications (deterministic, stochastic, adaptive, and robust formulations) and its applications (computer control, components and instruments, process control, social and economy control, etc.).

Automation and Remote Control is abstracted and/or indexed in *ACM Digital Library, BFI List, CLOCKSS, CNKI, CNPIEC Current Contents/ Engineering, Computing and Technology, DBLP, Dimensions, EBSCO Academic Search, EBSCO Advanced Placement Source, EBSCO Applied Science & Technology Source, EBSCO Computer Science Index, EBSCO Computers & Applied Sciences Complete, EBSCO Discovery Service, EBSCO Engineering Source, EBSCO STM Source, EI Compendex, Google Scholar, INSPEC, Japanese Science and Technology Agency (JST), Journal Citation Reports/Science Edition, Mathematical Reviews, Naver, OCLC WorldCat Discovery Service, Portico, ProQuest Advanced Technologies & Aerospace Database, ProQuest-ExLibris Primo, ProQuest-ExLibris Summon, SCImago, SCOPUS, Science Citation Index, Science Citation Index Expanded (Sci-Search), TD Net Discovery Service, UGC-CARE List (India), WTI Frankfurt eG, zbMATH.*

Journal website: http://ait.mtas.ru

0005-1179/23. *Automation and Remote Control* (ISSN: 0005-1179 print version, ISSN: 1608-3032 electronic version) is published monthly by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya street, Moscow 117997, Russia.

Volume 84 (12 issues) is published in 2023.

Publisher: Trapeznikov Institute of Control Sciences, Russian Academy of Sciences.

65 Profsoyuznaya street, Moscow 117997, Russia; e-mail: redacsia@ipu.rssi.ru; http://ait.mtas.ru, http://ait-arc.ru

Contents

Automation and Remote Control
Vol. 84, No. 7, 2023

Stochastic Systems

Control in Technical Systems

Intellectual Control Systems, Data Analysis

Optimization, System Analysis, and Operations Research

Obituary

Selected Articles from *Avtomatizatsiya v Promyshlennosti*

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 763–777.

STOCHASTIC SYSTEMS

Analytical Investigation of a Single-Server Queueing System with an Incoming MAP Event Flow

A. M. Gortsev[*,a] **and L. A. Nezhelskaya**[*,b]

[*] *National Research Tomsk State University, Tomsk, Russia*
e-mail: [a] *dekanat@fpmk.tsu.ru,* [b] *ludne@mail.ru*

Received February 27, 2023
Revised April 25, 2023
Accepted May 11, 2023

Abstract—This paper considers a single-server queueing system with an incoming Markovian Arrival Process (MAP) request flow with two states. Explicit expressions are derived for the stationary probability distribution of the states and several numerical characteristics of the system (the probability of idle time of the server, the expected number of requests in the system, and the mean queue length). The resulting numerical characteristics are presented in tables and plotted in graphical form as well. The recurrent MAP flow with two states as a special case of correlated MAP request flows is studied.

Keywords: MAP request flow, single-server queueing system, stationary probability distribution of system states, numerical characteristics

DOI: 10.25728/arcRAS.2023.62.11.001

1. INTRODUCTION

Mathematical models of queueing systems and networks (QSs, QNs) adequately describe the behavior of real physical, technical, economic, and other objects and systems. Therefore, they have become widespread in the scientific community. A basic element of QSs and QNs are random incoming request flows. Almost throughout the 20th century, research on QSs and QNs was based on the assumption of the uncorrelated nature of incoming request flows. In other words, the simplest request flows—stationary Poisson ones—were considered. However, at the end of the century, the stationary Poisson flow model lost its adequacy to real information request flows in telecommunication networks and systems, wireless and mobile communication networks due to their intensive development.

The rapid change of digital technologies ensured the penetration of digital networks into all spheres of human activity. It would be impossible without the use and development of mathematical modeling methods and algorithms for network technologies. Since the end of the 20th century, intensive research in modern queueing theory has been dealing with queueing systems with correlated flows (systems with doubly stochastic flows). The emergence of doubly stochastic event flows, a new mathematical model with the most adequate description of the correlated nature of real information flows, was motivated by the practical studies of modern telecommunication networks with essentially nonstationary and correlated heterogeneous information flows.

Doubly stochastic flows are characterized by two stochastics: requests in the flow arrive at random time instants (the first stochastics), and the flow intensity (the accompanying process) is a random process (the second stochastics). There are two types of doubly stochastic flows depending on their accompanying process (intensity): the ones with a continuous random process [1, 2] and

the ones with a piecewise constant random process with a finite (arbitrary) number of states. The studies of second-type flows were first presented almost simultaneously in 1979 in the papers [3–5]. In [3, 4], these flows were called Markov Chain (MC) flows; in [5], Markov Versatile Process (MVP) flows. In [6, 7], the above flows were termed Markovian Arrival Process (MAP) flows. Their main property is correlation. Note that MAP (MC) flows are the most appropriate mathematical model of correlated request flows in real telecommunication systems and networks [8].

The monograph [8], unique in the world literature, systematically presented QSs and QNs with correlated flows. As was emphasized in [8], the analytical investigation of QSs and QNs with correlated flows is a rather difficult process; finding the explicit-form characteristics of QSs and QNs is a nontrivial problem, sometimes unsolvable.

In this paper, we analytically investigate a single-server QS with waiting, the classical incoming MAP request flow with two states [6, 7], and exponential service.

For the stationary operation mode of this QS, explicit analytical formulas are derived for the probability of idle time of the server, the mean queue length, and the expected number of requests in the system.

Note that QSs and QNs with incoming MAP request flows have been analyzed since the 1990s. In particular, the states and parameters of an MAP request flow under perfect and incomplete observability conditions (in the presence of dead time) were estimated by the authors. In this regard, we refer to some publications [9–14].

In addition, the system under consideration differs from the systems operating in a synchronous random environment: in such an environment, synchronous flows are considered in which the state of the control process (accompanying process) changes at random time instants (the instants of events occurrence). Thus, a synchronous random environment always assumes a nonzero probability for changing the states of the control process at the instant of events occurrence in the synchronous flow. In an MAP flow, in contrast, a flow event not necessarily occurs at the instant of changing the state of the control process. (If the probability of event occurrence is always 1, we have a synchronous flow.) Thus, the mathematical model of a random environment considered below generalizes the mathematical model of a synchronous random environment, which is the novelty of this study.

The evolution from the simplest flow to modern mathematical models of information flows in telecommunication systems and networks (to the models of correlated flows, particularly MAP flows) can be traced in the monograph [8]. In addition, it provides an extensive bibliography on QSs and QNs. Among the recent works on this subject, let us mention the paper [15]. Note that numerical analysis is a common feature of research on QSs and QNs with an incoming MAP request flow. This paper continues the investigations initiated in [16].

2. MATHEMATICAL MODEL OF THE SYSTEM. PROBLEM STATEMENT

Consider a single-server QS with waiting. The server receives an incoming MAP flow of events (requests, messages, etc.) whose accompanying process $\lambda(t)$ is a piecewise constant random process with two states S_1 and S_2. If $\lambda(t) = \lambda_i$, then the process $\lambda(t)$ (flow) has the ith state (S_i), $i = 1, 2$; $\lambda_1 > \lambda_2 > 0$. The sojourn time of the process $\lambda(t)$ in the state S_i is a random variable with the exponential distribution function $F_i(t) = 1 - \exp\{-\lambda_i t\}$, $t \geqslant 0$, $i = 1, 2$.

When the ith state of the flow (process $\lambda(t)$) ends, the following instantaneous changes in the system state are possible:

1) A flow event occurs, and the process $\lambda(t)$ passes from the state S_i to the state S_j; the joint probability of this situation is $P_1(\lambda_j|\lambda_i)$, $i, j = 1, 2$.

2) No flow event occurs, and the process $\lambda(t)$ passes from the state S_i to the state S_j; the joint probability of this situation is $P_0(\lambda_j|\lambda_i)$, $i,j=1,2$ $(i\neq j)$.

Note that $P_0(\lambda_j|\lambda_i)+P_1(\lambda_j|\lambda_i)+P_1(\lambda_i|\lambda_i)=1$, $i,j=1,2$ $(i\neq j)$. Here, the occurrence (non-occurrence) of an event in the state S_i is primary, i.e., it precedes the transition from the flow state S_i to the flow state S_j with the probability $P_1(\lambda_j|\lambda_i)$ (the transition from the flow state S_i to the flow state S_j with the probability $P_0(\lambda_j|\lambda_i)$, respectively).

Let the QS operate in a stationary mode. Under the assumptions made, $\lambda(t)$ is the accompanying stationary, piecewise constant, and transitive Markov process with the two states S_1 and S_2. If the process $\lambda(t)$ is in the state S_i, then the request is served in a time $\tau\geqslant 0$ with the exponential distribution law $F^{(i)}(\tau)=1-\exp\{-\mu_i\tau\}$ with the intensity μ_i $(\mu_i>0)$, $i=1,2$.

Remark 1. For an MAP flow, the accompanying random process $\lambda(t)$ does not coincide with the flow intensity: in the states S_1 and S_2, the flow intensity takes the values $\lambda_1[1-P_0(\lambda_2|\lambda_1)]$ and $\lambda_2[1-P_0(\lambda_1|\lambda_2)]$, respectively. Then the mean intensity of this flow is [17]

$$\begin{gathered}\lambda=\lambda_1[1-P_0(\lambda_2|\lambda_1)]\pi_1+\lambda_2[1-P_0(\lambda_1|\lambda_2)]\pi_2,\\ \pi_1=\frac{\lambda_2[1-P_1(\lambda_2|\lambda_2)]}{\lambda_1[1-P_1(\lambda_1|\lambda_1)]+\lambda_2[1-P_1(\lambda_2|\lambda_2)]},\\ \pi_2=\frac{\lambda_1[1-P_1(\lambda_1|\lambda_1)]}{\lambda_1[1-P_1(\lambda_1|\lambda_1)]+\lambda_2[1-P_1(\lambda_2|\lambda_2)]},\end{gathered}\tag{1}$$

where π_1 and π_2 denote the prior probabilities of the states S_1 and S_2 of the process $\lambda(t)$ (flow), respectively, in the stationary mode.

Let $\tau_k=t_{k+1}-t_k$, $k=1,2,\ldots,$ be the duration of the kth interval between the arrival time instants t_k and t_{k+1} of flow requests $(\tau_k\geqslant 0)$. Due to the stationary mode, the probability density of the durations is $p(\tau_k)=p(\tau)$, $\tau\geqslant 0$, for any $k\geqslant 1$. Then, without loss of generality, t_k can be supposed 0, i.e., a request arrives at the time instant $\tau=0$. The following explicit formula for the probability density $p(\tau)$ was derived in [11]:

$$\begin{gathered}p(\tau)=\gamma z_1e^{-z_1\tau}+(1-\gamma)z_2e^{-z_2\tau},\quad \tau\geqslant 0,\\ \gamma=\{z_2-\lambda_1\pi_1(0)[1-P_0(\lambda_2|\lambda_1)]-\lambda_2\pi_2(0)[1-P_0(\lambda_1|\lambda_2)]\}(z_2-z_1)^{-1},\\ z_{1,2}=\left[(\lambda_1+\lambda_2)\mp\sqrt{(\lambda_1-\lambda_2)^2+4\lambda_1\lambda_2P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)}\right]/2,\\ \pi_1(0)=\frac{P_1(\lambda_1|\lambda_2)+P_1(\lambda_1|\lambda_1)P_0(\lambda_1|\lambda_2)}{P_1(\lambda_1|\lambda_2)+P_1(\lambda_2|\lambda_1)+P_1(\lambda_1|\lambda_1)P_0(\lambda_1|\lambda_2)+P_1(\lambda_2|\lambda_2)P_0(\lambda_2|\lambda_1)},\\ \pi_2(0)=1-\pi_1(0).\end{gathered}\tag{2}$$

In (2), $\pi_i(0)$ is the stationary probability that the process $\lambda(\tau)$ has the state S_i, $i=1,2$, at the time instant $\tau=0$ (the arrival of an MAP flow request); z_1 and z_2 are the roots of the characteristic equation $z^2-(\lambda_1+\lambda_2)z+\lambda_1\lambda_2[1-P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)]=0$, where $0<z_1<z_2$ due to (2); γ is a value that depends on the flow parameters.

Consider two adjacent intervals (t_k,t_{k+1}) and (t_{k+1},t_{k+2}) with the durations $\tau_k=t_{k+1}-t_k$ and $\tau_{k+1}=t_{k+2}-t_{k+1}$, respectively. Since the flow is stationary, they are located arbitrarily on the time axis. Letting $k=1$, we study two intervals (t_1,t_2) and (t_2,t_3) with the durations $\tau_1=t_2-t_1$ and $\tau_2=t_3-t_2$, respectively, where $\tau_1\geqslant 0$ and $\tau_2\geqslant 0$. In this case, $\tau_1=0$ corresponds to the

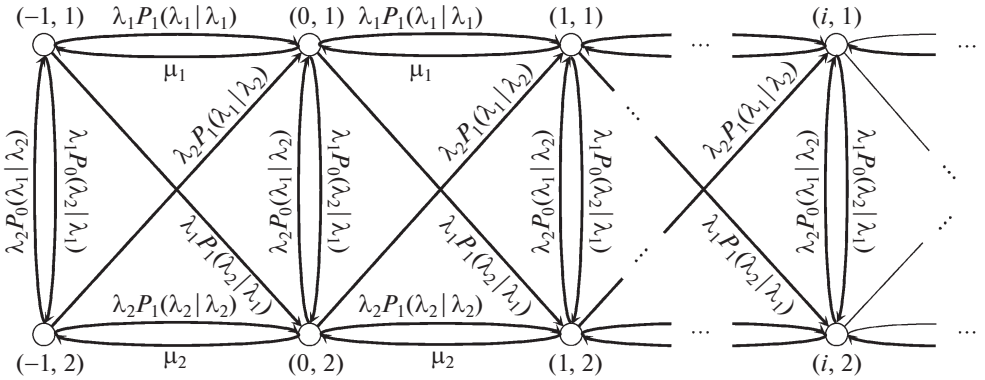


Fig. 1. The stochastic state transition graph for the process $\lambda(t)$.

arrival time instant t_1 of a flow request and $\tau_2 = 0$ to the arrival time instant t_2 of the next flow request. The joint probability density has the form [11, 13]

$$p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2) + \gamma(1-\gamma)\frac{P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1)}{1 - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)} \times \left(z_1 e^{-z_1\tau_1} - z_2 e^{-z_2\tau_1}\right)\left(z_1 e^{-z_1\tau_2} - z_2 e^{-z_2\tau_2}\right), \quad \tau_1 \geqslant 0, \quad \tau_2 \geqslant 0, \tag{3}$$

where z_1, z_2, and $p(\tau_k)$ are given by (2) for $\tau = \tau_k$, $k = 1, 2$.

According to (3), an MAP flow is generally a correlated flow; it turns recurrent or degenerates into elementary only in special cases.

Special case 1: $P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1) = 0$, a recurrent MAP request flow with two states. In this case, $p(\tau)$ is given by (2), where $\gamma = [z_2 - \lambda_1 P_1(\lambda_1|\lambda_1) - \lambda_2 P_1(\lambda_2|\lambda_2)](z_2 - z_1)^{-1}$.

From (3) it follows that $p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2)$. Since the arrival time instants $t_1, \ldots, t_k$ in the flow induce a nested Markov chain $\{\lambda(t_k)\}$, for an arbitrary number k, $k \geqslant 2$, we have $p(\tau_1, \ldots, \tau_k) = p(\tau_1) \ldots p(\tau_k)$.

The product $\gamma(1-\gamma)$ in (3) can be represented as

$$\gamma(1-\gamma) = \frac{z_1 z_2}{(z_2 - z_1)^2}\{\lambda_1[1 - P_0(\lambda_2|\lambda_1)] - \lambda_2[1 - P_0(\lambda_1|\lambda_2)]\} \times \{\pi_1(0)\lambda_1[1 - P_1(\lambda_1|\lambda_1)] - \pi_2(0)\lambda_2[1 - P_1(\lambda_2|\lambda_2)]\} \times \{\lambda_1\lambda_2[1 - P_0(\lambda_2|\lambda_1)][1 - P_1(\lambda_2|\lambda_2)] + \lambda_1\lambda_2[1 - P_0(\lambda_1|\lambda_2)][1 - P_1(\lambda_1|\lambda_1)]\}^{-1}. \tag{4}$$

The expression (4) implies special cases 2 and 3; see below.

Special case 2: $\lambda_1[1 - P_0(\lambda_2|\lambda_1)] - \lambda_2[1 - P_0(\lambda_1|\lambda_2)] = 0$, an elementary flow with a parameter z_1. From (2) it follows that $z_1 = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]$, $\gamma = 1$; $p(\tau) = z_1 e^{-z_1\tau}$, $\tau \geqslant 0$.

Special case 3: $\pi_1(0)\lambda_1[1 - P_1(\lambda_1|\lambda_1)] - \pi_2(0)\lambda_2[1 - P_1(\lambda_2|\lambda_2)] = 0$, an elementary flow with a parameter z_1. From (2) it follows that $z_1 = \lambda_2[P_1(\lambda_2|\lambda_1) + P_1(\lambda_2|\lambda_2)P_0(\lambda_2|\lambda_1)]$, $\gamma = 1$; $p(\tau) = z_1 e^{-z_1\tau}$, $\tau \geqslant 0$.

The problem is to find an explicit analytical form of the numerical characteristics of this QS:

(a) the probability of idle time of the server,

(b) the mean queue length,

(c) the expected number of requests in the system.

Let $i(t)$ be the number of requests in the queue at an arbitrary time instant t ($i(t) = 0, 1, \ldots$). Since the incoming MAP flow is correlated, the random process $i(t)$ is not Markovian. To construct a Markov process, it is necessary to consider the state of the incoming MAP flow. For this

purpose, we introduce an additional variable $j(t)$, i.e., the state of the incoming MAP flow (the state of the accompanying process $\lambda(t)$ at an arbitrary time instant t), $j(t) = 1, 2$. If $j(t) = 1$, then $\lambda(t) = \lambda_1$; if $j(t) = 2$, then $\lambda(t) = \lambda_2$, which ensures the Markov property of the two-dimensional process $(i(t), j(t))$.

Remark 2. Because the intensity of the server in the state S_j is μ_j ($\mu_j > 0$), $j = 1, 2$, the component $j(t)$ of the two-dimensional Markov process $(i(t), j(t))$ must be observable in the same way as the component $i(t)$ is. Then the accompanying process $\lambda(t)$, generally unobservable, must be treated as an observable process that controls the change of states in the MAP request flow.

Since the stationary operation mode is considered, the system state will be denoted by (i, j), $i = 0, 1, \ldots$, $j = 1, 2$. There are two more possible states, $(-1, 1)$ and $(-1, 2)$; in these states, the system receives no requests (the queue length is zero and the server is idle).

Under the prerequisites above, the mathematical model of the QS under study can be represented as a connected stochastic graph [18]; see Fig. 1. Here, the vertices reflect the states of the QS; each arc corresponds to infinitesimal characteristics (state transition intensities), without loops in each state; each vertex (each state) is reachable and recurrent.

3. DERIVATION OF NUMERICAL CHARACTERISTICS OF THE SYSTEM

We denote by $P(i, 1)$ and $P(i, 2)$ the stationary (final) probabilities of the system states ($i = -1, 0, \ldots$). The stochastic graph cutsets $G_{i1} = \{(i-1, 1; i, 1), (i, 1; i-1, 1), (i, 1; i+1, 1), (i+1, 1; i, 1), (i, 1; i, 2), (i, 2; i, 1), (i-1, 2; i, 1), (i, 1; i+1, 2)\}$, $G_{i2} = \{(i-1, 2; i, 2), (i, 2; i-1, 2), (i, 2; i+1, 2), (i+1, 2; i, 2), (i, 2; i, 1), (i, 1; i, 2), (i-1, 1; i, 2), (i, 2; i+1, 1)\}$, $i = 0, 1, \ldots$, satisfy the following infinite system of difference equations with constant coefficients:

$$\begin{gathered}
\mu_1 P(i+1, 1) - (\lambda_1 + \mu_1) P(i, 1) + \lambda_1 P_1(\lambda_1|\lambda_1) P(i-1, 1) \\
+ \lambda_2 P_0(\lambda_1|\lambda_2) P(i, 2) + \lambda_2 P_1(\lambda_1|\lambda_2) P(i-1, 2) = 0, \\
\mu_2 P(i+1, 2) - (\lambda_2 + \mu_2) P(i, 2) + \lambda_2 P_1(\lambda_2|\lambda_2) P(i-1, 2) \\
+ \lambda_1 P_0(\lambda_2|\lambda_1) P(i, 1) + \lambda_1 P_1(\lambda_2|\lambda_1) P(i-1, 1) = 0, \quad i = 0, 1, \ldots.
\end{gathered} \tag{5}$$

The solution of system (5) is found in the form $P(i, 1) = \xi^i$, $P(i, 2) = C\xi^i$ ($i = 0, 1, \ldots$). The characteristic equation for (5) is

$$\begin{gathered}
(\xi - 1)\Big\{\mu_1\mu_2\xi^3 - [\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)]\xi^2 \\
+ [\lambda_1\lambda_2 + \lambda_1\mu_2 P_1(\lambda_1|\lambda_1) + \lambda_2\mu_1 P_1(\lambda_2|\lambda_2) - \lambda_1\lambda_2 P_0(\lambda_1|\lambda_2) P_0(\lambda_2|\lambda_1)]\xi \\
- \lambda_1\lambda_2[P(\lambda_1|\lambda_1) P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2) P_1(\lambda_2|\lambda_1)]\Big\} = 0.
\end{gathered} \tag{6}$$

Consider conditions for the existence of the stationary operation mode of the QS (the existence of the probabilities $P(i, 1)$ and $P(i, 2)$, $i = -1, 0, \ldots$). The random variable $\boldsymbol{\tau}$, the duration of the time interval between sequential events in the MAP request flow, has the expectation

$$E(\boldsymbol{\tau}) = \int_0^\infty \tau p(\tau) d\tau, \tag{7}$$

where the density $p(\tau)$ is given by (2). Substituting this function into (7) yields $E(\boldsymbol{\tau}) = [\gamma z_2 + (1-\gamma) z_1]/z_1 z_2$. Then the expected number of requests in the incoming correlated MAP flow per unit time can be written as $\lambda = 1/E(\boldsymbol{\tau}) = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]\pi_1 + \lambda_2[1 - P_0(\lambda_1|\lambda_2)]\pi_2$, which coincides with (1). On the other hand, the expected number of requests served per unit time is $\mu = \mu_1\pi_1 + \mu_2\pi_2$.

Consider a situation where $\lambda = \mu$, or $(\mu_1 - \lambda_1[1 - P_0(\lambda_2|\lambda_1)])\pi_1 + (\mu_2 - \lambda_2[1 - P_0(\lambda_1|\lambda_2)])\pi_2 = 0$. Hence, this expression vanishes only if $\mu_1 = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]$, $\mu_2 = \lambda_2[1 - P_0(\lambda_1|\lambda_2)]$. Substituting these formulas for μ_1 and μ_2 into (6), we obtain the characteristic equation

$$\begin{gathered}\lambda_1\lambda_2(\xi-1)^2\Big\{[1-P_0(\lambda_1|\lambda_2)][1-P_0(\lambda_2|\lambda_1)]\xi^2 - [2-P_0(\lambda_1|\lambda_2)-P_0(\lambda_2|\lambda_1)]\xi \\ + [P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1)]\Big\} = 0.\end{gathered} \tag{8}$$

Since Eq. (8) has multiple roots, the general solution of system (5) with $\mu_1 = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]$ and $\mu_2 = \lambda_2[1 - P_0(\lambda_1|\lambda_2)]$ takes the form

$$\begin{gathered}P(i,1) = D_1\xi_1^i + D_2 i\xi_2^i + D_3\xi_3^i + D_4\xi_4^i, \\ P(i,2) = B_1D_1\xi_1^i + B_2D_2 i\xi_2^i + B_3D_3\xi_3^i + B_4D_4\xi_4^i, \quad i = 0,1,\ldots.\end{gathered} \tag{9}$$

In (9), $P_s(i,1) = D_s\xi_s^i$ and $P_s(i,2) = B_sD_s\xi_s^i$, $s = \overline{1,4}$, are partial solutions of system (5); their constants B_s and D_s are determined from the boundary conditions, $\xi_1 = \xi_2 = 1$, and

$$\begin{gathered}\xi_{3,4} = \Big\{[2 - P_0(\lambda_1|\lambda_2) - P_0(\lambda_2|\lambda_1)] \\ \mp \Big([2 - P_0(\lambda_1|\lambda_2) - P_0(\lambda_2|\lambda_1)]^2 - 4[1 - P_0(\lambda_1|\lambda_2)][1 - P_0(\lambda_2|\lambda_1)]b\Big)^{\frac{1}{2}}\Big\} \\ \times\Big\{2[1 - P_0(\lambda_1|\lambda_2)][1 - P_0(\lambda_2|\lambda_1)]\Big\}^{-1}, \\ b = P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1).\end{gathered} \tag{10}$$

Here, three cases are possible: $b > 0$, $b < 0$, and $b = 0$.

The case $b > 0$. From (10) it follows that $0 < \xi_3 < 1 < \xi_4$. Since $P(i,1)$ and $P(i,2)$ are probabilities, they must satisfy the normalization condition

$$\sum_{i=-1}^{\infty} P(i,1) + \sum_{i=-1}^{\infty} P(i,2) = 1.$$

A necessary condition for this equality is the limit relations $\lim P(i,1) = 0$ and $\lim P(i,2) = 0$ as $i \to \infty$. Otherwise, the series $\sum\limits_{i=-1}^{\infty} P(i,1)$ and $\sum\limits_{i=-1}^{\infty} P(i,2)$ will diverge. In view of the aforesaid, the general solution (9) with $D_1 = D_2 = D_4 = 0$ takes the form

$$P(i,1) = D_3\xi_3^i, \quad P(i,2) = B_3D_3\xi_3^i, \quad i = 0,1,\ldots. \tag{11}$$

We find the constant B_3. Substituting (11) into the first equation of system (5) with $\mu_1 = \lambda_1[1 - P_0(\lambda_2|\lambda_1)]$ and $\mu_2 = \lambda_2[1 - P_0(\lambda_1|\lambda_2)]$ gives $B_3 < 0$ after nontrivial transformations. Then (11) implies $D_3 < 0$. The inequality $D_3 < 0$ leads to a contradiction: $P(i,1) < 0$, $i \geqslant 0$; $P(i,2) > 0$, $i \geqslant 0$. Letting $D_3 = 0$ yields $P(i,1) = P(i,2) = 0$, $i \geqslant 0$; in other words, the contradiction is eliminated. Therefore, the final distribution $P(i,1)$, $P(i,2)$, $i \geqslant 0$, does not exist for $\lambda = \mu$ and, a fortiori, for $\lambda > \mu$.

We analyze the situation $\lambda < \mu$. Due to (6), the general solution of system (5) takes the form

$$\begin{gathered}P(i,1) = A_1\xi_1^i + A_2\xi_2^i + A_3\xi_3^i + A_4\xi_4^i, \\ P(i,2) = C_1A_1\xi_1^i + C_2A_2\xi_2^i + C_3A_3\xi_3^i + C_4A_4\xi_4^i, \quad i = 0,1,\ldots,\end{gathered} \tag{12}$$

where $P_s(i,1) = A_s\xi_s^i$ and $P_s(i,2) = C_sA_s\xi_s^i$ are partial solutions of system (5); their constants C_s and A_s, $s = \overline{1,4}$, are determined from the boundary conditions; $\xi_4 = 1$, ξ_1, ξ_2, and ξ_3 are the roots of the cubic equation in (6), positive real numbers: $0 < \xi_1 < \xi_2 < 1 < \xi_3$. In addition, the limit relations $\lim P(i,1) = \lim P(i,2) = 0$ as $i \to \infty$ hold (a necessary condition). Hence, $A_3 = A_4 = 0$, and the general solution of (12) takes the form

$$\begin{aligned} P(i,1) &= A_1\xi_1^i + A_2\xi_2^i, \\ P(i,2) &= C_1A_1\xi_1^i + C_2A_2\xi_2^i, \quad i = 0,1,\ldots. \end{aligned} \tag{13}$$

Substituting the partial solution $P_s(i,1) = A_s\xi_s^i$, $P_s(i,2) = C_sA_s\xi_s^i$, $i = 0,1,\ldots$, into the first equation of system (5), first for $s = 1$ and then for $s = 2$, we obtain the constants

$$C_s = -\frac{\mu_1\xi_s^2 - (\lambda_1+\mu_1)\xi_s + \lambda_1P_1(\lambda_1|\lambda_1)}{\lambda_2[P_0(\lambda_1|\lambda_2)\xi_s + P_1(\lambda_1|\lambda_2)]}, \quad s = 1,2. \tag{14}$$

The values A_i, $i = 1,2$, and the probabilities $P(-1,1)$ and $P(-1,2)$ are found using the boundary equations and the normalization condition. The stochastic graph cutsets

$$\begin{aligned} G_{-1,1} &= \{(-1,1;0,1),(0,1;-1,1),(-1,1;0,2),(-1,1;-1,2),(-1,2;-1,1)\}, \\ G_{-1,2} &= \{(-1,2;0,2),\ (0,2;-1,2),\ (-1,2;0,1),\ (-1,2;-1,1),(-1,1;-1,2)\}, \\ G &= \{(i,1;i+1,2),\ (i,1;i,2),\ (i,2;i+1,1),\ (i,2;i,1),\quad i = -1,0,1,\ldots\} \end{aligned}$$

determine the corresponding boundary equations:

$$\begin{aligned} \mu_1P(0,1) - \lambda_1P(-1,1) + \lambda_2P_0(\lambda_1|\lambda_2)P(-1,2) &= 0, \\ \mu_2P(0,2) - \lambda_2P(-1,2) + \lambda_1P_0(\lambda_2|\lambda_1)P(-1,1) &= 0, \\ \lambda_1[1 - P_1(\lambda_1|\lambda_1)]\sum_{i=-1}^{\infty}P(i,1) - \lambda_2[1 - P_1(\lambda_2|\lambda_2)]\sum_{i=-1}^{\infty}P(i,2) &= 0. \end{aligned} \tag{15}$$

Supplementing (15) with the normalization condition

$$P(-1,1) + P(-1,2) + \sum_{i=0}^{\infty}[P(i,1) + P(i,2)] = 1,$$

in view of (13), we arrive at the system of equations for the unknowns A_i, $i = 1,2$, $P(-1,1)$, and $P(-1,2)$. Solving (15) yields

$$\begin{aligned} P(-1,1) &= a_{11}A_1 + a_{12}A_2, \quad P(-1,2) = a_{21}A_1 + a_{22}A_2, \\ A_1 &= (1-\xi_1)\frac{\pi_1[C_2 + a_{22}(1-\xi_2)] - \pi_2[1 + a_{12}(1-\xi_2)]}{[1 + a_{11}(1-\xi_1)][C_2 + a_{22}(1-\xi_2)] - [1 + a_{12}(1-\xi_2)][C_1 + a_{21}(1-\xi_1)]}, \\ A_2 &= -(1-\xi_2)\frac{\pi_1[C_1 + a_{21}(1-\xi_1)] - \pi_2[1 + a_{11}(1-\xi_1)]}{[1 + a_{11}(1-\xi_1)][C_2 + a_{22}(1-\xi_2)] - [1 + a_{12}(1-\xi_2)][C_1 + a_{21}(1-\xi_1)]}, \\ a_{11} &= \frac{\mu_1 + \mu_2P_0(\lambda_1|\lambda_2)C_1}{\lambda_1[1 - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)]}, \quad a_{12} = \frac{\mu_1 + \mu_2P_0(\lambda_1|\lambda_2)C_2}{\lambda_1[1 - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)]}, \\ a_{21} &= \frac{\mu_2C_1 + \mu_1P_0(\lambda_2|\lambda_1)}{\lambda_2[1 - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)]}, \quad a_{22} = \frac{\mu_2C_2 + \mu_1P_0(\lambda_2|\lambda_1)}{\lambda_2[1 - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)]}. \end{aligned} \tag{16}$$

The values C_1 and C_2 are given by (14); the probabilities π_1 and π_2, by (1). The values ξ_1 and ξ_2 are the roots of the cubic equation in (6) ($0 < \xi_1 < \xi_2 < 1$).

Table 1. The probability of idle time $P(-1)$ depending on λ_1 for $b > 0$

$P_1(\lambda_1\|\lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0.780	0.718	0.680	0.651	0.627	0.616
1/6	0.787	0.728	0.693	0.667	0.645	0.636
1/8	0.790	0.734	0.700	0.675	0.654	0.645
1/10	0.792	0.737	0.704	0.679	0.659	0.651
1/12	0.794	0.739	0.706	0.682	0.663	0.655
1/13	0.794	0.739	0.707	0.684	0.664	0.656

Table 2. The mean queue length $E(I)$ depending on λ_1 for $b > 0$

$P_1(\lambda_1\|\lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0.052	0.097	0.145	0.196	0.249	0.276
1/6	0.047	0.085	0.125	0.167	0.209	0.231
1/8	0.045	0.080	0.116	0.153	0.191	0.210
1/10	0.043	0.076	0.110	0.145	0.180	0.198
1/12	0.042	0.074	0.106	0.140	0.173	0.190
1/13	0.042	0.073	0.105	0.138	0.171	0.187

Table 3. The expected number of requests $E(I+1)$ in the system depending on λ_1 for $b > 0$

$P_1(\lambda_1\|\lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0.272	0.379	0.465	0.545	0.622	0.659
1/6	0.260	0.357	0.432	0.500	0.564	0.595
1/8	0.254	0.346	0.416	0.478	0.537	0.565
1/10	0.251	0.340	0.406	0.466	0.521	0.547
1/12	0.249	0.335	0.400	0.457	0.510	0.536
1/13	0.248	0.334	0.398	0.454	0.506	0.531

Formulas (13) and (16) allow deriving explicit expressions for the numerical characteristics of the system: $P(-1)$ (the probability of idle time of the server), $E(I)$ (the mean queue length), and $E(I+1)$ (the expected number of requests in the system), where I is the random queue length in the QS. They are:

$$\begin{aligned} P(-1) &= (a_{11}+a_{21})A_1 + (a_{12}+a_{22})A_2, \\ E(I) &= A_1(1+C_1)\frac{\xi_1}{(1-\xi_1)^2} + A_2(1+C_2)\frac{\xi_2}{(1-\xi_2)^2}, \\ E(I+1) &= \frac{A_1(1+C_1)}{(1-\xi_1)^2} + \frac{A_2(1+C_2)}{(1-\xi_2)^2}, \end{aligned} \tag{17}$$

where C_1 and C_2 are given by (14); A_1, A_2, a_{11}, a_{21}, a_{12}, and a_{22}, by (16). The values ξ_1 and ξ_2 are the roots of the cubic equation in (6) ($0 < \xi_1 < \xi_2 < 1$).

The initial data for calculating the numerical characteristics (17), see the tables below, are chosen to assess the degree of their correspondence to the physical understanding of the service process in the QS.

Tables 1–3 present the characteristics $P(-1)$, $E(I)$, and $E(I+1)$ (17) depending on the parameter λ_1 ($\lambda_1 = 2, 4, \ldots, 10, 11$) under the fixed parameter values $\lambda_2 = 1$, $\mu_1 = 12$, $\mu_2 = 2$; $P_1(\lambda_1|\lambda_1) = P_1(\lambda_2|\lambda_1) = P_0(\lambda_2|\lambda_1) = P_1(\lambda_2|\lambda_2) = \frac{1}{3}$ for $b > 0$ and $P_1(\lambda_1|\lambda_2) = \frac{1}{4}$ ($P_0(\lambda_1|\lambda_2) = \frac{5}{12}$); $P_1(\lambda_1|\lambda_2) = \frac{1}{6}$

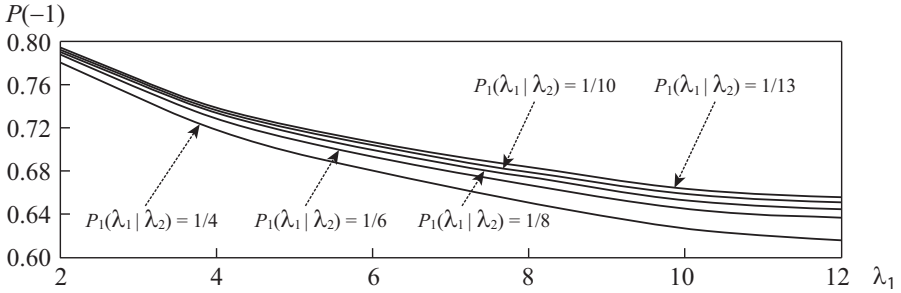


Fig. 2. The probability of idle time $P(-1)$ depending on λ_1 for $b > 0$.

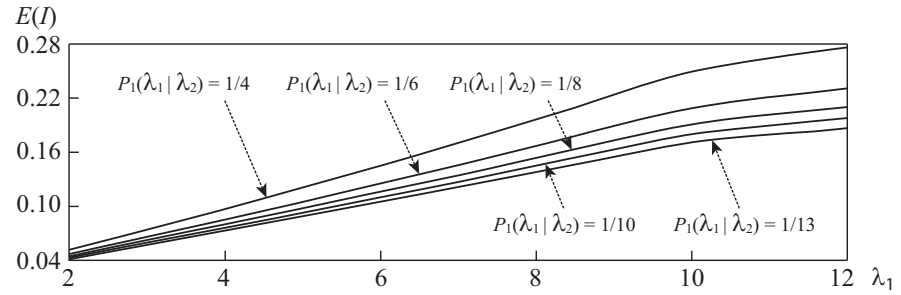


Fig. 3. The mean queue length $E(I)$ depending on λ_1 for $b > 0$.

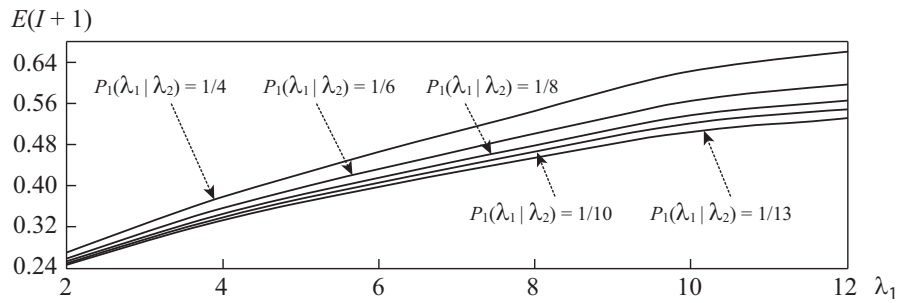


Fig. 4. The expected number of requests $E(I+1)$ in the system depending on λ_1 for $b > 0$

$(P_0(\lambda_1|\lambda_2) = \frac{1}{2})$; $P_1(\lambda_1|\lambda_2) = \frac{1}{8}$ $(P_0(\lambda_1|\lambda_2) = \frac{13}{24})$; $P_1(\lambda_1|\lambda_2) = \frac{1}{10}$ $(P_0(\lambda_1|\lambda_2) = \frac{17}{30})$; $P_1(\lambda_1|\lambda_2) = \frac{1}{12}$ $(P_0(\lambda_1|\lambda_2) = \frac{7}{12})$; $P_1(\lambda_1|\lambda_2) = \frac{1}{13}$ $(P_0(\lambda_1|\lambda_2) = \frac{23}{39})$.

The behavior of these characteristics depending on the parameter λ_1 for $b > 0$ matches the physical understanding of the service process in the single-server QS with an incoming correlated MAP request flow.

Figures 2–4 show the graphs of the numerical characteristics (17) plotted on the numerical values of Tables 1–3, respectively.

The case $b < 0$. First of all, we investigate the existence of the stationary mode, i.e., the situation $\lambda = \mu$. Then it follows from (10) that $\xi_3 < 0$ and $\xi_4 > 1$; similar to the case $b > 0$, the general solution of the system takes the form (11). Since $\xi_3 < 0$, this fact entails the negative probability $P(i, 1)$ for $i = 1, 3, \ldots$, (an obvious contradiction to its definition). This contradiction is eliminated by letting $D_3 = 0$: $P(i, 1) = P(i, 2) = 0$, $i \geqslant 0$. Therefore, in the case $b < 0$, the final distribution $P(i, 1)$, $P(i, 2)$, $i \geqslant 0$, does not exist for $\lambda = \mu$ and, a fortiori, for $\lambda > \mu$.

Now we study the situation $\lambda < \mu$. Due to (6), the general solution of system (5) takes the form (12). In the case $b < 0$, we have $\xi_4 = 1$, ξ_1, ξ_2, and ξ_3 are the real roots of the cubic equation in (6): $\xi_1 < 0$, $0 < \xi_2 < 1 < \xi_3$. Hence, it follows that $A_1 = A_3 = A_4 = 0$ in (12), and the general

Table 4. The probability of idle time $P(-1)$ depending on λ_1 for $b < 0$

$P_1(\lambda_2\|\lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0.796	0.729	0.687	0.654	0.627	0.615
1/6	0.815	0.748	0.705	0.672	0.645	0.633
1/8	0.824	0.757	0.714	0.681	0.654	0.642
1/10	0.830	0.762	0.719	0.686	0.659	0.647
1/12	0.833	0.765	0.722	0.690	0.662	0.650
1/13	0.834	0.767	0.724	0.691	0.664	0.652

Table 5. The mean queue length $E(I)$ depending on λ_1 for $b < 0$

$P_1(\lambda_2\|\lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0.044	0.091	0.143	0.200	0.260	0.291
1/6	0.035	0.077	0.124	0.177	0.232	0.261
1/8	0.031	0.071	0.116	0.166	0.220	0.248
1/10	0.029	0.067	0.112	0.161	0.213	0.241
1/12	0.027	0.065	0.109	0.157	0.209	0.236
1/13	0.027	0.064	0.108	0.156	0.207	0.234

Table 6. The expected number of requests $E(I+1)$ in the system depending on λ_1 for $b < 0$

$P_1(\lambda_2\|\lambda_2)$ \ λ_1	2	4	6	8	10	11
1/4	0.249	0.362	0.457	0.546	0.633	0.675
1/6	0.219	0.329	0.419	0.504	0.587	0.528
1/8	0.206	0.314	0.402	0.485	0.566	0.606
1/10	0.199	0.305	0.392	0.474	0.554	0.594
1/12	0.194	0.300	0.386	0.467	0.547	0.586
1/13	0.193	0.298	0.384	0.465	0.544	0.583

solution of (12) is written as

$$P(i,1) = A_2\xi_2^i, \quad P(i,2) = C_2A_2\xi_2^i, \quad i = 0,1,\ldots. \tag{18}$$

In (18), the constant C_2 is given by (14) for $s = 2$. The constant A_2 and the probabilities $P(-1,1)$ and $P(-1,2)$ are determined using Eqs. (15) and the normalization condition. As a result,

$$\begin{gathered} P(-1,1) = a_{12}A_2; \quad P(-1,2) = a_{22}A_2; \\ A_2 = \frac{1-\xi_2}{1+C_2+(a_{12}+a_{22})(1-\xi_2)}, \end{gathered} \tag{19}$$

where C_2 is given by (14) for $s = 2$; a_{12} and a_{22}, by (16). The value ξ_2 is the root of the cubic equation in (6) $(0 < \xi_2 < 1)$.

Formulas (18) and (19) allow deriving the system characteristics:

$$\begin{gathered} P(-1) = (a_{12}+a_{22})A_2; \\ E(I) = A_2\xi_2\frac{1+C_2}{(1-\xi_2)^2}, \quad E(I+1) = \frac{(1+C_2)A_2}{(1-\xi_2)^2}, \end{gathered} \tag{20}$$

where C_2 is given by (14) for $s = 2$; a_{12} and a_{22}, by (16); A_2, by (19). The value ξ_2 is the root of the cubic equation in (6) $(0 < \xi_2 < 1)$.

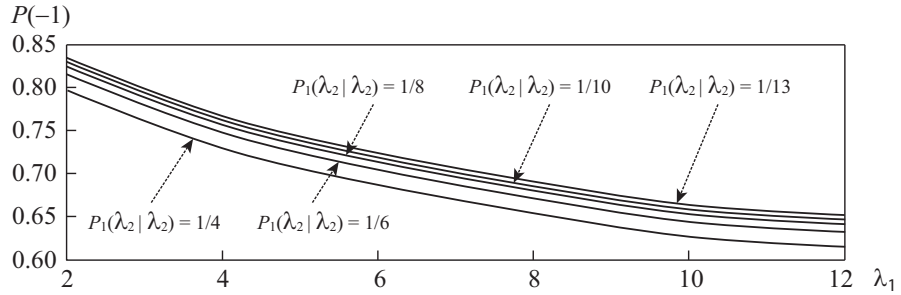


Fig. 5. The probability of idle time $P(-1)$ depending on λ_1 for $b < 0$.

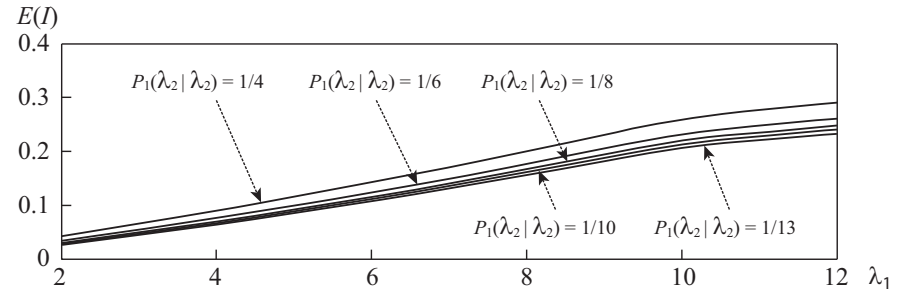


Fig. 6. The mean queue length $E(I)$ depending on λ_1 for $b < 0$.

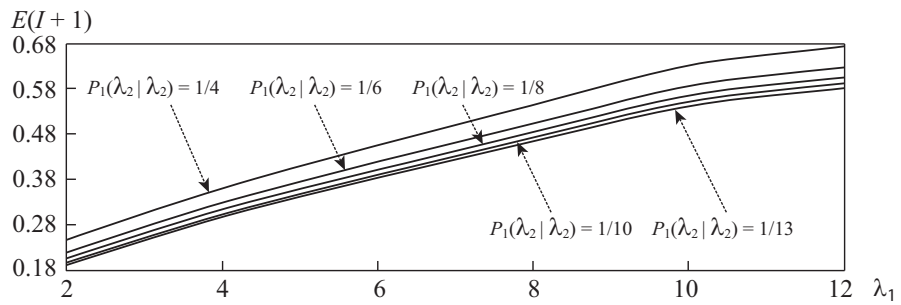


Fig. 7. The expected number of requests $E(I+1)$ in the system depending on λ_1 for $b < 0$.

Tables 4–6 present the characteristics $P(-1)$, $E(I)$, and $E(I+1)$ (20) depending on the parameter λ_1 ($\lambda_1 = 2, 4, \ldots, 10, 11$) for the fixed parameter values $\lambda_2 = 1$, $\mu_1 = 12$, $\mu_2 = 2$; $P_1(\lambda_1|\lambda_1) = P_1(\lambda_2|\lambda_1) = P_0(\lambda_2|\lambda_1) = P_1(\lambda_1|\lambda_2) = \frac{1}{3}$ for $b < 0$ and $P_1(\lambda_2|\lambda_2) = \frac{1}{4}$ ($P_0(\lambda_1|\lambda_2) = \frac{5}{12}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{6}$ ($P_0(\lambda_1|\lambda_2) = \frac{1}{2}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{8}$ ($P_0(\lambda_1|\lambda_2) = \frac{13}{24}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{10}$ ($P_0(\lambda_1|\lambda_2) = \frac{17}{30}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{12}$ ($P_0(\lambda_1|\lambda_2) = \frac{7}{12}$); $P_1(\lambda_2|\lambda_2) = \frac{1}{13}$ ($P_0(\lambda_1|\lambda_2) = \frac{23}{39}$).

Figures 5–7 show the graphs of the numerical characteristics (20) plotted on the numerical values of Tables 4–6, respectively.

The behavior of these characteristics depending on the parameter λ_1 for $b < 0$ also matches the physical understanding of the service process in the single-server QS with an incoming correlated MAP request flow.

4. A SPECIAL CASE: A RECURRENT MAP REQUEST FLOW

In this special case, we have $b = 0$, which implies the recurrence of the MAP request flow; see (3). Consider conditions for the existence of the stationary probabilities $P(i, 1)$ and $P(i, 2)$, $i \geqslant 0$. In the

Table 7. The probability of idle time $P(-1)$ depending on λ_1 for $b = 0$

$P_1(\lambda_1\|\lambda_1)$ \ λ_1	2	4	6	8	10	11
1/4	0.831	0.781	0.750	0.726	0.707	0.698
1/6	0.888	0.854	0.833	0.818	0.805	0.799
1/8	0.916	0.891	0.875	0.863	0.854	0.850
1/10	0.933	0.913	0.900	0.891	0.883	0.880
1/12	0.944	0.927	0.917	0.909	0.902	0.900
1/13	0.949	0.933	0.923	0.916	0.910	0.907

Table 8. The mean queue length $E(I)$ depending on λ_1 for $b = 0$

$P_1(\lambda_1\|\lambda_1)$ \ λ_1	2	4	6	8	10	11
1/4	0.030	0.056	0.083	0.111	0.139	0.152
1/6	0.013	0.023	0.033	0.043	0.053	0.058
1/8	0.007	0.012	0.018	0.023	0.028	0.030
1/10	0.004	0.008	0.011	0.014	0.017	0.019
1/12	0.003	0.005	0.008	0.010	0.0129	0.013
1/13	0.002	0.005	0.006	0.008	0.010	0.011

Table 9. The expected number of requests $E(I + 1)$ in the system depending on λ_1 for $b = 0$

$P_1(\lambda_1\|\lambda_1)$ \ λ_1	2	4	6	8	10	11
1/4	0.199	0.275	0.333	0.385	0.432	0.454
1/6	0.124	0.169	0.200	0.226	0.248	0.259
1/8	0.091	0.122	0.143	0.160	0.174	0.181
1/10	0.071	0.095	0.111	0.124	0.134	0.139
1/12	0.059	0.078	0.091	0.101	0.109	0.113
1/13	0.054	0.072	0.083	0.092	0.100	0.103

situation $\lambda = \mu$, the characteristic Eq. (8) takes the form

$$\lambda_1\lambda_2(\xi-1)^2\xi\Big\{[1-P_0(\lambda_2|\lambda_1)][1-P_0(\lambda_1|\lambda_2)]\xi-[2-P_0(\lambda_1|\lambda_2)-P_0(\lambda_2|\lambda_1)]\Big\}=0, \tag{21}$$

and the general solution of system (5) is (9). The characteristic Eq. (21) has the roots

$$\xi_1=\xi_2=1,\quad \xi_3=0,\quad \xi_4=\frac{1}{1-P_0(\lambda_1|\lambda_2)}+\frac{1}{1-P_0(\lambda_2|\lambda_1)}>1. \tag{22}$$

In view of (22), letting $D_1 = D_2 = D_4 = 0$ in the general solution (9) yields $P(i,1) = P(i,2) = 0$, $i \geqslant 0$. Therefore, in the case $b = 0$, the final distribution $P(i,1)$, $P(i,2)$, $i \geqslant 0$, does not exist for $\lambda = \mu$ and, a fortiori, for $\lambda > \mu$.

We analyze the situation $\lambda < \mu$. In the case $b = 0$, the characteristic Eq. (6) is written as

$$\begin{aligned}&\xi(\xi-1)\Big\{\mu_1\mu_2\xi^2-[\lambda_1\mu_2+\mu_1(\lambda_2+\mu_2)]\xi\\&+[\lambda_1\lambda_2+\lambda_1\mu_2P_1(\lambda_1|\lambda_1)+\lambda_2\mu_1P_1(\lambda_2|\lambda_2)-\lambda_1\lambda_2P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)]\Big\}=0.\end{aligned} \tag{23}$$

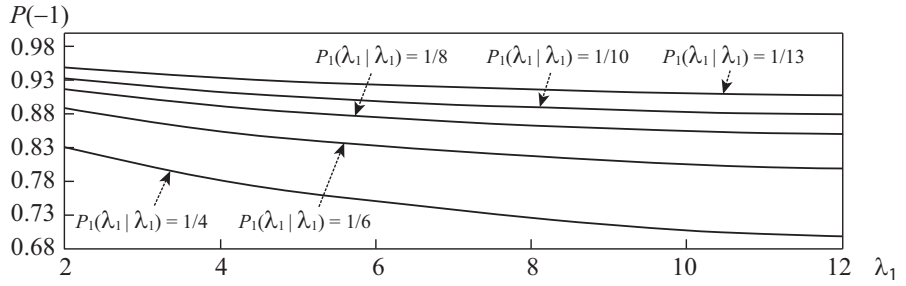


Fig. 8. The probability of idle time $P(-1)$ depending on λ_1 for $b = 0$.

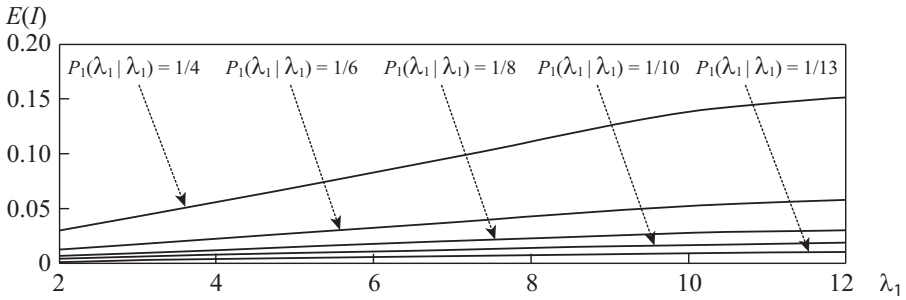


Fig. 9. The mean queue length $E(I)$ depending on λ_1 for $b = 0$.

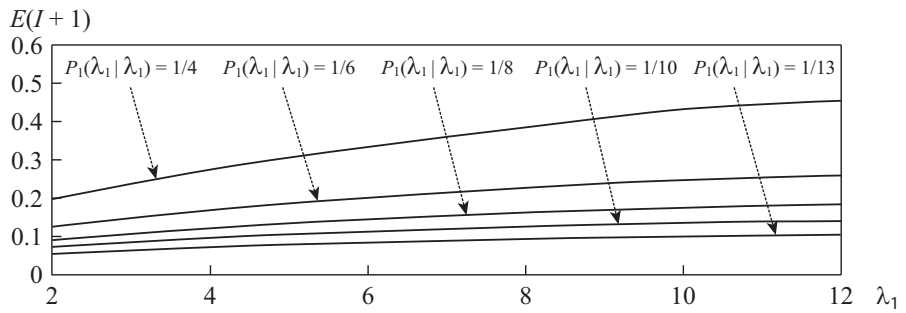


Fig. 10. The expected number of requests $E(I+1)$ in the system depending on λ_1 for $b = 0$.

The characteristic Eq. (23) has the roots $\xi_3 = 0$, $\xi_4 = 1$, and

$$\xi_{1,2} = \Big\{(\lambda_1\mu_2 + \lambda_2\mu_1 + \mu_1\mu_2) \mp \Big[(\lambda_1\mu_2 + \lambda_2\mu_1 + \mu_1\mu_2)^2 - 4\mu_1\mu_2\Big(\lambda_1\lambda_2 + \lambda_1\mu_2 P_1(\lambda_1|\lambda_1) + \lambda_2\mu_1 P_1(\lambda_2|\lambda_2) - \lambda_1\lambda_2 P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)\Big)\Big]^{\frac{1}{2}}\Big\}/2\mu_1\mu_2, \tag{24}$$

$0 < \xi_1 < 1 < \xi_2$. Due to (23) and (24), the general solution (12) of system (5) takes the form

$$P(i,1) = A_1\xi_1^i, \quad P(i,2) = C_1A_1\xi_1^i, \quad i = 0,1,\ldots. \tag{25}$$

In (25), the constant C_1 is given by (14) for $s = 1$. The constant A_1 and the probabilities $P(-1,1)$ and $P(-1,2)$ are determined using Eqs. (15) and the normalization condition. As a result,

$$P(-1,1) = a_{11}A_1; \quad P(-1,2) = a_{21}A_1;$$
$$A_1 = \frac{1-\xi_1}{1 + C_1 + (a_{11} + a_{21})(1-\xi_1)}, \tag{26}$$

where C_1 is given by (14) for $s = 1$; a_{21} and a_{11}, by (16); ξ_1, by (24).

Formulas (25) and (26) allow deriving the system characteristics:

$$P(-1) = (a_{21} + a_{11})A_1;$$
$$E(I) = A_1\xi_1\frac{1+C_1}{(1-\xi_1)^2}, \quad E(I+1) = \frac{(1+C_1)A_1}{(1-\xi_1)^2}, \tag{27}$$

where C_1 is given by (14) for $s = 1$; a_{21} and a_{11}, by (16); A_1, by (26); ξ_1, by (24).

Tables 7–9 present the characteristics $P(-1)$, $E(I)$, and $E(I+1)$ (27) depending on the parameter λ_1 ($\lambda_1 = 2, 4, \ldots, 10, 11$) for the fixed parameter values $\lambda_2 = 1$, $\mu_1 = 12$, $\mu_2 = 2$ for $b = 0$ and $(P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{4}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{1}{2})$; $(P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{6}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{2}{3})$; $(P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{8}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{3}{4})$; $(P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{10}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{4}{5})$; $(P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{12}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{5}{6})$; $(P_1(\lambda_i|\lambda_i) = P_1(\lambda_j|\lambda_i) = \frac{1}{13}$; $P_0(\lambda_1|\lambda_2) = P_0(\lambda_2|\lambda_1) = \frac{11}{13})$; $i, j = 1, 2$ $(i \neq j)$.

As in the cases $b > 0$ and $b < 0$, the behavior of these characteristics depending on the parameter λ_1 for $b = 0$ matches the physical understanding of the service process in the single-server QS with an incoming correlated MAP request flow.

Figures 8–10 show the graphs of the numerical characteristics (27) plotted on the numerical values of Tables 7–9, respectively.

5. CONCLUSIONS

The paper has considered a single-server QS with an incoming correlated MAP request flow with two states. The analysis problems formulated in Section 2 have been completely solved for this queueing system.

Let us summarize the results and present the final formulas.

The case $b > 0$. The stationary probabilities $P(i, 1)$ and $P(i, 2)$, $i = 0, 1, \ldots$, are given by $P(i, 1) = A_1\xi_1^i + A_2\xi_2^i$ and $P(i, 2) = C_1A_1\xi_1^i + C_2A_2\xi_2^i$, respectively, where: the constants C_s, $s = 1, 2$, are calculated using (14); ξ_1 and ξ_2 $(0 < \xi_1 < \xi_2 < 1)$ are the roots of the cubic Eq. (6); the probabilities $P(-1, 1)$ and $P(-1, 2)$ as well as the constants A_1 and A_2 are calculated using (16). The numerical characteristics $P(-1)$, $E(I)$, and $E(I+1)$ are given by (17).

The case $b < 0$. The stationary probabilities $P(i, 1)$ and $P(i, 2)$, $i = 0, 1, \ldots$, are given by $P(i, 1) = A_2\xi_2^i$ and $P(i, 2) = C_2A_2\xi_2^i$, respectively, where: the constant C_2 is calculated using (14) for $s = 2$; ξ_2 $(0 < \xi_2 < 1)$ is the root of the cubic Eq. (6); the probabilities $P(-1, 1)$ and $P(-1, 2)$ as well as the constant A_2 are calculated using (19). The numerical characteristics $P(-1)$, $E(I)$, and $E(I+1)$ are given by (20).

The case $b = 0$. The stationary probabilities $P(i, 1)$ and $P(i, 2)$, $i = 0, 1, \ldots$, are given by $P(i, 1) = A_1\xi_1^i$ and $P(i, 2) = C_1A_1\xi_1^i$, respectively, where: the constant C_1 is calculated using (14) for $s = 1$; ξ_1 $(0 < \xi_1 < 1)$ is the root (24) of the characteristic Eq. (23); the probabilities $P(-1, 1)$ and $P(-1, 2)$ as well as the constant A_1 are calculated using (26). The numerical characteristics $P(-1)$, $E(I)$, and $E(I+1)$ are given by (27).

Formulas (17), (20), and (27) have been derived by introducing an additional variable and using the method of transition intensity diagrams (the method of stochastic graph cutsets) [8]. The case $b = 0$ degenerates the incoming correlated MAP request flow into a recurrent one.

The analytical formulas (17), (20), and (27) serve to calculate the numerical characteristics of an MAP request flow with given parameters without involving numerical methods. The graphs of the numerical characteristics presented above match the physical understanding of the service process in this QS.

REFERENCES

1. Cox, D.R., The Analysis of Non-Markovian Stochastic Processes by the Inclusion of Supplementary Variables, *Proceedings of the Cambridge Philosophical Society*, 1955, vol. 51, no. 3, pp. 433–441.
2. Kingman, J.F.C., On Doubly Stochastic Poisson Processes, *Proceedings of the Cambridge Philosophical Society*, 1964, vol. 60, no. 4, pp. 923–930.
3. Basharin, G.P., Kokotushkin, V.A., and Naumov, V.A., On the Equivalent Change Method to Calculate Fragments of Communication Networks. Part 1, *Izv. Akad. Nauk SSSR. Tekh. Kibern.*, 1979, no. 6, pp. 92–99.
4. Basharin, G.P., Kokotushkin, V.A., and Naumov, V.A., On the Equivalent Change Method to Calculate Fragments of Communication Networks. Part 2, *Izv. Akad. Nauk SSSR. Tekh. Kibern.*, 1980, no. 1, pp. 55–61.
5. Neuts, M.F., A Versatile Markov Point Process, *J. Appl. Probab.*, 1979, vol. 16, no. 4, pp. 764–779.
6. Lucantoni, D.M., New Results on the Single Server Queue with a Batch Markovian Arrival Process, *Communications in Statistics. Stochastic Models*, 1991, vol. 7, no. 1, pp. 1–46.
7. Lucantoni, D.M. and Neuts, M.F., Some Steady-State Distributions for the MAP/SM/1 Queue, *Communications in Statistics. Stochastic Models*, 1994, vol. 10, no. 3, pp. 575–598.
8. Vishnevsky, V.M., Dudin, A.N., and Klimenok, V.I., *The Theory of Queuing Systems with Correlated Flows*, Heidelberg: Springer, 2020.
9. Gortsev, A.M. and Solov'ev, A.A., The Probability of Error in Estimating the States of Physical Event Flows, *Izv. Vuzov. Fizika*, 2016, vol. 59, no. 5, pp. 54–60.
10. Gortsev, A.M., Nezhel'skaya, L.A., and Solov'ev, A.A., Optimal State Estimation in MAP Event Flows with Unextendable Dead Time, *Autom. Remote Control*, 2012, vol. 73, no. 8, pp. 1316–1326.
11. Gortzev, A.M. and Soloviev, A.A., The Joint Density of Probability for Intervals of MAP Event Flows and Conditions of Its Recurrence, *Tomsk State Univ. J. Control Comp. Sci.*, 2012, no. 3(20), pp. 32–41.
12. Gortzev, A.M. and Solovev, A.A., Comparison of ML- and M-Estimates for Unextandable Dead Time in MAP Event Flow, *Tomsk State Univ. J. Control Comp. Sci.*, 2015, no. 4(33), pp. 13–22.
13. Gortsev, A.M. and Solov'ev, A.A., The Joint Probability Density of Physical Event Flow Intervals with Unextendable Dead Time, *Izv. Vuzov. Fizika*, 2014, vol. 57, no. 7, pp. 103–111.
14. Gortsev, A.M. and Solov'ev, A.A., A Maximum Likelihood Estimate for Unextendable Dead Time in a Physical Event Flow, *Izv. Vuzov. Fizika*, 2015, vol. 58, no. 11, pp. 141–149.
15. Bin Sun, Dudin, S.A., Dudina, O.S., and Dudin, A.N., A Customer Service Model in an Adaptive-Modulation Mobile Communication Cell with Allowance for Random Environment, *Autom. Remote Control*, 2021, vol. 82, no. 5, pp. 812–826.
16. Gortsev, A.M. and Nezhelskaya, L.A., Analytical Investigation of a Single-Channel QS with Incoming Asynchronous Event Flow, *Autom. Remote Control*, 2022, vol. 83, no. 8, pp. 1200–1212.
17. Nezhel'skaya, L.A., *Otsenka sostoyanii dvazhdy stokhasticheskikh potokov sobytii* (State Estimation for Doubly Stochastic Event Flows), Tomsk: Tomsk State Univ., 2020.
18. Medvedev, G.A., Analysis of Stochastic Graphs Describing the Behavior of Stepped Automatic Search Systems, *Avtomat. Vychisl. Tekh.*, 1968, no. 4, pp. 27–30.

This paper was recommended for publication by V.M. Vishnevsky, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 778–790.

STOCHASTIC SYSTEMS

On Continuous Random Processes with Fuzzy States

V. L. Khatskevich

Military Training and Research Center of the Air Force,
Air Force Academy named after N.E. Zhukovsky and Yu.A. Gagarin, Voronezh, Russia
e-mail: vlkhats@mail.ru

Received August 3, 2022
Revised January 12, 2023
Accepted March 30, 2023

Abstract—Continuous random processes with fuzzy states are studied. The properties of their numerical characteristics (expectations and correlation functions) corresponding to those of numerical random processes are established. The results obtained are based on the properties of fuzzy random variables. Applications to the problem of transforming a random signal with fuzzy states by a linear dynamic system are considered.

Keywords: continuous random processes, fuzzy states, fuzzy expectations, correlation functions

DOI: 10.25728/arcRAS.2023.73.23.001

1. INTRODUCTION

Fuzzy modeling is actively used in various applications with incomplete or weakly formalized initial data [1–3]. In particular, note the recent works on the fuzzy approach in control theory [4, 5].

On the other hand, when studying dynamic processes under limited initial information, a possible approach is to treat their parameters as realizations of some random processes [6].

This paper combines both approaches: continuous random processes with fuzzy states are investigated below. More precisely put, time and the set of possible fuzzy states are supposed to be continuous. (In other words, this set is uncountable.) The cross-section of a continuous fuzzy random process at any time instant is a fuzzy random variable. This study involves well-known results from the theory of fuzzy random variables [7–9] and the differentiability and integrability of fuzzy-valued functions ([10, 11] and [12, 13], respectively).

The properties established below are modifications of the well-known results [6, Chapter 1] on the expectations and correlation functions of standard continuous random processes. Seemingly, they have not been observed previously. In this case, an important role is played by fuzzy expectations (nonrandom fuzzy-valued functions of time) reflecting the trends of fuzzy random processes.

One application considered in this paper is the problem of transforming a fuzzy random signal by a linear dynamic system described by a linear differential equation with constant coefficients. Here, the characteristics of an input signal (fuzzy expectation and correlation function) are used to determine similar characteristics of an output signal. These results develop the well-known results for real stationary random processes; for example, see [6, Chapter 7].

Note the difference between the approach and results presented below and the works on random processes with discrete fuzzy states and continuous time (e.g., [14, 15]) that discussed the properties of the probabilities of discrete fuzzy states depending on time.

In what follows, a fuzzy number $\tilde{z}$ defined on the universal space R of real numbers will be understood as a set of ordered pairs $(x, \mu_{\tilde{z}}(x))$, where the membership function $\mu_{\tilde{z}} : R \to [0, 1]$ $\forall\, x \in R$ determines the grade of membership to the set $\tilde{z}$ [1, Chapter 5].

The interval representation of fuzzy numbers will be employed: to each fuzzy number we assign the set of its α-intervals.

As is known, the α-level set of a fuzzy number $\tilde{z}$ with a membership function $\mu_{\tilde{z}}(x)$ is defined as

$$Z_\alpha = \{x \,|\, \mu_{\tilde{z}}(x) \geqslant \alpha\} \quad (\alpha \in (0,1]), \qquad Z_0 = cl\{x \,|\, \mu\tilde{z}(x) > 0\},$$

where cl indicates the closure of an appropriate set.

Assume that all α-levels of a fuzzy number are closed and bounded intervals on the entire real axis. Let $z^-(\alpha)$ and $z^+(\alpha)$ denote the left and right bounds of an interval $Z_\alpha : Z_\alpha = [z^-(\alpha), z^+(\alpha)]$. Sometimes, $z^-(\alpha)$ and $z^+(\alpha)$ are called the left and right α-indices of a fuzzy number, respectively. Suppose that they are measurable and bounded on $[0,1]$. The set of such fuzzy numbers will be denoted by J.

The sum of fuzzy numbers is understood as a fuzzy number whose indices are the sums of the corresponding indices of the terms. Multiplication of a fuzzy number by a positive number means multiplying its indices by this number. Multiplication by a negative real number means multiplying the indices by this number and reversing them. Equality for fuzzy numbers is understood as equality for all the corresponding α-indices $\forall\, \alpha \in [0,1]$.

The set J can be metrized in various ways. In particular, for fuzzy numbers $\tilde{z}_1, \tilde{z}_2 \in J$, one metric is given by

$$d(\tilde{z}_1, \tilde{z}_2) = \left(\int_0^1 \left((z_1^-(\alpha) - z_2^-(\alpha))^2 + (z_1^+(\alpha) - z_2^+(\alpha))^2 \right) d\alpha \right)^{1/2}, \tag{1}$$

where $z_i^-(\alpha)$ and $z_i^+(\alpha)$ are the left and right indices of $\tilde{z}_i$ $(i = 1,2)$, respectively; for example, see [16].

2. FUZZY EXPECTATION, THE EXPECTATION AND COVARIANCE OF FUZZY RANDOM VARIABLES

Let (Ω, Σ, P) be a probability space with Ω as the set of elementary events, Σ as a σ-algebra consisting of the subsets of the set Ω, and P as a probability measure.

Consider a mapping $\widetilde{X} : \Omega \to J$. For a fixed event $\omega \in \Omega$, its α-level intervals $X_\alpha(\omega)$ are given by

$$X_\alpha(\omega) = \left\{ r \in R : \mu_{\widetilde{X}(\omega)}(r) \geqslant \alpha \right\} \quad \alpha \in (0,1], \qquad X_0(\omega) = cl\left\{ r \in R : \mu_{\widetilde{X}(\omega)}(r) > 0 \right\},$$

where $\mu_{\widetilde{X}(\omega)}(r)$ is the membership function of a fuzzy number $\widetilde{X}(\omega)$. An interval $X_\alpha(\omega)$ can be written as $X_\alpha(\omega) = [X^-(\omega,\alpha), X^+(\omega,\alpha)]$, where the bounds $X^-(\omega,\alpha)$ and $X^+(\omega,\alpha)$ are called the left and right indices of the mapping $\widetilde{X}$, respectively.

A mapping $\widetilde{X} : \Omega \to J$ is called a fuzzy random variable (FRV) if the real-valued functions $X^\pm(\omega,\alpha)$ are measurable for all $\alpha \in [0,1]$; for example, see [7, 8]. In this case, the indices are real random variables for any $\alpha \in [0,1]$.

We will study the class $\mathcal{A}$ of all fuzzy random variables for which the indices $X^-(\omega,\alpha)$ and $X^+(\omega,\alpha)$ are square summable functions on $\Omega \times [0,1]$.

Note that an FRV can be interpreted as a random element in the metric space J with the metric (1).

For an FRV $\widetilde{X}(\omega)$, let

$$x^-(\alpha) = \int_\Omega X^-(\omega,\alpha) dP, \qquad x^+(\alpha) = \int_\Omega X^+(\omega,\alpha) dP. \tag{2}$$

A fuzzy number with α-indices (2) is called the fuzzy expectation of an FRV $\widetilde{X}$; for example, see [9]. In what follows, it will be denoted by $M(\widetilde{X})$ and its indices by $[M(\widetilde{X})]^{\pm}_{\alpha}$.

For FRVs $\widetilde{X}_1$ and $\widetilde{X}_2$ on the set $\mathcal{X}$, we consider the metric

$$\rho(\widetilde{X}_1,\widetilde{X}_2)=\left(\int\limits_0^1\int\limits_\Omega\left((X_1^-(\omega,\alpha)-X_2^-(\omega,\alpha))^2+(X_1^+(\omega,\alpha)-X_2^+(\omega,\alpha))^2\right)dPd\alpha\right)^{1/2}. \tag{3}$$

The fuzzy expectation defined by (2) has similar properties as the expectations of real random variables.

Proposition 1 (e.g., see [17, 18]). *The fuzzy expectation defined by* (2) *possesses the following properties:*

1. *If $\widetilde{X}(\omega)=\widetilde{X}$ for almost all $\omega\in\Omega$, then $M(\widetilde{X})=\widetilde{X}$ (idempotency).*
2. *The fuzzy expectation $M:\mathcal{X}\to J$ is additive, i.e., $M(\widetilde{X}_1+\widetilde{X}_2)=M(\widetilde{X}_1)+M(\widetilde{X}_2)\ \forall\widetilde{X}_1,\widetilde{X}_2\in\mathcal{X}$.*
3. *The fuzzy expectation $M:\mathcal{X}\to J$ is homogeneous, i.e., $M(C\widetilde{X})=CM(\widetilde{X})\ \forall\widetilde{X}\in\mathcal{X},\ \forall C\in R$.*
4. *The fuzzy expectation $M:\mathcal{X}\to J$ is continuous as a mapping from $\mathcal{X}$ with the metric* (3) *into J with the metric* (1).
5. *The fuzzy expectation $M(\widetilde{X})$ of a given FRV $\widetilde{X}\in\mathcal{X}$ satisfies the inequality*

$$\rho^2(\widetilde{X},M(\widetilde{X}))=\int\limits_0^1\int\limits_\Omega\left(\left(X^-(\omega,\alpha)-[M(\widetilde{X})]^-_\alpha\right)^2+\left(X^+(\omega,\alpha)-[M(\widetilde{X})]^+_\alpha\right)^2\right)dPd\alpha$$
$$\leqslant\int\limits_0^1\int\limits_\Omega\left((X^-(\omega,\alpha)-V^-_\alpha)^2+(X^+(\omega,\alpha)-V^+_\alpha)^2\right)dPd\alpha=\rho^2(\widetilde{X},\widetilde{V})\quad\forall\widetilde{V}\in J,$$

where $X^{\pm}(\omega,\alpha)$ and $V^{\pm}_\alpha$ are the indices of the FRV $\widetilde{X}(\omega)$ and the fuzzy number $\widetilde{V}$, respectively (the extremal property).

Consider now the defuzzification of fuzzy expectations.

According to [19], within the interval approach, the mean value of a fuzzy number $\tilde{z}$ is given by

$$z_{\text{mean}}=\frac{1}{2}\int\limits_0^1(z^-(\alpha)+z^+(\alpha))d\alpha, \tag{4}$$

where $z^{\pm}(\alpha)$ are the indices of $\tilde{z}$.

In view of (4), the expectation $m(\widetilde{X})$ of an FRV $\widetilde{X}\in\mathcal{X}$ is defined as an averaging functional by

$$m(\widetilde{X})=\frac{1}{2}\int\limits_0^1\left([M(\widetilde{X})]^-(\alpha)+[M(\widetilde{X})]^+(\alpha)\right)d\alpha, \tag{5}$$

where $M^{\pm}(\alpha)$ are the indices of the fuzzy expectation $M(\widetilde{X})$ given by (2). As a matter of fact, this is a defuzzification method for fuzzy expectations.

The next result is immediate from the definition (5) and Proposition 1.

Proposition 2 (e.g., see [17, 18]). *The expectation* (5) *of an FRV possesses the following properties:*

1. *If $\widetilde{X}(\omega)=\widetilde{X}$ for almost all $\omega\in\Omega$, then $m(\widetilde{X})=X_{\text{mean}}$ (idempotency).*
2. *The expectation $m:\mathcal{X}\to R$ is additive, i.e., $m(\widetilde{X}_1+\widetilde{X}_2)=m(\widetilde{X}_1)+m(\widetilde{X}_2)\ \forall\widetilde{X}_1,\widetilde{X}_2\in\mathcal{X}$.*

3. *The expectation $m : \mathcal{A} \to R$ is homogeneous, i.e., $m(C\widetilde{X}) = Cm(\widetilde{X})\ \forall\, \widetilde{X} \in \mathcal{A}$, $\forall\, C \in R$.*

4. *The expectation $m : \mathcal{X} \to R$ is continuous.*

5. *For $\widetilde{X} \in \mathcal{A}$, the expectation $m(\widetilde{X})$ satisfies the inequality*

$$\rho\left(\widetilde{X}, m(\widetilde{X})\right) \leqslant \rho(\widetilde{X}, a) \quad \forall\, a \in R,$$

where $a_\alpha^- = a_\alpha^+ = a$ for a real number a (the extremal property).

The covariance of FRVs $\widetilde{X}$ and $\widetilde{Y}$ is defined by [20]

$$cov(\widetilde{X}, \widetilde{Y}) = \frac{1}{2} \int_0^1 \left(cov(X_\alpha^-, Y_\alpha^-) + cov(X_\alpha^+, Y_\alpha^+)\right)\, d\alpha; \tag{6}$$

the variance of an FRV $\widetilde{X}$, by the formula $D(\widetilde{X}) = cov(\widetilde{X}, \widetilde{X})$. In (6), the covariances of real random variables $X_\alpha^\pm$ and $Y_\alpha^\pm$ are defined in the standard way [21, Chapter I]:

$$cov(X_\alpha^\pm, Y_\alpha^\pm) = E(X_\alpha^\pm - E(X_\alpha^\pm))(Y_\alpha^\pm - E(Y_\alpha^\pm)).$$

Hereinafter, the symbol E stands for the expectation of a scalar random variable: $E\xi = \int_\Omega \xi(\omega)\, dP$ for a random variable $\xi(\omega)$.

Proposition 3 [20]. *The covariance* (6) *of an FRV possesses the following properties:*

1. $cov(\widetilde{X}, \widetilde{Y}) = cov(\widetilde{Y}, \widetilde{X})$, $\forall\, \widetilde{X}, \widetilde{Y} \in \mathcal{A}$ (*symmetry*).
2. $cov(\widetilde{X} + \widetilde{Y}, \widetilde{Z}) = cov(\widetilde{X}, \widetilde{Y}) + cov(\widetilde{Y}, \widetilde{Z})$, $\forall\, \widetilde{X}, \widetilde{Y}, \widetilde{Z} \in \mathcal{A}$ (*additivity*).
3. $cov(C_1\widetilde{X}, C_2\widetilde{Y}) = C_1C_2cov(\widetilde{X}, \widetilde{Y})$, $\forall\, \widetilde{X}, \widetilde{Y} \in \mathcal{A}$, $\forall\, C_1, C_2 \in R\colon C_1C_2 > 0$ (*positive homogeneity*).
4. $D(C\widetilde{X}) = C^2D(\widetilde{X})$, $\forall\, \widetilde{X} \in \mathcal{A}$, $\forall\, C \in R$.
5. $D(\widetilde{X} + \widetilde{Y}) = D(\widetilde{X}) + D(\widetilde{Y}) + 2cov(\widetilde{X}, \widetilde{Y})$, $\forall\, \widetilde{X}, \widetilde{Y} \in \mathcal{A}$.

3. CONTINUOUS RANDOM PROCESSES WITH FUZZY STATES

Let $[t_0, T]$ be an extended segment of the real axis. As in Section 2, (Ω, Σ, P) is a probability space and $\mathcal{A}$ is the set of FRVs with square summable indices on $\Omega \times [0, 1]$.

A continuous random process with fuzzy states or a fuzzy random process (FRP) is a mapping $\widetilde{X} : [t_0, T] \to \mathcal{A}$, i.e., a function $\widetilde{X}(\omega, t)$ with FRVs as its values $\forall\, t \in [t_0, T]$. The α-indices of an FRP $\widetilde{X}(\omega, t)$ will be denoted by $X_\alpha^\pm(\omega, t)$.

For the FRPs considered below, the functions $X_\alpha^\pm(\omega, t)$ are jointly square summable on $\Omega \times [0, 1] \times [t_0, T]$.

For each $t \in [t_0, T]$, the fuzzy expectation $M(\widetilde{X}(\omega, t))$ of an FRP $\widetilde{X}(\omega, t)$ is defined as the fuzzy expectation of the corresponding FRV, i.e., a fuzzy-valued function with the indices $\left[M(\widetilde{X}(\omega, t))\right]_\alpha^\pm = \int_\Omega X_\alpha^\pm(\omega, t)\, dP$.

The definition of the fuzzy expectation of an FRP and the properties of the fuzzy expectations of FRVs imply the following result.

Proposition 4. 1. *The fuzzy expectation of a nonrandom fuzzy-valued function $\tilde{z} : [t_0, T] \to J$ coincides with this function: $M(\tilde{z}(t)) = \tilde{z}(t)$.*

2. *A nonrandom scalar function $\varphi : [t_0, T] \to R$ can be factored outside the fuzzy expectation sign:*

$$M(\varphi(t)\widetilde{X}(t)) = \varphi(t)M(\widetilde{X}(t)),$$

where $\widetilde{X}(t)$ is a fuzzy random process.

3. *The fuzzy expectation of the sum of two FRPs equals the sum of their fuzzy expectations:*

$$M(\widetilde{X}(t)+\widetilde{Y}(t))=M(\widetilde{X}(t))+M(\widetilde{Y}(t)).$$

4. *For an FRP $\widetilde{X}(t)$, the fuzzy expectation $M(\widetilde{X}(t))$ satisfies the inequality*

$$\int_{t_0}^{T}\rho^2\left(\widetilde{X}(t),M(\widetilde{X}(t))\right)\,dt\leqslant\int_{t_0}^{T}\rho^2\left(\widetilde{X}(t),\widetilde{V}(t)\right)\,dt.$$

(*the extremal property*). *Here $\widetilde{V}:[t_0,T]\to J$ is an arbitrary nonrandom fuzzy-valued function with square summable indices on $[t_0,T]$ and the metric ρ is given by* (3).

In accordance with (5), the expectation of an FRP $\widetilde{X}(t)$ is defined as

$$m(\widetilde{X}(t))=\frac{1}{2}\int_0^1\left([M(\widetilde{X}(t))]_\alpha^-+[M(\widetilde{X}(t))]_\alpha^+\right)\,d\alpha.$$

Note that Proposition 2 holds for $m(\widetilde{X}(t))\ \forall\, t\in[t_0,T]$.

This paper involves the mean-square differentiability of a scalar random process [21, Chapter II]. A scalar random process $\xi(t)$ is said to be mean-square differentiable at a point $t\in R$ if there exists a random variable $\xi'(t)$ such that the expectation

$$E\left|\frac{\xi(t+h)-\xi(t)}{h}-\xi'(t)\right|^2\to 0\quad\text{as}\quad h\to 0.$$

Also, the derivative of a fuzzy-valued function will be used; for example, see [10, 11].

An FRP $\widetilde{X}(\omega,t)$ with α-intervals $[X_\alpha^-(\omega,t),X_\alpha^+(\omega,t)]$ is said to be differentiable at a point t (in the Seikkala sense) if its α-indices are mean-square differentiable with respect to t as scalar random processes and $\frac{\partial}{\partial t}X_\alpha^-(\omega,t)$ and $\frac{\partial}{\partial t}X_\alpha^+(\omega,t)$ are the lower and upper α-indices of some FRV called the derivative (cf. with [11] for a fuzzy-valued function). In this case, the time derivative of an FRP $\widetilde{X}(t)$ will be denoted by $\widetilde{X}'(t)=\frac{\partial}{\partial t}\widetilde{X}(\omega,t)$.

Theorem 1. *Let an FRP $\widetilde{X}(t)$ be differentiable for $t\in(t_0,T)$. Assume that $\forall\,\alpha\in[0,1]$ there exist summable functions $\varphi_\alpha^\pm(\omega)$ on Ω such that $\left|\frac{\partial}{\partial t}X_\alpha^\pm(\omega,t)\right|\leqslant\varphi_\alpha^\pm(\omega)\ \forall\, t\in(t_0,T)$, $\omega\in\Omega$. Then the fuzzy expectation of the derivative of this FRP equals the derivative of its fuzzy expectation:*

$$M\widetilde{X}'(t)=\left(M\widetilde{X}(t)\right)'.\tag{7}$$

Indeed, by the definition of derivative and fuzzy expectation, $\left[M(\widetilde{X}'(t))\right]_\alpha^\pm=\int_\Omega\frac{\partial}{\partial t}X_\alpha^\pm(\omega,t)\,dP$ $\forall\,\alpha\in[0,1]$.

Under the conditions of Theorem 1, the derivative on the right-hand side of this equality can be taken outside the sign of the integral; see the theorem on differentiation by a parameter under the sign of a Lebesgue integral. Then, using the interval test of equality for fuzzy numbers, we obtain (7).

The considerations below will employ the integral of a fuzzy-valued function; for example, see [12, 13].

The integral of an FRP $\widetilde{X}(\omega,t)$ between the limits of a segment $[t_0,T]$ is an FRV $\widetilde{Y}(\omega)=\int_{t_0}^T\widetilde{X}(\omega,t)\,dt$ with the α-indices

$$Y_\alpha^\pm(\omega)=\int_{t_0}^{T}X_\alpha^\pm(\omega,t)\,dt,$$

where $X_\alpha^\pm(t)$ denote the corresponding α-indices of the FRP $\widetilde{X}(t)$ and the integrals of the random processes $X_\alpha^\pm(t)$ are understood in the mean-square sense [21, Chapter II]. If such an integral exists, then the FRP is said to be integrable on $[t_0, T]$.

Theorem 2. *Let $\widetilde{X}(\omega, t)$ be an integrable FRP on $[t_0, T]$. Then*

$$M\left(\int_{t_0}^{T} \widetilde{X}(\omega, t)\, dt\right) = \int_{t_0}^{T} M\left(\widetilde{X}(\omega, t)\right)\, dt. \tag{8}$$

Indeed, by the definition of fuzzy expectation, the left-hand side of (8) has the α-indices $\left[M\left(\int_{t_0}^{T} \widetilde{X}(\omega, t)\, dt\right)\right]_\alpha^\pm = \int\limits_\Omega \left[\int_{t_0}^{T} \widetilde{X}(\omega, t)\, dt\right]_\alpha^\pm dP = \int\limits_\Omega \left(\int_{t_0}^{T} \widetilde{X}_\alpha^\pm(\omega, t) dt\right) dP$. Since the random variables $X_\alpha^\pm(\omega, t)$ are square summable on $\Omega \times [t_0, T]$, the order of integration in the last expression can be interchanged by Fubini's theorem. As a result, based on the definition of fuzzy expectation,

$$\left[M\left(\int_{t_0}^{T} \widetilde{X}(\omega, t)\, dt\right)\right]_\alpha^\pm = \int_{t_0}^{T} \left(\int_\Omega X_\alpha^\pm(\omega, t)\, dP\right) dt = \int_{t_0}^{T} \left[M(\widetilde{X}(\omega, t))\right]_\alpha^\pm dt = \left[\int_{t_0}^{T} M\left(\widetilde{X}(\omega, t)\right) dt\right]_\alpha^\pm.$$

Then formula (8) follows from the equality of the corresponding indices $\forall\, [0, 1]$.

Example 1. Let numerical random processes $\xi_i(\omega, t)$ $(i = 1, 2, 3;\ \omega \in \Omega,\ t \in [t_0, T])$ be square summable on $\Omega \times [t_0, T]$ and $\xi_1(\omega, t) < \xi_2(\omega, t) < \xi_3(\omega, t)$ for all $\omega \in \Omega$, $t \in [t_0, T]$.

Consider an FRP $\widetilde{X}(t)$ in which the fuzzy number $\widetilde{X}(\omega, t)$ for each $\omega \in \Omega$, $t \in [t_0, T]$, has the triangular form $(\xi_1(\omega, t), \xi_2(\omega, t), \xi_3(\omega, t))$. In other words, for any $\omega \in \Omega$, $t \in [t_0, T]$, the membership function $\widetilde{X}(\omega, t)$ is described by

$$\mu_{\omega,t}(x) = \begin{cases} \dfrac{x - \xi_1(\omega, t)}{\xi_2(\omega, t) - \xi_1(\omega, t)} & \text{if } x \in [\xi_1(\omega, t), \xi_2(\omega, t)]; \\ \dfrac{x - \xi_3(\omega, t)}{\xi_2(\omega, t) - \xi_3(\omega, t)} & \text{if } x \in [\xi_2(\omega, t), \xi_3(\omega, t)]; \\ 0 & \text{otherwise.} \end{cases}$$

In this case,

$$X_\alpha^-(\omega, t) = (1 - \alpha)\xi_1(\omega, t) + \alpha\xi_2(\omega, t), \quad X_\alpha^+(\omega, t) = (1 - \alpha)\xi_3(\omega, t) + \alpha\xi_2(\omega, t).$$

Then the fuzzy expectation $M(\widetilde{X}(t))$ is given by the formulas for the α-indices

$$[M(\widetilde{X})]_\alpha^-(t) = (1 - \alpha) \int\limits_\Omega \xi_1(\omega, t)\, dP + \alpha \int\limits_\Omega \xi_2(\omega, t)\, dP = (1 - \alpha)E\xi_1(t) + \alpha E\xi_2(t) \quad \forall\, \alpha \in [0, 1]$$

and

$$[M(\widetilde{X})]_\alpha^+(t) = (1 - \alpha) \int\limits_\Omega \xi_3(\omega, t)\, dP + \alpha \int\limits_\Omega \xi_2(\omega, t)\, dP = (1 - \alpha)E\xi_3(t) + \alpha E\xi_2(t) \quad \forall\, \alpha \in [0, 1],$$

where E denotes the expectation of a real random variable.

Example 2. Within Example 1, let the random processes $\xi_i(\omega, t)$, $i = 1, 2, 3$, be mean-square differentiable with respect to t. In addition, assume that the derivatives satisfy the relation $\frac{\partial}{\partial t}\xi_1(\omega, t) \leqslant \frac{\partial}{\partial t}\xi_2(\omega, t) \leqslant \frac{\partial}{\partial t}\xi_3(\omega, t)$ for all $\omega \in \Omega$, $t \in [t_0, T]$. Then the derivative $\widetilde{X}'(t)$ of the FRP $\widetilde{X}(t)$ has the triangular form $\left(\frac{\partial}{\partial t}\xi_1(\omega, t), \frac{\partial}{\partial t}\xi_2(\omega, t), \frac{\partial}{\partial t}\xi_3(\omega, t)\right)$.

In particular, let random variables $\xi_1(\omega) < \xi_2(\omega) < \xi_3(\omega)$ be given. Consider the triangular fuzzy process $(e^t\xi_1(\omega), e^t\xi_2(\omega), e^t\xi_3(\omega))$. Its derivative exists and coincides with the initial FRP.

Example 3. Within Example 1, let the random processes $\xi_i(\omega, t)$ be integrable in t on a segment $[t_0, T]$. Then the FRP $\int_{t_0}^{T} \widetilde{X}(t)dt$ has the triangular form $\left(\frac{\partial}{\partial t}\xi_1(\omega, t)dt, \frac{\partial}{\partial t}\xi_2(\omega, t)dt, \frac{\partial}{\partial t}\xi_3(\omega, t)dt\right)$.

In what follows, we present the covariance function of an FRP and its properties. The covariance function of an FRP $\widetilde{X}(t)$ is the value

$$K_{\widetilde{X}}(t, s) = cov(\widetilde{X}(t), \widetilde{X}(s)) = \frac{1}{2}\int_0^1 \left(K_{X_\alpha^-}(t, s) + K_{X_\alpha^+}(t, s)\right) d\alpha. \tag{9}$$

Here, $K_{X_\alpha^-}(t, s)$ and $K_{X_\alpha^+}(t, s)$ are the covariance functions of random processes $X_\alpha^-(\omega, t)$ and $X_\alpha^+(\omega, t)$, respectively:

$$K_{X_\alpha^\pm}(t, s) = E\left[(X_\alpha^\pm(\omega, t) - EX_\alpha^\pm(\omega, t))(X_\alpha^\pm(\omega, s) - EX_\alpha^\pm(\omega, s))\right]. \tag{10}$$

The variance of an FRP $\widetilde{X}(t)$ is given by $D_{\widetilde{X}}(t) = K_{\widetilde{X}}(t, t)$.

Note that (9) is a fuzzy modification of the conventional covariance function of scalar random processes; for example, see [21, Chapter II].

The expressions (9) and (10) and the properties of the covariance of an FRV (Proposition 3) lead to the following result.

Proposition 5. *The covariance function of an FRP possesses the following properties:*

1. *For a continuous FRP* $\widetilde{X}(t)$, $K_{\widetilde{X}}(t_1, t_2) = K_{\widetilde{X}}(t_2, t_1)\ \forall\, t_1, t_2 \in [t_0, T]$ *(symmetry).*

2. *If* $\widetilde{X}(t)$ *is a continuous FRP and* $\varphi(t)$ *is a nonrandom numerical function, then the covariance function* $K_{\widetilde{Y}}(t_1, t_2)$ *of the FRP* $\widetilde{Y}(t) = \varphi(t)\widetilde{X}(t)$ *has the form* $K_{\widetilde{Y}}(t_1, t_2) = \varphi(t_1)\varphi(t_2)K_{\widetilde{X}}(t_1, t_2)$.

3. *If* $\widetilde{Y}(t) = \widetilde{X}(t) + \varphi(t)$, *then* $K_{\widetilde{Y}}(t_1, t_2) = K_{\widetilde{X}}(t_1, t_2)$.

4. $|K_{\widetilde{X}}(t_1, t_2)| \leqslant \sqrt{D_{\widetilde{X}}(t_1)D_{\widetilde{X}}(t_2)}$.

The next result characterizes the connection between the correlation functions of a differentiable FRP and its derivative.

Theorem 3. *Let the second derivatives* $\frac{\partial^2 K_{X_\alpha^-}(t,s)}{\partial t\partial s}$ *and* $\frac{\partial^2 K_{X_\alpha^+}(t,s)}{\partial t\partial s}$ *of the covariance functions* (10) *of an FRP* $\widetilde{X}(t)$ *be defined and jointly continuous in the variables* t, s, α. *Then the covariance function* $K'_{\widetilde{X}}(t, s)$ *of the derivative* $\widetilde{X}'(t)$ *of the FRP* $\widetilde{X}(t)$ *is given by*

$$K'_{\widetilde{X}}(t, s) = \frac{\partial^2 K_{\widetilde{X}}(t, s)}{\partial t\partial s}. \tag{11}$$

Proof. By definition (9),

$$K_{\widetilde{X}'}(t, s) = \frac{1}{2}\int_0^1 \left(K_{(\widetilde{X}_\alpha^-)'}(t, s) + K_{(\widetilde{X}_\alpha^+)'}(t, s)\right) d\alpha.$$

According to the well-known property of (scalar) random processes,

$$K_{(\widetilde{X}_\alpha^-)'}(t, s) = \frac{\partial^2 K_{X_\alpha^-}(t, s)}{\partial t\partial s}, \quad K_{(\widetilde{X}_\alpha^+)'}(t, s) = \frac{\partial^2 K_{X_\alpha^+}(t, s)}{\partial t\partial s}.$$

Then $K_{\widetilde{X}'}(t, s) = \frac{1}{2}\int_0^1 \frac{\partial^2}{\partial t\partial s}\left(K_{X_\alpha^-}(t, s) + K_{X_\alpha^+}(t, s)\right) d\alpha$. Taking the second mixed variable outside the integral sign yields formula (11).

Note that the last operation is valid due to the joint continuity of $\frac{\partial^2 K_{X_\alpha^-}(t,s)}{\partial t \partial s}$ and $\frac{\partial^2 K_{X_\alpha^+}(t,s)}{\partial t \partial s}$ in the variables t, s, α.

Consider the integral $\widetilde{Y}(t) = \int_{t_0}^t \widetilde{X}(\omega, s)\, ds$ of an FRP $\widetilde{X}(t)$ with a variable upper limit.

Theorem 4. *Let the covariance functions $K_{X_\alpha^\pm}(t,s)$ of the α-indices $X_\alpha^\pm(t)$ of an FRP $\widetilde{X}(t)$ be jointly summable in the variables t, s, α. Then the covariance function of the integral $\widetilde{Y}(t)$ is*

$$K_{\widetilde{Y}}(t,s) = \int_{t_0}^{t}\int_{t_0}^{s} K_{\widetilde{X}}(\tau_1,\tau_2)\, d\tau_1 d\tau_2. \tag{12}$$

Proof. By definition (9), $K_{\widetilde{Y}}(t,s) = \frac{1}{2}\int_0^1 \left(K_{Y_\alpha^-}(t,s) + K_{Y_\alpha^+}(t,s)\right) d\alpha$. Due to the definition of a fuzzy integral, $Y_\alpha^\pm(t) = \int_{t_0}^t X_\alpha^\pm(\tau)\, d\tau$. Using the well-known property of the integral of a scalar random process, we obtain

$$K_{Y_\alpha^-}(t,s) = \int_{t_0}^{t}\int_{t_0}^{s} K_{X_\alpha^-}(\tau_1,\tau_2)\, d\tau_1 d\tau_2, \quad K_{Y_\alpha^+}(t,s) = \int_{t_0}^{t}\int_{t_0}^{s} K_{X_\alpha^+}(\tau_1,\tau_2)\, d\tau_1 d\tau_2.$$

Hence, $K_{\widetilde{Y}}(t,s) = \frac{1}{2}\int_0^1 \left(\int_{t_0}^t \int_{t_0}^s (K_{X_\alpha^-}(\tau_1,\tau_2) + K_{X_\alpha^+}(\tau_1,\tau_2)\right) d\tau_1 d\tau_2)\, d\alpha$. The desired result (12) is established by interchanging the order of integration on the right-hand side (based on Fubini's theorem) and employing (9).

Note that the last operation is valid due to the joint summability of $K_{X_\alpha^-}(\tau_1,\tau_2)$ and $K_{X_\alpha^+}(\tau_1,\tau_2)$ in the variables τ_1, τ_2, α.

4. TRANSFORMATION OF A FUZZY RANDOM SIGNAL BY A LINEAR DYNAMIC SYSTEM

Consider some device A (Fig. 1) with continuous random signals $y(t)$ and $z(t)$ at its input and output, respectively.

Device A is called a linear dynamic system if the relationship between the input and output signals is described by an nth order differential equation with constant coefficients.

The literature (e.g., see [6, Chapter 7]) considers the problem of establishing connections between the numerical characteristics (expectations and covariance functions) of the input and output random signals. Assuming the stationarity of the random signals, this problem is solved using the frequency response of the system, the direct and inverse Fourier transforms, and the Wiener–Khinchin theorem. A random process is called stationary (in the broad sense) if its expectation does not depend on time and the covariance function depends only on the difference of the arguments.

Consider a similar problem when the input and output signals are continuous random signals with fuzzy states (fuzzy random signals, FRSs). In this case, stationarity in any sense is not supposed. In contrast to well-known techniques, the Green function method is used below. The proposed approach will be illustrated on examples.

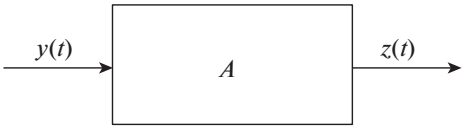


Fig. 1.

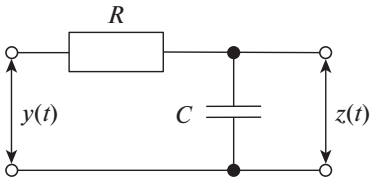


Fig. 2.

Example 4. Consider an RC circuit (Fig. 2) described by the differential equation

$$z'(t) + \beta z(t) = \beta y(t), \quad \beta = \frac{1}{RC} > 0,$$

where R and C denote the resistance and capacity, respectively.

Let a continuous FRS $\widetilde{Y}(t)$ be supplied to the input. We will determine the connection between the fuzzy expectations (as well as expectations) of the output and input signals of this system. By convention, the output FRS $\widetilde{X}(t)$ satisfies the fuzzy random differential equation

$$\widetilde{X}'(t) + \beta \widetilde{X}(t) = \beta \widetilde{Y}(t). \tag{13}$$

Note that fuzzy differential equations were studied, e.g., in [22–24].

In view of the additivity and homogeneity of the expectation of an FRP (Proposition 4) and Theorem 1, taking the fuzzy expectation of both sides of (13) gives

$$(M\widetilde{X})'(t) + \beta M\widetilde{X}(t) = \beta M\widetilde{Y}(t).$$

By the definition of equality for fuzzy numbers and operations between them, the definition of derivative for α-indices, and $\beta > 0$, this equation is equivalent to the following set of relations for the indices:

$$\frac{\partial}{\partial t}(M\widetilde{X})^{\pm}_{\alpha}(t) + \beta(M\widetilde{X})^{\pm}_{\alpha}(t) = \beta(M\widetilde{Y})^{\pm}_{\alpha}(t) \quad \forall\, \alpha \in [0,1]. \tag{14}$$

In addition, assume that the functions $(M\widetilde{Y})^{\pm}_{\alpha}(t)$ are bounded in t on the entire real axis. According to [25, Chapter II], there exists a unique and asymptotically Lyapunov stable solution of Eq. (14) that is bounded on the entire real axis. This solution has the form

$$(M\widetilde{X})^{\pm}_{\alpha}(t) = \beta \int\limits_{-\infty}^{\infty} G_1(t-s)(M\widetilde{Y})^{\pm}_{\alpha}(s)\, ds = \beta \int\limits_{-\infty}^{t} e^{-\beta(t-s)}(M\widetilde{Y})^{\pm}_{\alpha}(s)\, ds \quad \forall\, \alpha \in [0,1], \tag{15}$$

where $G_1 = \begin{cases} e^{-\beta t} & \text{for } \ t > 0; \\ 0 & \text{for } \ t < 0 \end{cases}$ is the Green function of the problem on bounded solutions of the (scalar) Eq. (14).

Formula (15) characterizes the connection between the fuzzy expectations of the input and output FRSs of the system described by Eq. (13).

Due to (15) and (5), the expectations of the input and output FRSs of system (13) have the connection

$$m\left(\widetilde{X}(t)\right) = \beta \int\limits_{-\infty}^{t} e^{-\beta(t-s)} m\left(\widetilde{Y}(s)\right)\, ds.$$

By the definition of equality for fuzzy numbers and the definition of derivative for FRPs, from (13) it follows that the α-indices of $\widetilde{X}(t)$ satisfy

$$(X_\alpha^\pm)'(t) + \beta X_\alpha^\pm(t) = \beta Y_\alpha^\pm(t) \quad \forall\, \alpha \in [0,1]. \tag{16}$$

The relation (16) can be treated as an equation in the Hilbert space $\mathcal{H}$ of all random variables with a finite second moment. In addition, assume that the functions $Y_\alpha^\pm(t)$ are bounded in t in the space $\mathcal{H}$ on the entire real axis. According to [25, Chapter II], there exists a unique and asymptotically Lyapunov stable solution of Eq. (16) that is bounded on the entire real axis. This solution has the form

$$X_\alpha^\pm(t) = \beta \int\limits_{-\infty}^{t} e^{-\beta(t-s)} Y_\alpha^\pm(s) ds \quad \forall\, \alpha \in [0,1]. \tag{17}$$

Note that $X_\alpha^\pm(t)$ in (17) determine the α-indices of the fuzzy number $\widetilde{X}(t)$. In particular, $X_\alpha^+(t)$ and $X_\alpha^-(t)$ are monotonically nonincreasing and nondecreasing in α, respectively, due to the monotonicity of the integral and the corresponding properties of the α-indices $Y_\alpha^+(s)$ and $Y_\alpha^-(s)$, respectively; for details, see [1, Chapter 5].

Now, we calculate the covariance function of the output FRP $\widetilde{X}$ of the system described by Eq. (13). (Recall that this process is bounded on the entire real axis.) By (10) and (17), the covariance function $K_{X_\alpha^\pm}(t,s)$ has the form

$$\begin{aligned} K_{X_\alpha^\pm}(t,s) = \beta^2 E \Bigg[\Bigg(\int\limits_{-\infty}^{t} e^{-\beta(t-\tau_1)} Y_\alpha^\pm(\omega,\tau_1)\, d\tau_1 - E \int\limits_{-\infty}^{t} e^{-\beta(t-\tau_1)} Y_\alpha^\pm(\omega,\tau_1)\, d\tau_1 \Bigg) \\ \times \Bigg(\int\limits_{-\infty}^{s} e^{-\beta(s-\tau_2)} Y_\alpha^\pm(\omega,\tau_2)\, d\tau_2 - E \int\limits_{-\infty}^{s} e^{-\beta(s-\tau_2)} Y_\alpha^\pm(\omega,\tau_2)\, d\tau_2 \Bigg) \Bigg]. \end{aligned}$$

Interchanging the expectation (E) and integration operations in the inner parentheses yields

$$\begin{aligned} K_{X_\alpha^\pm}(t,s) &= \beta^2 E \Bigg[\Bigg(\int\limits_{-\infty}^{t} e^{-\beta(t-\tau_1)} \left(Y_\alpha^\pm(\omega,\tau_1) - E(Y_\alpha^\pm(\omega,\tau_1)) \right) d\tau_1 \Bigg) \\ &\quad \times \Bigg(\int\limits_{-\infty}^{s} e^{-\beta(s-\tau_2)} \left(Y_\alpha^\pm(\omega,\tau_2) - E(Y_\alpha^\pm(\omega,\tau_2)) \right) d\tau_2 \Bigg) \Bigg] \\ &= \beta^2 \int\limits_{-\infty}^{t} \int\limits_{-\infty}^{s} e^{-\beta(t-\tau_1)} e^{-\beta(s-\tau_2)} K_{Y_\alpha^\pm}(\tau_1,\tau_2)\, d\tau_1 d\tau_2. \end{aligned} \tag{18}$$

Due to (18) and (9), the covariance functions of the output and input FRSs of system (13) are related by

$$\begin{aligned} K_{\widetilde{X}}(t,s) &= \frac{\beta^2}{2} \int\limits_{-\infty}^{t} \int\limits_{-\infty}^{s} e^{-\beta(t-\tau_1)} e^{-\beta(s-\tau_2)} \int\limits_0^1 \left(K_{Y_\alpha^-}(\tau_1,\tau_2) + K_{Y_\alpha^+}(\tau_1,\tau_2) \right) d\alpha d\tau_1 d\tau_2 \\ &= \beta^2 \int\limits_{-\infty}^{t} \int\limits_{-\infty}^{s} e^{-\beta(t-\tau_1)} e^{-\beta(s-\tau_2)} K_{\widetilde{Y}}(\tau_1,\tau_2) d\tau_1 d\tau_2. \end{aligned}$$

The convergence of these improper integrals and the validity of the considerations above are ensured by the exponential estimates of the corresponding integrands.

Example 5. An FRS $\widetilde{Y}(t)$ is supplied to the input of a linear dynamic system described by the differential equation

$$z''(t) + a_1 z'(t) + a_2 z(t) = y(t). \tag{19}$$

It is required to characterize its output FRS $\widetilde{X}(t)$.

From (19) we have the fuzzy differential equation

$$\widetilde{X}''(t) + a_1 \widetilde{X}'(t) + a_2 \widetilde{X}(t) = \widetilde{Y}(t). \tag{20}$$

By analogy with Example 4, the expectation $M(\widetilde{X})(t)$ of the output signal satisfies the differential equation

$$(M\widetilde{X})''(t) + a_1 (M\widetilde{X})'(t) + a_2 M(\widetilde{X})(t) = M(\widetilde{Y})(t).$$

Let the coefficients be $a_1, a_2 > 0$. Then, for the α-indices of the expectation,

$$\left[(M\widetilde{X})^{\pm}_{\alpha}\right]''(t) + a_1 \left[(M\widetilde{X})^{\pm}_{\alpha}\right]'(t) + a_2 (M\widetilde{X})^{\pm}_{\alpha}(t) = M(\widetilde{Y})^{\pm}_{\alpha}(t) \quad \forall\, \alpha \in [0,1]. \tag{21}$$

Let the functions $(M\tilde{Y})^{\pm}_{\alpha}(t)$ be bounded in t on the entire real axis. In addition, suppose that the roots of the characteristic equation $\lambda^2 + a_1\lambda + a_2 = 0$ corresponding to (19) are real, negative, and $\lambda_1 < \lambda_2 < 0$. Then there exists a unique and asymptotically Lyapunov stable solution of Eq. (21) that is bounded on the entire real axis. This solution has the form

$$(M\widetilde{X})^{\pm}_{\alpha}(t) = \int\limits_{-\infty}^{t} G_2(t-s)(M\widetilde{Y})^{\pm}_{\alpha}(s)\, ds \quad \forall\, \alpha \in [0,1], \tag{22}$$

where G_2 is the Green function of the problem on bounded solutions of the scalar Eq. (19); for example, see [26, Chapter 2, § 8]. Under the assumptions accepted above, it has the form

$$G_2(t) = \begin{cases} (e^{\lambda_2 t} - e^{\lambda_1 t})(\lambda_2 - \lambda_1)^{-1} & \text{for} \quad t > 0; \\ 0 & \text{for} \quad t < 0. \end{cases}$$

Formula (22) characterizes the connection between the fuzzy expectations of the input and output FRSs of the system described by Eq. (20).

Assume that the α-indices $Y^{\pm}_{\alpha}(t)$ are bounded in the space $\mathcal{H}$ on the entire real axis. Following the considerations of Example 4, we easily arrive at

$$\begin{aligned} K_{\widetilde{X}}(t,s) &= \frac{1}{2} \int\limits_{-\infty}^{t} \int\limits_{-\infty}^{s} G_2(t-\tau_1) G_2(s-\tau_2) \times \int\limits_{0}^{1} \left(K_{\widetilde{Y}^{-}_{\alpha}}(\tau_1,\tau_2) + K_{\widetilde{Y}^{+}_{\alpha}}(\tau_1,\tau_2) \right) d\alpha d\tau_1 d\tau_2 \\ &= \int\limits_{-\infty}^{t} \int\limits_{-\infty}^{s} G_2(t-\tau_1) G_2(s-\tau_2) K_{\widetilde{Y}}(\tau_1,\tau_2)\, d\tau_1 d\tau_2. \end{aligned} \tag{23}$$

Formula (23) characterizes the connection between the correlation functions of the input and output FRSs of the system described by Eq. (20).

Note that within Examples 4 and 5, the output processes are asymptotically Lyapunov stable in the sense that their α-indices are asymptotically Lyapunov stable; for example, see [25, Chapter II]. Only such processes are physically realizable. Other approaches to the stability of solutions of fuzzy differential equations are also considered in the literature; for example, see [3, Chapter 8].

5. CONCLUSIONS

The starting point for this study is the paper [20], where the covariances of FRVs were discussed and the covariance function of an FRP was introduced. Such functions have been investigated above.

The properties of the fuzzy expectations of FRPs (Proposition 4) and covariance functions (Proposition 5) naturally follow from the corresponding properties of the fuzzy expectations and covariances of fuzzy random variables (Propositions 1–3).

The essential content and scientific novelty of this paper are Theorems 1–4, which present the characteristics of differentiable and integrable FRPs. Their proof involves the definitions and properties of differentiability and integrability of fuzzy-valued functions. Theorems 1–4 generalize the well-known results for standard continuous random processes; for example, see [6]. Also, note the extremal property of the fuzzy expectations of FRPs (Proposition 4), which seems new to the author.

Examples 1–3 are illustrative. Examples 4–5 show the possible use of this theory in applications, particularly the problem of transforming an FRS by a linear dynamic system. The results of Section 4 can be extended to the case of periodic and almost periodic FRSs, including the problem of spectral decompositions of FRSs.

REFERENCES

1. Averkin, A.N., *Nechetkie mnozhestva v modelyakh upravleniya i iskusstvennogo intellekta* (Fuzzy Sets in Models of Control and Artificial Intelligence), Moscow: Nauka, 1986.
2. Buckley, J.J., Eslami, E., and Feuring, T., *Fuzzy Mathematics in Economics and Engineering*, Heidelberg–New York: Physica-Verl., 2002.
3. Pegat, A., *Nechetkoe modelirovanie i upravlenie* (Fuzzy Modeling and Control), Moscow: BINOM. Laboratoriya Znanii, 2015.
4. Mochalov, I.A., Khrisat, M.S., and Shihab Eddin, M.Ya., Fuzzy Differential Equationx s in Control. Part II, *Information Technologies*, 2015, vol. 21, no. 4, pp. 243–250.
5. Demenkov, N.P., Mikrin, E.A., and Mochalov, I.A., Fuzzy Optimal Control of Linear Systems. Part 1. Positional Control, *Information Technologies*, 2019, vol. 25, no. 5, pp. 259–270.
6. Venttsel', E.S. and Ovcharov, L.A., *Teoriya sluchainykh protsessov i ikh inzhenernye prilozheniya* (Theory of Random Processes and Their Engineering Applications), Moscow: Knorus, 2016.
7. Puri, M.L. and Ralesku, D.A., Fuzzy Random Variables, *Journal of Mathematical Analysis and Applications*, 1986, vol. 114, pp. 409–422.
8. Nguyen, H.T. and Wu, B., *Fundamentals of Statistics with Fuzzy Data*, Berlin–Heidelberg: Springer-Verlag, 2006.
9. Shvedov, A.A., Estimating the Means and the Covariances of Fuzzy Random Variables, *Applied Econometrics*, 2016, vol. 42, pp. 121–138.
10. Puri, M.L. and Ralescu, D.A., Differential of Fuzzy Functions, *J. Math. Anal. Appl.*, 1983, vol. 91, pp. 552–558.
11. Seikkala, S., On the Fuzzy Initial Value Problem, *Fuzzy Sets Syst.*, 1987, vol. 24, no. 3, pp. 319–330.
12. Aumann, R.J., Integrals of Set-Valued Functions, *J. Math. Anal. Appl.*, 1987, no. 3, pp. 1–12.
13. Hukuhara, M., Integration des applications mesurables dont la valeur est un compact convexe, *Func. Ekvacioj.*, 1967, no. 11, pp. 205–223.
14. Demenkov, N.P., Mikrin, E.A., and Mochalov, I.A., Markov and Semi-Markov Processes with Fuzzy States. Part 1. Markov Processes, *Information Technologies*, 2020, vol. 26, no. 6, pp. 323–334.

15. Demenkov, N.P., Mikrin, E.A., and Mochalov, I.A., Markov and Semi-Markov Processes with Fuzzy States. Part 2. Semi-Markov Processes, *Information Technologies*, 2020, vol. 26, no. 7, pp. 387–393.

16. Diamond, P. and Kloeden, P., Metric Spaces of Fuzzy Sets, *Fuzzy Sets Syst.*, 1990, vol. 35, no. 2, pp. 241–249.

17. Khatskevich, V.L., On Some Properties of Fuzzy Expectations and Nonlinear Fuzzy Expectations of Fuzzy-Random Variables, *Russian Math. (Iz. VUZ)*, 2022, vol. 66, no. 11, pp. 86–96.

18. Khatskevich, V.L., Extremal Properties of Means of Fuzzy Random Variables, in *Itogi nauki i tekhniki. Sovremmennaya matematika i prilozheniya. Tematicheskie obzory* (Results of Science and Engineering. Contemporary Mathematics and Applications. Thematic Surveys), vol. 204, Moscow: VINITI, Moscow, 2022, pp. 160–169.

19. Dubois, D. and Prade, H., The Mean Value of Fuzzy Number, *Fuzzy Sets and Syst.*, 1987, vol. 24, no. 3, pp. 279–300.

20. Feng, Y., Hu, L., and Shu, H., The Variance and Covariance of Fuzzy Random Variables, *Fuzzy Syst.*, 2001, vol. 120, no. 3, pp. 487–497.

21. Rozanov, Yu.A., *Teoriya veroyatnostei, sluchainye protsessy i matematicheskaya statistika* (Probability Theory, Random Processes, and Mathematical Statistics), Moscow: Nauka, 1985.

22. Kaleva, O., Fuzzy Differential Equations, *Fuzzy Sets and Syst.*, 1987, vol. 24, no. 3, pp. 301–317.

23. Park, J.Y. and Han, H., Existence and Uniqueness Theorem for a Solution of Fuzzy Differential Equations, *Int. J. Math. Mathem. Sci.*, 1999, no. 22(2), pp. 271–280.

24. Ahmad, L., Farooq, M., and Abdullah, S., Solving nth Order Fuzzy Differential Equation by Fuzzy Laplace Transform, *Ind. J. Pure Appl. Math.*, 2014, no. 2, pp. 1–20.

25. Daletskii, Yu.L. and Krein, M.G., *Ustoichivost' reshenii differentsial'nykh uravnenii v banakhovom prostranstve* (Stability of Solutions of Differential Equations in a Banach Space), Moscow: Nauka, 1970.

26. Krasnosel'skii, M.A., Burd, V.Sh., and Kolesov, Yu.S., *Nelineinye pochti periodicheskie kolebaniya* (Nonlinear Almost Periodic Oscillations), Moscow: Nauka, 1970.

This paper was recommended for publication by A.I. Kibzun, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 791–809.

CONTROL IN TECHNICAL SYSTEMS

The Guaranteeing Estimation Method to Calibrate a Gyro Unit

P. A. Akimov[*,a] **and A. I. Matasov**[*,b]

[*] *Moscow State University, Faculty of Mechanics and Mathematics, Laboratory of Control and Navigation, Moscow, Russia*
e-mail: [a] *akmpavel@rambler.ru,* [b] *alexander.matasov@gmail.com*

Received November 7, 2022
Revised March 14, 2023
Accepted March 30, 2023

Abstract—This paper is devoted to the guaranteeing estimation method with application to the calibration problem of a gyro unit. Mathematical models are constructed to describe the kinematics of the gyro unit on a test bench. The applicability limits and errors of the models are investigated. A numerical solution procedure is developed for guaranteeing estimation problems based on their reduction to l_1-approximation problems.

Keywords: the guaranteeing approach to estimation, convex optimization, gyro unit, calibration, Python

DOI: 10.25728/arcRAS.2023.10.39.001

1. INTRODUCTION

This paper proposes a calibration procedure for a unit of angular rate sensors (ARSs), i.e., gyros, based on the guaranteeing estimation method [1–5]. The purpose of calibration is to estimate parametric errors (biases, scale factors, and misalignment angles) in a unit consisting of three ARSs. These parameters are determined through a series of measurements on a bench with high-accuracy control of the angular rate and orientation of the unit. During calibration, rotation modes are selected and the resulting signals are processed.

Many traditional calibration methods involve dynamic models for estimating unknown parameters based on measurements of inertial navigation system (INS) sensors (accelerometers and gyros); for details, see [6–8]. Nevertheless [9], in some cases, it is preferable to determine ARS errors without involving information from accelerometers. Such a situation arises, first, when the accuracy of accelerometers is too low to use their signals in bench tests and, second, when calibrating laser gyros on vibration suspensions. The Kalman filter and the least squares method [10] are the main tool for estimation. The calibration problem includes a large number of unknown parameters (sensor errors, bench errors) that nonlinearly affect the measurement results. Therefore, two questions are essential here as follows. How can one construct a mathematical model of bench tests to consider all these factors? How can one minimize the impact of errors on the estimation result? A possible answer to these questions is presented in this paper, which continues the earlier research on the application of the guaranteeing approach in inertial navigation; for example, see [11], where the calibration of an accelerometer unit was considered. However, in contrast to the cited paper, this method is applied below to a different class of systems with a large number of unknown parameters and nonlinear effects. This class requires constructing other models and leads to other, structurally more complex estimation problems.

The guaranteeing estimation method allows estimating unknown parameters under the "worst-case" realizations of measurement errors with minimal estimation accuracy. In this case, typical

maximum values of the bench errors are used instead of probabilistic hypotheses. The resulting solution sets the directions of rotation of the gyro unit on the bench, i.e., explicitly describes the optimal calibration modes.

This paper is organized as follows. In Section 1, the mathematical model of bench tests of the gyro unit is constructed; the transition from the basic kinematic equations to several linear models with signal averaging is performed; the applicability limits of these models are investigated. Sections 2 and 3 formulate the guaranteeing estimation problems and the corresponding auxiliary variational l_1-approximation problems that can be solved numerically. In Section 4, we determine optimal experiment plans, apply the estimation algorithms to model problems, and analyzing the accuracy of the resulting solutions.

2. MATHEMATICAL MODEL

2.1. Basic Assumptions and Kinematic Relations for a Gyro Unit

Consider a mathematical model describing the basic kinematic relations of a gyro unit on a test bench during calibration [12]. Let an experiment be conducted at the point M rigidly coupled to the Earth on its surface. We introduce the following notations:

$Mz = Mz_1z_2z_3$ is the instrumental frame rigidly coupled to the gyro unit;

$Mx = Mx_1x_2x_3$ is the frame rigidly coupled to the bench base fixed relative to the Earth;

$D(t)$ is the orthogonal orientation matrix of Mz relative to Mx. By definition of an orientation matrix, for any vector l, its coordinates in the reference frames Mz, Mx have the relation $l_z = D(t)\, l_x$ and the rows of the orientation matrix consist of the coordinates of the basis vectors of Mz in the frame Mx;

$\Omega(t)$ is the angular rate vector of the gyro unit relative to the bench;

$\omega(t)$ is the absolute angular rate vector of the gyro unit;

u_x are the coordinates of the angular rate of the Earth in the frame Mx and u is the absolute angular rate of the Earth.

The absolute angular rate of the gyro unit is described in projections onto Mx by the relation

$$\omega_x(t) = \Omega_x(t) + u_x.$$

In projections onto the axis of the frame Mz, this equality takes the form

$$\omega_z(t) = D(t)\,(\Omega_x(t) + u_x). \tag{1}$$

At the initial time instant, the orientation matrix D is known with some accuracy. We denote this estimate $D(0)$ by D_{init}:

$$D(0) = D_{\text{init}}(I_3 + \hat{\beta}), \quad \beta = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}}, \quad \hat{\beta} = \begin{pmatrix} 0 & \beta_3 & -\beta_2 \\ -\beta_3 & 0 & \beta_1 \\ \beta_2 & -\beta_1 & 0 \end{pmatrix}, \tag{2}$$
$$|\beta_i| \leqslant \beta_{\max}, \quad i = 1, 2, 3.$$

The initial alignment errors of the gyro unit, i.e., the small rotation angles β_i, are unknown but their absolute values are bounded by $\beta_{\max}$. Throughout this paper, I_n stands for an identity matrix of dimensions $n \times n$ and $\hat{\beta}$ denotes the skew-symmetric matrix constructed from the vector β according to the above rule.

For the gyro unit, the output signals (measurements) are the readings of each ARS, i.e., the components of the vector $\omega_z(t)$. Let us introduce the measurement model

$$\zeta(t) = \omega_z(t) + \Gamma\omega_z(t) + \nu_0 + \delta\nu(t). \tag{3}$$

Here: $\zeta(t) \in \mathbf{R}^3$ are measurement values; $\Gamma \in \mathbf{R}^{3\times 3}$ is an unknown matrix describing the scale factor errors and the orientation errors of the sensitivity axes; $\nu_0 \in \mathbf{R}^3$ are unknown zero biases in the sensor readings; finally, $\delta\nu(t) \in \mathbf{R}^3$ are nonparametric measurement errors (fluctuations). Without loss of generality, the matrix Γ is supposed symmetric [5].

The calibration problem consists in determining the values of Γ and ν_0 from the set of available measurements $\zeta(t)$. Note that the angular rate $\Omega_x(t)$ is controlled on the bench, and the optimal strategy of the unit's motion on the bench is one purpose of the mathematical calibration problem. Let the angular rate $\Omega_x(t)$ be written as

$$\Omega_x(t) = s'(t)w,$$

where $s'(t) \in \mathbf{R}$ denotes the angular rate and $w \in \mathbf{R}^3$ is the unit direction vector of the angular rate of Mz in projections onto Mx. The considerations below concern a special case of motion of the gyro unit on the bench that consists of several same-type experiments. Within each of them, the unit is rotated about a fixed axis with a given angular rate; then the unit is placed in a new position, a new direction of its rotation is set, and the experiment is repeated. The direction of the rotation axis and the angular rate are known with some errors. In other words, the function $s(t) \geqslant 0$ and vector y ($\|y\|_2 = 1$) are given in the expressions

$$s'(t) = s(t) + \varepsilon(t), \quad w = (I_3 + \hat{\alpha})y, \quad \Omega_x(t) = (s(t) + \varepsilon(t))(I_3 + \hat{\alpha})y, \tag{4}$$

which relate them to their true counterparts. The small rotation angles $\alpha \in \mathbf{R}^3$ and the corresponding skew-symmetric matrix $\hat{\alpha}$ determine the unknown errors in the rotation vector of the gyro unit whereas the scalar function $\varepsilon(t)$ determines the error in the angular rate value. As in the case of the angles β, the maximum possible values for α are known: $|\alpha_i| \leqslant \alpha_{\max}$, $i = 1, 2, 3$.

At each test stage, the errors α and β are constant but nonidentical in different experiments: by assumption, the programmed (target) angular rate and the unit orientation are set independently in each experiment. Thus, several series of measurements $\zeta(t)$ are formed that correspond to different rotation modes and different error realizations.

2.2. Linearization of the Equations and Signal Averaging

Substituting (1) and (4) into (3) yields the measurements

$$\zeta(t) = (I_3 + \Gamma)\, D(t)\ \left(s'(t)w + u_x\right) + \nu_0 + \delta\nu(t). \tag{5}$$

In addition to the signal $\zeta(t)$, two angular rate components, $D(t)u_x$ and $s'(t)D(t)w$, as well as the uncertain errors $\delta\nu(t)$ and $\varepsilon(t)$ depend on time. This section considers a mathematical model corresponding to the rotation mode of the gyro unit with a directionally constant angular rate on a time interval T. The calibration procedure will consist of a sequence of such rotation modes with different directions.

Next, we construct a "time-averaged" analog of the measurement equation by considering the averaging effect when Mz rotates relative to Mx. Averaging means calculating arithmetic means from a series of measurements on a long time interval T (several tens of minutes), during which the system makes multiple complete revolutions.

Recall that $D(t)$ is the transition matrix from Mx to Mz, and Mz is rigidly coupled to the gyro unit rotating relative to Mx with the angular rate Ω_x. The transition from Mx to Mz consists of three stages as follows.

1. Transition from Mx to Mx_{fix}, the stationary frame relative to the bench base, whose unit basis vector $e_{\text{fix}3}$ coincides with w in direction. We denote by $D_{\text{fix}} = (d_{\text{fix}1}; d_{\text{fix}2}; d_{\text{fix}3})$ the corresponding

transition matrix; its third row is $d_{\text{fix}3} = w^{\mathrm{T}}$, and the first and second rows are orthogonal to it and to each other and can be chosen in any suitable way.

2. Rotation about the axis $Mx_{\text{fix}3} = Mz_{\text{cir}3}$ with the angular rate $s'(t)$, translating Mx_{fix} into the frame Mz_{cir} rigidly coupled to the gyro unit. We denote by $\psi(t)$ the time-dependent rotation angle in the plane $Mx_{\text{fix}1}x_{\text{fix}2}$. Therefore, $\frac{d\psi(t)}{dt} = s'(t)$, and the transition matrix from Mx_{fix} to Mz takes the form

$$D_{\text{cir}}(t) = \begin{pmatrix} \cos\psi(t) & -\sin\psi(t) & 0 \\ \sin\psi(t) & \cos\psi(t) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

3. Transition from Mz_{cir} to Mz through an inexactly known orthogonal matrix. For convenience of further calculations, this matrix is represented as $D' = (d'_1, d'_2, d'_3)$.

Thus, the matrix $D(t)$ can be written as a product of the fixed and time-dependent transition matrices:

$$D(t) = D' D_{\text{cir}}(t) D_{\text{fix}}.$$

By definition, the first two rows of the matrix D_{fix} are orthogonal to w. Using this fact, we obtain the following expression for $s'(t)D(t)w$:

$$s'(t)D(t)w = s'(t)D'D_{\text{cir}}(t)D_{\text{fix}}w = s'(t)D' \begin{pmatrix} \cos\psi(t) & -\sin\psi(t) & 0 \\ \sin\psi(t) & \cos\psi(t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = s'(t)d'_3. \qquad (6)$$

By the definition of the transition matrix D', the column d'_3 consists of the projections of the unit basis vectors of the instrumental frame onto the axis $Mz_{\text{cir}3}$. Since the direction of $Mz_{\text{cir}3}$ does not change in time and the rotation is about this axis, the projections of the unit basis vectors of the instrumental frame onto this direction will also remain constant. Therefore, d'_3 can be determined from the a priori information (2):

$$d'_3 = D(t)w = D(0)w = D_{\text{init}}(I_3 + \hat{\beta})w. \qquad (7)$$

In other words, when rotating about a fixed axis, the direction of the vector $D(t)\Omega_x(t)$ remains constant; its averaging yields the vector $s'D_{\text{init}}(I_3 + \hat{\beta})w$, where s' is the mean value of $s'(t)$.

Due to the motion mode under consideration, the averaging result for the vector $D(t)u_x$ will have a special structure as follows.

Lemma 1. *Let the angular rate of the rotating frame be described by the function $s'(t) = s + \varepsilon(t)$, the "programmed" angular rate s be constant and $s > \varepsilon_{\max}$, and the rotation occur about a fixed direction w. Then, under the time averaging of the signal, the projections of the Earth's rotation rate onto the axes of Mz are described by*

$$\bar{u}_z = D_{\text{init}}(I_3 + \hat{\beta})ww^{\mathrm{T}}u_x + u^{\perp}.$$

In addition, the unknown vector $u^{\perp}$ is orthogonal to the vector $D_{\text{init}}(I_3 + \hat{\beta})w$ and its components can be estimated as

$$|u_i^{\perp}| \leqslant u \left(\frac{4}{T(s - \varepsilon_{\max})} + C\frac{\varepsilon_{\max}}{s} \right) \stackrel{\text{def}}{=} u_{\max},$$

where C is a bounded value depending on the initial and final values of the rotation angle on the bench.

Table 1. The orders of values for model parameters

Parameter	The order of value
$\alpha, \alpha_{\max}$	1' $\approx 2.9 \times 10^{-4}$
$\beta, \beta_{\max}$	5' $\approx 1.5 \times 10^{-3}$
$\varepsilon(t)$	5×10^{-6} 1/s
$\varepsilon, \varepsilon_{\max}$	1×10^{-8} 1/s
s	17.5 1/s
T	600–1200 s
Γ_{ii}	5×10^{-3} (5×10^{-5})
$\Gamma_{ij},\ i \neq j$	5×10^{-3} (5×10^{-5})
ν_0	2.4×10^{-7} 1/s (5×10^{-8} 1/s)
$\nu_{\max}$	1.2×10^{-8} 1/s
u	7.292115 $\times 10^{-5}$ 1/s

We obtain an explicit form of the measurement equation by passing to the mean values in (5) and using the expressions (4) and (7) with Lemma 1:

$$\begin{aligned}\tilde{\zeta} &= (I_3+\Gamma)\left(s' D_{\text{init}}(I_3+\hat{\beta})w + D_{\text{init}}(I_3+\hat{\beta})\, ww^{\mathrm{T}}u_x + u^{\perp}\right) + \nu_0 + \delta\tilde{\nu} \\ &= (I_3+\Gamma)\, D_{\text{init}}(I_3+\hat{\beta})\left((s+\varepsilon)(I_3+\hat{\alpha})y + (I_3+\hat{\alpha})yy^{\mathrm{T}}(I_3+\hat{\alpha})^{\mathrm{T}}u_x\right) \\ &\quad + (I_3+\Gamma)\, u^{\perp} + \nu_0 + \delta\tilde{\nu}.\end{aligned} \tag{8}$$

Here, ε (the mean value of the noise $\varepsilon(t)$) and the error $\delta\tilde{\nu}$ (the mean value of the noise $\delta\nu(t)$) are supposed to be bounded:

$$|\varepsilon| \leqslant \varepsilon_{\max}, \quad |\delta\tilde{\nu}_j| \leqslant \nu_{\max}, \quad j = 1,2,3,$$

where a known constant $\nu_{\max}$ characterizes the a priori knowledge of the gyro error.

Depending on the scales of the variables α, β, Γ, ε, and ν, the measurement model can be simplified in different ways by neglecting one or another group of variables. In Table 1 below, we fix the characteristic scales of the model parameters corresponding to the typical accuracies of benches and gyros as well as the accuracy requirements for estimating the parameters Γ and ν_0 (indicated in parentheses).

Note that in the case under consideration, the values $\nu_{\max}$ and $\varepsilon_{\max}$ are much smaller than the characteristic amplitudes of $\delta\nu(t)$ and $\varepsilon(t)$. This corresponds to the averaging of the original signal. The approach proposed in this paper is applicable to other scales of variables as well; the original measurement model can be simplified in other ways depending on the real problem.

With the selected values, the terms in the measurements expression are divided into several groups: non-small terms, such as sy and $\tilde{\zeta}$; terms with a linear dependence on the small parameters $\alpha,\ \beta,\ \varepsilon$, and u; negligibly small second- and third-order infinitesimals not exceeding $\nu_{\max}$; nonlinear terms that cannot be neglected due to their dependence on $s\Gamma D_{\text{init}}\hat{\alpha},\ s\Gamma D_{\text{init}}\hat{\beta},\ sD_{\text{init}}\hat{\alpha}\hat{\beta},\ \Gamma u^{\perp},\ u^{\perp}$ and (possibly) higher values than $\nu_{\max}$.

After eliminating the small terms, Eq. (8) takes the form

$$\begin{aligned}\tilde{\zeta} &= D_{\text{init}}\left(s(I_3+\hat{\alpha}+\hat{\beta})y + s\hat{\beta}\hat{\alpha}y + \varepsilon y + (I_3+\hat{\alpha}+\hat{\beta})yy^{\mathrm{T}}u_x - yy^{\mathrm{T}}\hat{\alpha}u_x\right) \\ &\quad + u^{\perp} + \Gamma u^{\perp} + \Gamma D_{\text{init}}(sy + yy^{\mathrm{T}}u_x) + \Gamma D_{\text{init}}s(\hat{\alpha}+\hat{\beta})y + \nu_0 + \delta\tilde{\nu}.\end{aligned}$$

The terms $sy^{\mathrm{T}}\hat{\beta}\hat{\alpha}y$ and $\Gamma D_{\mathrm{init}}s(\hat{\alpha}+\hat{\beta})y$ can be neglected if their value is comparable with the unremovable noise $\delta\tilde{\nu}$, i.e., $\|sy^{\mathrm{T}}\hat{\beta}\hat{\alpha}y\|_\infty \leqslant \nu_{\max}$ and $\|\Gamma D_{\mathrm{init}}s(\hat{\alpha}+\hat{\beta})y\|_\infty \leqslant \nu_{\max}$. The components of the matrix Γ have a known scale: $|\Gamma_{ij}| \leqslant \Gamma_{\max}$; see Table 1. Therefore, we introduce an additional constraint on the angular rate s:

$$s \leqslant \frac{\nu_{\max}}{2\max\{\alpha_{\max}\beta_{\max}, \Gamma_{\max}(\alpha_{\max}+\beta_{\max})\}} \stackrel{\mathrm{def}}{=} s_{\max}. \tag{9}$$

In other words, the errors α and β will have a smaller effect on the estimation result when rotating the gyro unit on the bench with a lower angular rate.

Under too slow rotation, it may turn out that the averaging error (the term $\Gamma u^{\perp}$) exceeds the required estimation accuracy. Hence, we obtain the second constraint on the parameter s: $\|\Gamma u^{\perp}\|_\infty \leqslant \nu_{\max}$. Due to the a priori known scale of the components of the matrix $\Gamma_{\max}$ and Lemma 1,

$$\|\Gamma u^{\perp}\|_\infty \leqslant 3\Gamma_{\max}\, u\, \left(\frac{4}{(s-\varepsilon_{\max})T} + C\frac{\varepsilon_{\max}}{s}\right) \leqslant \nu_{\max}.$$

The term $\Gamma_{\max} C\, u\, \varepsilon_{\max}/s$ is small compared to $\nu_{\max}$. Hence, the constraint on s takes the form

$$3\Gamma_{\max}\, u\, \frac{4}{(s-\varepsilon_{\max})T} \leqslant \nu_{\max}, \quad \text{or} \quad s \geqslant \frac{12\Gamma_{\max}u}{\nu_{\max}T} + \varepsilon_{\max} \stackrel{\mathrm{def}}{=} s_{\min}. \tag{10}$$

The effect of $u^{\perp}$ can be compensated by radically increasing T. However, see below, this is achieved in a different way (through scalarization).

In addition to rotation with the angular rate sy, another mode of bench tests is possible: the ARS unit is stationary relative to the bench base and the gyros measure the angular rate of the Earth's rotation. In this case, the expression for measurements can be obtained from (5) by substituting $s' = 0$, $s = 0$, $\varepsilon(t) = 0$, and $D(t) = D_{\mathrm{init}}(I_3 + \hat{\beta})$ and passing to the averaged signals $\tilde{\zeta}$:

$$\tilde{\zeta} = (I_3+\Gamma)\, D_{\mathrm{init}}(I_3+\hat{\beta})\, u_x + \nu_0 + \delta\tilde{\nu} = D_{\mathrm{init}}\, u_x + \Gamma\, D_{\mathrm{init}}u_x + D_{\mathrm{init}}\hat{\beta}\, u_x + \nu_0 + \delta\tilde{\nu}$$

within second-order infinitesimals.

Summarizing the results of this section, we formulate the averaged gyro signal model with the constraints (9) and (10):

$$\begin{aligned}\tilde{\zeta} &= D_{\mathrm{init}}\left(s(I_3+\hat{\alpha}+\hat{\beta})y + \varepsilon y + (I_3+\hat{\alpha}+\hat{\beta})yy^{\mathrm{T}}u_x - yy^{\mathrm{T}}\hat{\alpha}u_x\right)\\ &+ \Gamma D_{\mathrm{init}}(sy + yy^{\mathrm{T}}u_x) + u^{\perp} + \nu_0 + \delta\tilde{\nu}, \quad s \in \{0\}\cup[s_{\min}, s_{\max}].\end{aligned} \tag{11}$$

2.3. Measurement Models and Scalarization

In Eq. (11), the input information is the terms $\tilde{\zeta}$ and $D_{\mathrm{init}}(sy + yy^{\mathrm{T}}u_x)$, whereas the "useful signal" is the terms $\Gamma D_{\mathrm{init}}(sy + yy^{\mathrm{T}}u_x) + \nu_0$. The measurement errors consist of the vector $\delta\tilde{\nu}$ arising when averaging the fluctuation noise $\delta\tilde{\nu}(t)$, and unknown systematic errors due to bench inaccuracy (varying with each new bench test). After rearranging the known terms to the left-hand side of Eq. (11), we obtain the linear measurement model

$$z(s,y) = \Gamma D_{\mathrm{init}}(sy + yy^{\mathrm{T}}u_x) + \nu_0 + r + \delta\nu', \tag{12}$$

with the "measurements" $z = z(s,y)$ and their errors $r = r(s,y,\alpha,\beta,\varepsilon)$ and $\delta\nu'$ given by

$$z = \tilde{\zeta} - D_{\mathrm{init}}(sy + yy^{\mathrm{T}}u_x), \quad r = D_{\mathrm{init}}\left(s(\hat{\alpha}+\hat{\beta})y + \varepsilon y + (\hat{\alpha}+\hat{\beta})yy^{\mathrm{T}}u_x - yy^{\mathrm{T}}\hat{\alpha}u_x\right), \tag{13}$$

$$\delta\nu' = \delta\tilde{\nu} + u^{\perp}, \quad |\delta\nu'_j| \leqslant \nu_{\max} + u_{\max} \stackrel{\mathrm{def}}{=} \nu'_{\max}, \quad j = 1,2,3. \tag{14}$$

For definiteness, this model will be called three-dimensional (3D) (since $z(s,y)\in\mathbf{R}^3$) or Model-1.

In model (12)–(14), the component $\delta\tilde{\nu}$ has no intelligible spectrum due to averaging and is therefore exhaustively described by the inequality $|\delta\tilde{\nu}_j|\leqslant\nu_{\max}$. The component $u^\perp$ is not equally arbitrary; see Lemma 1. Therefore, the constraint $|\delta\nu'_j|\leqslant\nu'_{\max}$ in (14) is coarse.

We use the scalarization method [11]: the original 3D measurement equations are multiplied by a known vector (in this case, $\tilde{y}=D_{\text{init}}y$). After this step, all terms representing the product of a skew-symmetric matrix by the vector y in r are reduced. The scalar product $y^{\mathrm{T}}D_{\text{init}}^{\mathrm{T}}u^\perp$ can also be considered approximately equal to 0 by Lemma 1:

$$0=w^{\mathrm{T}}(I_3-\hat{\beta})D_{\text{init}}^{\mathrm{T}}u^\perp=y^{\mathrm{T}}(I_3-\hat{\alpha})(I_3-\hat{\beta})D_{\text{init}}^{\mathrm{T}}u^\perp\approx y^{\mathrm{T}}D_{\text{init}}^{\mathrm{T}}u^\perp \tag{15}$$

(within the infinitesimals of order $(\alpha_{\max}+\beta_{\max})u(\frac{4}{T(s-\varepsilon_{\max})}+C\frac{\varepsilon_{\max}}{s})$). This approximation accuracy is sufficient because

$$(\alpha_{\max}+\beta_{\max})u\left(\frac{4}{T(s-\varepsilon_{\max})}+C\frac{\varepsilon_{\max}}{s}\right)\leqslant\nu_{\max}$$

for the scales of the parameters $\alpha_{\max}$, $\beta_{\max}$, s, T, and $\varepsilon_{\max}$.

Consequently, it becomes possible to pass to a one-dimensional (scalar) measurement model with smaller scale errors:

$$z_{\text{scal}}=z_{\text{scal}}(s,y)=\tilde{y}^{\mathrm{T}}\Gamma D_{\text{init}}(sy+yy^{\mathrm{T}}u_x)+\tilde{y}^{\mathrm{T}}\nu_0+r_{\text{scal}}+\tilde{y}^{\mathrm{T}}\delta\tilde{\nu}, \tag{16}$$

where

$$z_{\text{scal}}=\tilde{y}^{\mathrm{T}}\tilde{\zeta}-s-y^{\mathrm{T}}u_x,\quad r_{\text{scal}}=\varepsilon-y^{\mathrm{T}}\hat{\alpha}u_x. \tag{17}$$

This measurement model will be called scalar or Model-2. It covers the specifics of the term containing the value $u^\perp$, which almost vanishes during scalarization (see (15)). Therefore, the model for the measurement noise z_{scal} is more adequate in this case than for Model-1 (the 3D model), which explains its better accuracy.

3. GUARANTEED ESTIMATION: PROBLEM STATEMENTS

Following the ideas presented in [5, 11], we obtain the unknown matrix Γ and vector ν_0 through guaranteeing estimation. The vector of unknown parameters in this estimation problem consists of the components of the errors Γ and ν_0:

$$\gamma=(\Gamma_{11},\Gamma_{21},\Gamma_{31},\Gamma_{12},\Gamma_{22},\Gamma_{32},\Gamma_{13},\Gamma_{23},\Gamma_{33})^{\mathrm{T}},\quad q=(\gamma^{\mathrm{T}},\nu_{01},\nu_{02},\nu_{03})^{\mathrm{T}}\in\mathbf{R}^{12}.$$

It is required to estimate the scalar value $a^{\mathrm{T}}q$ with a given vector $a\in\mathbf{R}^{12}$. For example, $a=(1,0,\ldots,0)^{\mathrm{T}}$ when estimating the component Γ_{11}.

The desired estimate is a linear functional of the measurements:

$$l(\Phi)=\int\limits_{(y,s)\in\mathrm{S}}\Phi_0^{\mathrm{T}}(y,s)z(y,s)dyds+\sum_{k=1}^{K}\Phi^T(k)z(y(k),s(k)),$$

where the integral is taken over the set

$$\mathrm{S}=\{y\in\mathbf{R}^3:\|y\|_2=1\}\times\{\{0\}\cup[s_{\min},s_{\max}]\},$$

$\Phi_0(\cdot)\colon \mathrm{S} \to \mathbf{R}^3$, $\Phi^{(k)} \in \mathbf{R}^3$, and the aggregate $\{y(k), s(k)\}_{k=1}^K$ specifies the set of isolated positions and angular rates of rotation within the experiment. (For brevity, the set over which the integral is taken will be omitted below.) To each element (y, s) we assign the measurements $z(y, s)$ obtained by processing the gyro signals during rotation with the corresponding angular rate. The structure of measurements can be defined by Model-1 or Model-2 (see the previous section). Let us first formulate the guaranteeing estimation problem for the 3D model and then, by analogy, for the scalar model.

With a formal notation using the Dirac delta function, $l(\Phi)$ can be compactly written as

$$l(\Phi) = \int\limits_{y,s} \Phi^{\mathrm{T}}(y, s) z(y, s) dy ds, \quad \Phi(y, s) = \Phi_0(y, s) + \sum_{k=1}^{K} \Phi^T(k)\, \delta(y - y(k), s - s(k)).$$

We denote by $\mathcal{F}$ the set of all such functions $\Phi(\cdot)$.

Consider the guaranteeing estimation problem for the scalar parameter $a^{\mathrm{T}} q$: find an estimator Φ minimizing the (guaranteeing estimation) error [11]

$$I(\Phi) \to \inf_{\Phi \in \mathcal{F}}, \tag{18}$$

where the objective functional $I(\Phi)$ is the supremum of the error $|l(\Phi) - a^{\mathrm{T}} q|$, i.e.,

$$I(\Phi) = \sup_{(q,\alpha,\beta,\varepsilon,\delta\nu') \in \mathcal{B}'} |l(\Phi) - a^{\mathrm{T}} q|. \tag{19}$$

This supremum is calculated over $\mathcal{B}'$, the set of all admissible values of the unknown parameters $(q, \alpha, \beta, \varepsilon, \delta\tilde{\nu})$:

$$q \in \mathbf{R}^{12}, \quad |\alpha_j| \leqslant \alpha_{\max}, \quad |\beta_j| \leqslant \beta_{\max}, \quad |\delta\nu'_j| \leqslant \nu'_{\max}, \quad j = 1, 2, 3, \quad |\varepsilon| \leqslant \varepsilon_{\max}. \tag{20}$$

The solution of problem (18)–(20) determines the optimal plan of the experiment. In practice, it is often necessary to estimate each component of the vector q. For these purposes, 12 separate problems of the form (20) are solved; in each of them, only one component a is nonzero.

For the guaranteeing estimation problem, an equivalent l_1-approximation problem can be formulated and numerically solved.

Proposition 1. *A function $\Phi(y, s) \in \mathcal{F}$ is the solution of the guaranteeing estimation problem* (18)–(20) *for Model-1* (12)–(13) *if and only if it is the solution of the optimization problem*

$$\int \left(\nu'_{\max} \|\Phi\|_1 + \alpha_{\max} \|C_\alpha \Phi\|_1 + \beta_{\max} \|C_\beta \Phi\|_1 + \varepsilon_{\max} \left| y^{\mathrm{T}} D_{\mathrm{init}}^{\mathrm{T}} \Phi \right| \right) dy ds \to \inf_{\Phi \in \mathcal{F}} \tag{21}$$

subject to the constraints

$$\begin{pmatrix} \int v(y, s) \otimes \Phi \, dy ds \\ \int \Phi \, dy ds \end{pmatrix} = a \tag{22}$$

with the following notations:

$$\begin{aligned} v =& D_{\mathrm{init}}(sy + yy^{\mathrm{T}} u_x), \\ C_\alpha =& \left(s\hat{y} + y^{\mathrm{T}} u_x \hat{y} - \hat{u}_x y y^{\mathrm{T}} \right) D_{\mathrm{init}}^{\mathrm{T}}, \\ C_\beta =& \left(s\hat{y} + y^{\mathrm{T}} u_x \hat{y} \right) D_{\mathrm{init}}^{\mathrm{T}}. \end{aligned} \tag{23}$$

Here, the symbol $\otimes$ stands for the Kronecker product; the vector v and the matrices C_α and C_β are functions of s, y. The proof of Proposition 1 is given in the Appendix.

Similarly, we formulate the guaranteeing estimation problems for Model-2 by defining the one-dimensional estimator

$$\chi(y,s) = \chi_0(y,s) + \sum_{k=1}^{K} \chi(k)\,\delta(y - y(k), s - s(k)).$$

The resulting estimate of the unknown scalar parameter $a^{\mathrm{T}}q$ has the form

$$l(\chi) = \int\limits_{y,s} \chi(y,s) z_{\mathrm{scal}}(y,s)\,dyds.$$

We denote by $\mathcal{X}$ the set of all such functions $\chi(y,s)$ with the described structure.

Proposition 2. *A function $\chi(y,s) \in \mathcal{X}$ is the solution of the guaranteeing estimation problem* (18)–(20) *for Model-2* (16)–(17) *if and only if it is the solution of the optimization problem*

$$\int \left(\nu_{\max}\|D_{\mathrm{init}}y\|_1 + \alpha_{\max}\|\hat{u}_x y\|_1 + \varepsilon_{\max}\right)|\chi(y,s)|\,dyds \to \inf_{\chi\in\mathcal{X}} \tag{24}$$

subject to the constraints

$$\int \chi(y,s) \begin{pmatrix} v(y,s)\otimes D_{\mathrm{init}}y \\ D_{\mathrm{init}}y \end{pmatrix} dyds = a. \tag{25}$$

Proposition 2 is established by analogy to Proposition 1.

4. DISCRETE OPTIMIZATION PROBLEMS

In the variational l_1-approximation problem (21)–(22), the desired variable is the vector function $\Phi(y,s)$ whose argument takes a continuum set of values. This fact complicates numerical solution. For similar guaranteeing estimation problems, it was proved [5, 11] that the optimal estimator Φ differs from zero on a finite set of points. In this paper, we will not obtain an analytical solution: consider a discrete analog of the above problems and solve them numerically instead. Let a function $\Phi(y,s)$ be nonzero on a finite set of points $\{y(k), s(k)\}_{k=1}^{K}$ corresponding to different values of the angular rate vector of the unit on the bench and take values $\Phi(k)$ at them. This discrete set of positions can be specified by introducing spherical coordinates to define the vector y and choose a "value grid" for latitude and longitude with a given step.

Transition from integrals to finite sums in (21)–(22) yields the optimization problem

$$\sum_{k=1}^{K}\Big(\nu'_{\max}\|\Phi(k)\|_1 + \alpha_{\max}\|C_\alpha(k)\Phi(k)\|_1 + \beta_{\max}\|C_\beta(k)\Phi(k)\|_1 + \varepsilon_{\max}|y^{\mathrm{T}}(k)D_{\mathrm{init}}^{\mathrm{T}}\Phi(k)|\Big) \to \inf_{\Phi(1),\ldots,\Phi(K)} \tag{26}$$

subject to the constraints

$$\begin{pmatrix} \sum\limits_{k=1}^{K} v(k)\otimes\Phi(k) \\ \sum\limits_{k=1}^{K}\Phi(k) \end{pmatrix} = a, \tag{27}$$

where the vectors $v(k)$ and the matrices $C_\alpha(k)$ and $C_\beta(k)$ depend on the known arguments $y(k)$ and $s(k)$:

$$v(k) = D_{\text{init}}(s(k)y(k) + y(k)y^{\mathrm{T}}(k)u_x),$$
$$C_\alpha(k) = \Big(s(k)\hat{y}(k) + y^{\mathrm{T}}(k)u_x\hat{y}(k) - \hat{u}_x y(k)y^{\mathrm{T}}(k)\Big)\, D_{\text{init}}^{\mathrm{T}},$$
$$C_\beta(k) = \Big(s(k)\hat{y}(k) + y^{\mathrm{T}}(k)u_x\hat{y}(k)\Big)\, D_{\text{init}}^{\mathrm{T}}.$$

Note that in this problem, the discrete value grid $\{y(k), s(k)\}_{k=1}^{K}$ is considered a priori given and only the values of $\Phi(k)$ have to be found. Multiplying the objective function (26) by the constant $\nu'^{-1}_{\max}$ and introducing the notations

$$\begin{gathered}\frac{\alpha_{\max}}{\nu'_{\max}}\, C_\alpha(k)\Phi(k) = x_\alpha(k), \quad \frac{\beta_{\max}}{\nu'_{\max}}\, C_\beta(k)\Phi(k) = x_\beta(k), \quad \frac{\varepsilon_{\max}}{\nu'_{\max}}\, y^{\mathrm{T}}(k)D_{\text{init}}^{\mathrm{T}}\Phi(k) = x_\varepsilon(k), \\ x = \Big(\Phi^{\mathrm{T}}(1), \ldots, \Phi^{\mathrm{T}}(K), x_\alpha^{\mathrm{T}}(1), \ldots, x_\alpha^{\mathrm{T}}(K), x_\beta^{\mathrm{T}}(1), \ldots, x_\beta^{\mathrm{T}}(K), x_\varepsilon(1), \ldots, x_\varepsilon(K)\Big)^{\mathrm{T}} \in \mathbf{R}^{10K},\end{gathered} \tag{28}$$

we write problem (26)–(27) in a compact form corresponding to the classical l_1-approximation problem

$$\|x\|_1 \to \inf_{x \in \mathbf{R}^{10K}} \tag{29}$$

subject to the linear constraints $A_{eq}x = a_{eq}$.

The matrix and vector from the constraint equation can be represented in the block form:

$$A_{eq} = \begin{pmatrix} A_\alpha & I_{3K} & 0_{3K\times 3K} & 0_{K\times K} \\ A_\beta & 0_{3K\times 3K} & I_{3K} & 0_{K\times K} \\ A_\varepsilon & 0_{K\times 3K} & 0_{K\times 3K} & I_K \\ A_\Phi & 0_{12\times 3K} & 0_{12\times 3K} & 0_{12\times K} \end{pmatrix} \in \mathbf{R}^{(7K+12)\times 10K}, \quad a_{eq} = \begin{pmatrix} 0_{7K\times 1} \\ a \end{pmatrix} \in \mathbf{R}^{7K+12},$$

where, due to (27) and (28),

$$A_\alpha = \frac{\alpha_{\max}}{\nu_{\max}} \begin{pmatrix} C_\alpha(1) & 0_{3\times 3} & \ldots & 0_{3\times 3} \\ 0_{3\times 3} & C_\alpha(2) & \ldots & 0_{3\times 3} \\ \ldots & \ldots & \ddots & \ldots \\ 0_{3\times 3} & \ldots & 0_{3\times 3} & C_\alpha(K) \end{pmatrix},$$

$$A_\beta = \frac{\beta_{\max}}{\nu_{\max}} \begin{pmatrix} C_\beta(1) & 0_{3\times 3} & \ldots & 0_{3\times 3} \\ 0_{3\times 3} & C_\beta(2) & \ldots & 0_{3\times 3} \\ \ldots & \ldots & \ddots & \ldots \\ 0_{3\times 3} & \ldots & 0_{3\times 3} & C_\beta(K) \end{pmatrix},$$

$$A_\varepsilon = \frac{\varepsilon_{\max}}{\nu_{\max}} \begin{pmatrix} y^{\mathrm{T}}(1)D_{\text{init}}^{\mathrm{T}} & 0_{1\times 3} & \ldots & 0_{1\times 3} \\ 0_{1\times 3} & y^{\mathrm{T}}(2)D_{\text{init}}^{\mathrm{T}} & \ldots & 0_{1\times 3} \\ \ldots & \ldots & \ddots & \ldots \\ 0_{1\times 3} & \ldots & 0_{1\times 3} & y^{\mathrm{T}}(K)D_{\text{init}}^{\mathrm{T}} \end{pmatrix},$$

$$A_\Phi = \begin{pmatrix} v_1(1) & v_1(2)I_3 & \ldots & v_1(K)I_3 \\ v_2(1)I_3 & v_2(2)I_3 & \ldots & v_2(K)I_3 \\ v_3(1)I_3 & v_3(2)I_3 & \ldots & v_3(K)I_3 \\ I_3 & I_3 & \ldots & I_3 \end{pmatrix}.$$

A discrete analog of the scalarized model (24)–(25) can be formulated by analogy. In this problem, it is required to minimize the sum of the moduli of the components of the unknown vector under linear equality constraints:

$$\sum_{k=1}^{K} \left(\nu_{\max}\|D_{\text{init}}y(k)\|_1 + \alpha_{\max}\|\hat{u}_x y(k)\|_1 + \varepsilon_{\max}\right)|\chi(k)| \to \inf_{\chi(1),\ldots,\chi(K)} \tag{30}$$

subject to

$$\sum_{k=1}^{K} \chi(k) \begin{pmatrix} v(k) \otimes D_{\text{init}}y(k) \\ D_{\text{init}}y(k) \end{pmatrix} = a. \tag{31}$$

In matrix form, the problem is written as

$$\|x_\chi\|_1 \to \inf_{x_\chi \in \mathbf{R}^K} \tag{32}$$

subject to the constraints $A_\chi x_\chi = a$.

Here, we adopt the notations

$$x_{\chi k} = \rho_k \chi(k), \quad \rho_k = \nu_{\max}\|D_{\text{init}}y(k)\|_1 + \alpha_{\max}\|\hat{u}_x y(k)\|_1 + \varepsilon_{\max}, \quad k = 1, \ldots, K;$$

$$A_\chi = \begin{pmatrix} \rho_1^{-1}v_1(1)D_{\text{init}}y(1) & \rho_2^{-1}v_1(2)D_{\text{init}}y(2) & \ldots & \rho_K^{-1}v_1(K)D_{\text{init}}y(K) \\ \rho_1^{-1}v_2(1)D_{\text{init}}y(1) & \rho_2^{-1}v_2(2)D_{\text{init}}y(2) & \ldots & \rho_K^{-1}v_2(K)D_{\text{init}}y(K) \\ \rho_1^{-1}v_3(1)D_{\text{init}}y(1) & \rho_2^{-1}v_3(2)D_{\text{init}}y(2) & \ldots & \rho_K^{-1}, v_3(K)D_{\text{init}}y(K) \\ \rho_1^{-1}D_{\text{init}}y(1) & \rho_2^{-1}D_{\text{init}}y(2) & \ldots & \rho_K^{-1}D_{\text{init}}y(K) \end{pmatrix} \in \mathbf{R}^{12\times K}.$$

Such convex optimization problems can be solved by various numerical methods, e.g., the interior point method [13], ADMM [14], and the method of variationally weighted quadratic approximations [5]. Unlike the problem for the 3D model (29), problems (32) have a smaller dimension of the unknown vector and constraint matrices (7–10 times less variables and constraints). Therefore, they better suit numerical solution in the case of large values K.

Thus, the optimal estimators $\Phi(k)$ and $\chi(k)$ obtained by solving the l_1-approximation problems (29) or (32) yield the target values of the angular rates $s(k)y(k)$ of the gyro unit on the bench. As a rule, a small number of angular rate positions correspond to non-zero values of $\Phi(k)$ or $\chi(k)$. (This is a common property of guaranteeing estimation solutions; for example, see justification in the book [5].) We denote this subset by $\mathcal{K}$, $\mathcal{K} \subset \{1, \ldots, K\}$.

The guaranteeing estimation algorithm for the ARS unit errors is a series of steps. At each step, the following operations are carried out for each $k \in \mathcal{K}$:

(1) The gyro unit is rotated with the angular rate $s(k)y(k)$, and the set of gyro readings $\zeta(t) \in \mathbf{R}^3$ is formed.

(2) The signal $\zeta(t)$ is averaged on the time interval T of the fixed-rate rotation:

$$\tilde{\zeta} = \left(\sum_{t=0}^{\mathrm{T}} \zeta(t)\right)/(T+1).$$

(3) According to (13)–(17), the measurements $z(y(k), s(k))$ and $z_{\text{scal}}(y(k), s(k))$ are formed for the linear estimation models.

Then the unknown parameter $a^{\mathrm{T}}q$ is estimated as

$$\sum_{k\in\mathcal{K}} \Phi^{\mathrm{T}}(k) z(y(k), s(k)) \quad \text{or} \quad \sum_{k\in\mathcal{K}} \chi(k) z_{\text{scal}}(y(k), s(k)).$$

5. NUMERICAL EXAMPLES

Consider several numerical examples illustrating the application of the guaranteeing estimation methods proposed in this paper. The practical implementation of the algorithm includes several stages as follows: solving the guaranteeing estimation problems; modeling the signal, i.e., the measurements $\zeta(t)$ for given "true" values of the errors and the unknown parameters Γ and ν_0; building the estimates Γ and ν_0 and comparing them with the "true" values. The corresponding code was implemented in Python and standard procedures from CVXPY[1] were used to solve the l_1-approximation problems (29) and (32).

The typical values of the errors and model parameters were selected according to Table 1. The admissible limits for the angular rate were defined by formulas (9) and (10): $s_{\min} = 1.25\,^\circ/\text{s}$ and $s_{\max} = 3.28\,^\circ/\text{s}$. In the model example, we supposed the following: the absolute angular rate takes the value $s = 0$ (no rotation) and two values from the segment $[s_{\min}, s_{\max}]$, i.e., $s_1 = 1.5\,^\circ/\text{s}$ and $s_2 = 2\,^\circ/\text{s}$; the vectors $y(k)$ are uniformly located on the unit sphere; $D_{\text{init}} = I_3$. Let us describe the resulting solutions for each group of the unknown parameters Γ and ν_0.

For the full and scalarized models, the optimal estimators for the diagonal components Γ_{ii} have the form

$$\begin{aligned}\Phi_{ii}(y,s) &= \chi_0 e_i \delta(y - e_i, s - s_2) - \chi_0 e_i \delta(y + e_i, s - s_2),\\ \chi_{ii}(y,s) &= \chi_0 \delta(y - e_i, s - s_2) + \chi_0 \delta(y + e_i, s - s_2),\end{aligned}$$

where e_i is a unit vector with ith component equal to 1 and χ_0 is some value numerically determined in the solution of the optimization problem.

In other words, to estimate, e.g., the component Γ_{11} (the scaling factor of the first gyro's error), it is necessary to carry out two series of measurements, rotating the unit along the sensitivity axis of this gyro with the maximum angular rate $s = s_2$ first in one direction ($y(1) = (1,0,0)^{\mathrm{T}}$) and then in the other ($y(2) = (-1,0,0)^{\mathrm{T}}$).

For the off-diagonal elements $\Gamma_{12} = \Gamma_{21}$, the optimal estimators are given by

$$\begin{aligned}\Phi_{12}(y,s) = &\begin{pmatrix}\Phi_1\\ \Phi_1\\ 0\end{pmatrix}\delta(y - e(\pi/4), s - s_2) + \begin{pmatrix}\Phi_2\\ -\Phi_1\\ 0\end{pmatrix}\delta(y - e(3\pi/4), s - s_2)\\ &+ \begin{pmatrix}-\Phi_2\\ -\Phi_2\\ 0\end{pmatrix}\delta(y - e(5\pi/4), s - s_2) + \begin{pmatrix}-\Phi_1\\ \Phi_2\\ 0\end{pmatrix}\delta(y - e(7\pi/4), s - s_2),\\ \chi_{12}(y,s) = &\chi_0\delta(y - e(\pi/4), s - s_2) + \chi_0\delta(y - e(3\pi/4), s - s_2)\\ &+ \chi_0\delta(y - e(5\pi/4), s - s_2) + \chi_0\delta(y - e(7\pi/4), s - s_2)\end{aligned}$$

with the following notations: $e(\theta) = (\cos\theta, \sin\theta, 0)^{\mathrm{T}}$ are the unit vectors corresponding to the rotation by the angle θ in the plane e_1e_2; Φ_1 and Φ_2 are the values numerically determined in the solution of the optimization problem; χ_0 is the same value as for the diagonal elements. The estimators for the components $\Gamma_{13} = \Gamma_{31}$ and $\Gamma_{23} = \Gamma_{32}$ are determined by analogy within the rearrangements of the indices in the corresponding vectors. Thus, the optimal experiment for estimating the misalignment angles between the gyro sensitivity axes consists of four series of measurements; in each of them, rotation is performed along the bisector of the angle between the coordinate axes with the maximum admissible angular rate (by absolute value).

[1] An open source Python-embedded modeling language for convex optimization problems; https://web.stanford.edu/boyd/papers/pdf/cvxpy_rewriting.pdf.

Table 2. Guaranteed estimation errors

Model	Variable		
	Γ_{ii}	$\Gamma_{ij},\ i \neq j$	ν_{0i}
The order of value	$[0.7;\ 1.3] \times 10^{-3}$	$[4;\ 6] \times 10^{-3}$	$[2; 3] \times 10^{-7}$
Required accuracy	5×10^{-5}	5×10^{-5}	5×10^{-8}
Model-1	2.01×10^{-4}	5.67×10^{-4}	7.01×10^{-6}
Model-2	2.23×10^{-6}	3.45×10^{-6}	5.16×10^{-8}

Table 3. Average estimation errors

Model	Variable		
	Γ_{ii}	$\Gamma_{ij},\ i \neq j$	ν_{0i}
The order of value	$[0.7;\ 1.3] \times 10^{-3}$	$[4;\ 6] \times 10^{-3}$	$[2; 3] \times 10^{-7}$
Model-1	1.25×10^{-6} (0.13%)	1.33×10^{-6} (0.02%)	4.88×10^{-8} (22%)
Model-2	1.20×10^{-6} (0.13%)	1.24×10^{-6} (0.03%)	2.89×10^{-8} (14%)

The optimal estimators for the zero biases ν_{0i} have the following structure:

$$\Phi_{\nu i}(y,s) = \phi_1 e_i \delta(y - e_i, s - s_2) + \phi_2 e_i \delta(y + e_i, s - s_2) + \phi_3 e_i \delta(y - e_i, s - s_1) + \phi_4 e_i \delta(y + e_i, s - s_1),$$
$$\chi_{\nu i}(y,s) = \chi_1 \delta(y - e_i, s - s_2) - \chi_2 \delta(y + e_i, s - s_2) + \chi_3 \delta(y - e_i, s - s_1) - \chi_4 \delta(y + e_i, s - s_1),$$

where the values $\phi_i > 0$ and $\chi_i > 0$ are numerically determined in the solution of the optimization problem. They are close to 1/4.

Clearly, the rotation directions and weight coefficients obtained by guaranteeing estimation have a simple geometric structure: the optimal motion modes imply rotation with the maximum admissible angular rate, in the direction coinciding (within the initial alignment error) either with the gyro sensitivity axes or with the bisectors of the angles between these axes. A similar result was established using guaranteeing estimation in the calibration problem of an accelerometer unit [11]. However, unlike the latter problem, the mathematical model of measurements in this paper is significantly more complex and depends on a larger number of parameters. Therefore, it is difficult to justify the optimal structure of the estimator analytically, and numerical methods are employed to find solutions.

The main quality indicator of the solutions is the guaranteeing estimation error, which does not depend on particular realizations of errors and measurements. Table 2 shows the guaranteeing estimation errors corresponding to the optimal estimators for the components Γ and ν_0.

For the model parameters given in Table 1, the required guaranteeing estimation accuracy is achieved for all components of the matrix Γ within Model-2. Model-1 can lead to errors of about 10% of the parameter estimated; for the components ν_{0i}, the required accuracy is achieved within Model-2.

We present the estimation results for the components Γ, ν_0 from a series of numerical experiments with this procedure under particular realizations of the systematic errors and noises in the measurements. The "true" values of these parameters and errors α, β, $\varepsilon(t)$, and $\delta\nu(t)$ were outputted using a random number generator. For each parameter Γ_{ij}, ν_{0i}, the modeling and estimation procedure was repeated 20 times to evaluate typical deviations of the estimate from the true value. Table 3 shows the average absolute ($|\Gamma^0_{ij} - \Gamma_{ij}|$) and relative ($|\Gamma^0_{ij} - \Gamma_{ij}|/|\Gamma_{ij}|$, in parentheses) deviations of the estimates obtained with Models-1 and -2 from the true values. Unlike guaranteeing errors, these deviations do not determine an upper bound on the error but characterize both the accuracy of the models and solutions of the estimation problems. This is important because the input information $\zeta(t)$ is constructed using a nonlinear model of the original signal (3), and the

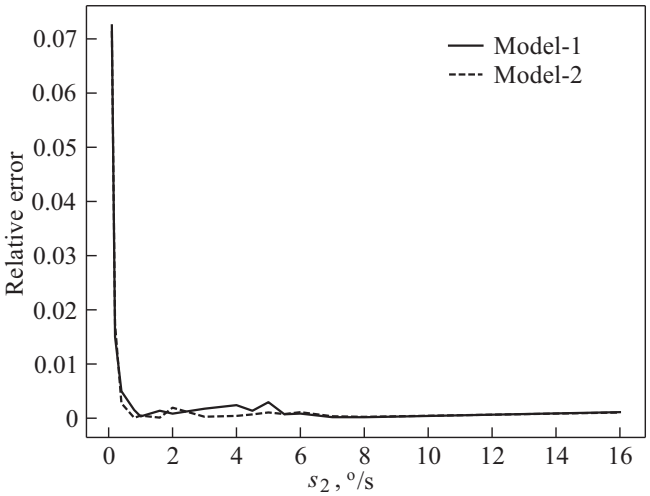


Fig. 1.

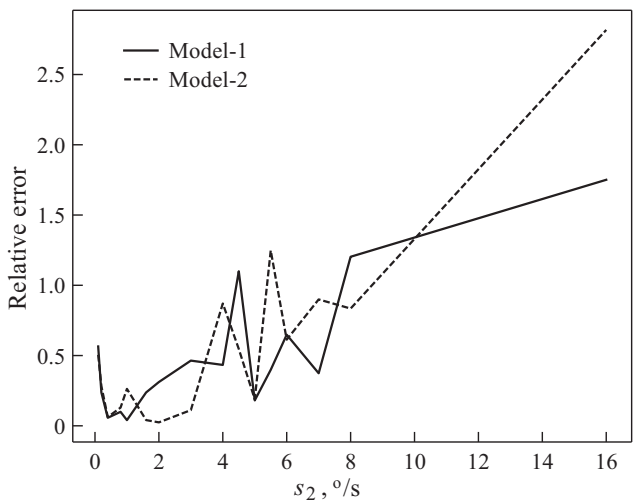


Fig. 2.

optimal solution is the result of applying guaranteeing estimation to linearized Models-1 and -2. In other words, this experiment reflects the effect of different factors on the estimation result: the errors and noises in the measurements and the errors due to the transition to simpler linear models.

According to the series of numerical examples, the estimates of the unknown parameters based on the proposed algorithm with the measurement information are close to their "true" values and the deviations lie within the guaranteeing estimation errors.

Let us illustrate how the solution accuracy depends on the choice of the angular rate s_2. Consider an experiment in which, for the same pre-selected values Γ and ν_o, the unknown parameters are estimated using the two models described above for 16 different values of s_2. The relative accuracy of the resulting estimates is presented in the graphs below: the estimation errors for Γ_{11} (Fig. 1; for Γ_{12} the results are similar) and the estimation errors for ν_{01} (Fig. 2).

According to the graph, the critical drop in accuracy is observed under the minimum values of the angular rate (less than the threshold $s_{\min}$ obtained in Section 2.2). When estimating the matrix Γ, the approaches appear to be insensitive to the increase in the angular rate s_2. However, when estimating ν_0, the accuracy deteriorates as s_2 increases, especially in the case of exceeding the threshold $s_2 > s_{\max} = 3.28\ ^\circ/\text{s} = 0.057\ 1/\text{s}$.

6. CONCLUSIONS

This paper has developed a calibration procedure for a gyro unit with three main ideas as follows: derivation of linear measurement models for averaged signals, scalarization, and guaranteeing estimation. The guaranteeing estimation problems have been reduced to discrete l_1-approximation problems, which are solved using numerical algorithms. An important advantage is that this procedure yields an optimal experiment plan as a result of solving the estimation problem. Guaranteed estimation leads to simple-structure solutions: from a large set of admissible directions and angular rates, an optimal combination contains 2–4 rotation modes. This calibration procedure and the corresponding software implementations can be extended with minimum changes to more complex systems, e.g., the ones with a limited number of rotation directions and with temperature-dependent sensor errors.

Besides inertial navigation, the approach proposed above or its modifications can be used in other applications requiring an optimal set of measurements to estimate unknown parameters or an experiment plan from a set of admissible scenarios.

APPENDIX

Proof of Lemma 1. On the time interval T the rotation occurs about a fixed direction. Therefore, the rotation matrix D_{cir} is decomposed by averaging as follows:

$$\bar{D}_{\text{cir}} = \bar{D}_{\text{cir1}} + \bar{D}_{\text{cir2}}, \quad \bar{D}_{\text{cir1}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \bar{D}_{\text{cir2}} = \begin{pmatrix} c_1 & -c_2 & 0 \\ c_2 & c_1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{A.1}$$

where c_i is the result of the time averaging of the functions $\sin\psi(t)$ and $\cos\psi(t)$.

Due to formulas (6), (7), and (A.1), in the course of averaging, the vector u_z is represented as the sum of two terms, one proportional to w and the other orthogonal to $D(0)w$:

$$\begin{gathered} \bar{u}_z = \bar{D}u_x = D'\bar{D}_{\text{cir1}}D_{\text{fix}}u_x + D'\bar{D}_{\text{cir2}}D_{\text{fix}}u_x \\ = D'\begin{pmatrix} 0_{1\times 3} \\ 0_{1\times 3} \\ w^{\mathrm{T}} \end{pmatrix} u_x + u^{\perp} = d_3' w^{\mathrm{T}}u_x + u^{\perp} = D_{\text{init}}(I_3 + \hat{\beta})ww^{\mathrm{T}}u_x + u^{\perp}, \end{gathered} \tag{A.2}$$

where $u^{\perp} = D'\bar{D}_{\text{cir2}}D_{\text{fix}}u_x$.

The orthogonality of $u^{\perp}$ to the direction $D(0)w = D_{\text{init}}(I_3 + \hat{\beta})w$ can be established using formulas (2) and (6): $D(0)w = D'D_{\text{cir}}(0)D_{\text{fix}}w = D'\,(0,0,1)^{\mathrm{T}}$; the corresponding scalar product is explicitly calculated as

$$w^{\mathrm{T}}D(0)^{\mathrm{T}}u^{\perp} = (0,0,1)D'^{\mathrm{T}}D'\begin{pmatrix} c_1 & -c_2 & 0 \\ c_2 & c_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} D_{\text{fix}}u_x = (0,0,1)^{\mathrm{T}}\begin{pmatrix} c_1 & -c_2 & 0 \\ c_2 & c_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} D_{\text{fix}}u_x = 0.$$

Using the component c_1 as an example, we explain the idea of estimating from above the result of the time averaging of the function $\cos\psi(t)$. Consider the continuous case of averaging:

$$c_1 = \frac{1}{T}\int_0^{\mathrm{T}} \cos\psi(t)dt.$$

The dynamics of the angle ψ are described by a differential equation and constraints on the functions on its right-hand side:

$$\frac{d\psi(t)}{dt} = s + \varepsilon(t), \quad \psi(0) = \psi_0, \quad |\varepsilon(t)| \leqslant \varepsilon_{\max}, \quad s + \varepsilon(t) > 0.$$

The change of variables $t = t(\psi)$, $\epsilon(\psi) = \varepsilon(t(\psi))$, $|\epsilon(\psi)| \leqslant \varepsilon_{\max}$, in the integral yields

$$\int\limits_0^{\mathrm{T}} \cos\psi(t)dt = \int\limits_{\psi_0}^{\psi(T)} \frac{\cos\psi}{s+\epsilon(\psi)}d\psi.$$

This integral can be written as the sum of integrals on the half-periods of the function $\cos\psi$ (intervals where the function has a fixed sign) and two integrals corresponding to the time intervals at the beginning and end of the interval $[\psi_0, \psi(T)]$. For example, if $\psi_0 < \pi/2$, this interval is represented as follows:

$$[\psi_0, \psi(T)] = [\psi_0, \pi/2] \cup [\pi/2, 3\pi/2] \cup [3\pi/2, 5\pi/2] \cup \ldots \cup [\pi/2 + 2\pi n_{\mathrm{cir}}, \psi(T)],$$

where n_{cir} is the number of complete revolutions of the system about the axis of rotation and the length of the last interval does not exceed π, i.e., $\pi/2 + 2\pi n_{\mathrm{cir}} \leqslant \psi(T) \leqslant 3\pi/2 + 2\pi n_{\mathrm{cir}}$.

The integrand on each such interval has a fixed sign, and the maximum value of the integrand (hence, that of the integral) is achieved at $\epsilon(\psi) = -\mathrm{sgn}(\cos\psi)\,\varepsilon_{\max}$:

$$\int \frac{\cos\psi}{s+\epsilon(\psi)}d\psi \leqslant \int \frac{\cos\psi}{\min_{|\epsilon|\leqslant\varepsilon_{\max}}(s+\epsilon)}d\psi = \int \frac{\cos\psi}{s - \mathrm{sgn}(\cos\psi)\,\epsilon_{\max}}d\psi.$$

Therefore, each integral can be estimated bilaterally (from below and above):

$$\left|\int\limits_{\psi_0}^{\pi/2} \frac{\cos\psi}{s+\epsilon(\psi)}d\psi\right| \leqslant \frac{2}{s-\varepsilon_{\max}}, \quad \left|\int\limits_{\pi/2+2\pi n_{\mathrm{cir}}}^{\psi(T)} \frac{\cos\psi}{s+\epsilon(\psi)}d\psi\right| \leqslant \frac{2}{s-\varepsilon_{\max}},$$

$$\frac{-2}{s-\varepsilon_{\max}} \leqslant \int\limits_{\pi/2}^{3\pi/2} \frac{\cos\psi}{s+\epsilon(\psi)}d\psi \leqslant \frac{-2}{s+\varepsilon_{\max}}, \quad \frac{2}{s+\varepsilon_{\max}} \leqslant \int\limits_{3\pi/2}^{5\pi/2} \frac{\cos\psi}{s+\epsilon(\psi)}d\psi \leqslant \frac{2}{s-\varepsilon_{\max}}.$$

As a result, the absolute value of the integral on the averaging interval admits the following upper bound:

$$\left|\int\limits_{\psi_0}^{\psi(T)} \frac{\cos\psi}{s+\varepsilon(\psi)}d\psi\right| \leqslant \left|\int\limits_{\psi_0}^{\pi/2} \frac{\cos\psi}{s+\varepsilon(\psi)}d\psi\right| + \left|\int\limits_{\pi/2+2\pi n_{\mathrm{cir}}}^{\psi(T)} \frac{\cos\psi}{s+\varepsilon(\psi)}d\psi\right|$$

$$+ \left|\sum_{j=1}^{n_{\mathrm{cir}}} \left(\int\limits_{\pi/2}^{3\pi/2} \frac{\cos\psi}{s+\varepsilon(\psi)}d\psi + \int\limits_{3\pi/2}^{5\pi/2} \frac{\cos\psi}{s+\varepsilon(\psi)}d\psi\right)\right|$$

$$\leqslant \frac{4}{s-\varepsilon_{\max}} + \left|\sum_{j=1}^{n_{\mathrm{cir}}} \frac{-2}{s+\varepsilon_{\max}} + \frac{2}{s-\varepsilon_{\max}}\right| \leqslant \frac{4}{s-\varepsilon_{\max}} + \frac{n_{\mathrm{cir}}\,4\varepsilon_{\max}}{(s+\varepsilon_{\max})(s-\varepsilon_{\max})}.$$

The angular rate and the number of complete revolutions of the system are related by

$$sT = 2\pi\, n_{\mathrm{cir}} + \Delta\psi$$

for some $\Delta\psi \leqslant 2\pi$. Consequently,

$$|c_1| = \left|\frac{1}{T}\int\limits_0^{\mathrm{T}} \cos\psi(t)dt\right| \leqslant \frac{4}{T\,(s-\varepsilon_{\max})} + \frac{n_{\mathrm{cir}}\,4\varepsilon_{\max}}{T\,(s^2-\varepsilon_{\max}^2)}$$
$$= \frac{4}{T\,(s-\varepsilon_{\max})} + \frac{(sT-\Delta\psi)\,4\varepsilon_{\max}}{2\pi\;sT\;s(1-\varepsilon_{\max}^2/s^2)} = \frac{4}{T\,(s-\varepsilon_{\max})} + \frac{2\,(1-\Delta\psi/(sT))}{\pi\,(1-\varepsilon_{\max}^2/s^2)}\;\frac{\varepsilon_{\max}}{s}.$$

Thus, we obtain

$$|c_1| \leqslant \frac{4}{T\,(s-\varepsilon_{\max})} + C\frac{\varepsilon_{\max}}{s},$$

where the parameter $C = \frac{2}{\pi\,(1-\varepsilon_{\max}^2/s^2)}$ is an upper bound for the fraction $\frac{2(1-\Delta\psi/(sT))}{\pi\,(1-\varepsilon_{\max}^2/s^2)}$.

Proof of Proposition 1. We transform the integrand of the objective function into problem (18) by substituting formulas (12) and (13) with the additional notation $v = v(s,y) = D_{\mathrm{init}}(sy+yy^{\mathrm{T}}u_x)$:

$$\Phi^{\mathrm{T}}z = \Phi^{\mathrm{T}}\left(\Gamma D_{\mathrm{init}}(sy+yy^{\mathrm{T}}u_x) + \nu_0 + r + \delta\nu'\right)$$
$$= \Phi^{\mathrm{T}}\Gamma v + \Phi^{\mathrm{T}}\nu_0 + \Phi^{\mathrm{T}}\delta\nu' + \Phi^{\mathrm{T}}D_{\mathrm{init}}\left(-s(\hat{y}\alpha+\hat{y}\beta) + \varepsilon y + y^{\mathrm{T}}u_x(\hat{\alpha}+\hat{\beta})y + yy^{\mathrm{T}}\hat{u}_x\alpha\right)$$
$$= (v\otimes\Phi)^{\mathrm{T}}\gamma + \Phi^{\mathrm{T}}\nu_0 + \Phi^{\mathrm{T}}\delta\nu' + \Phi^{\mathrm{T}}D_{\mathrm{init}}\left(-s(\hat{y}\alpha+\hat{y}\beta) + \varepsilon y - y^{\mathrm{T}}u_x(\hat{y}\alpha+\hat{y}\beta) + yy^{\mathrm{T}}\hat{u}_x\alpha\right).$$

Hence,

$$\Phi^{\mathrm{T}}z = (v\otimes\Phi)^{\mathrm{T}}\gamma + \Phi^{\mathrm{T}}\nu_0 + \Phi^{\mathrm{T}}\delta\nu' + \varepsilon\Phi^{\mathrm{T}}D_{\mathrm{init}}y$$
$$+ \Phi^{\mathrm{T}}D_{\mathrm{init}}\left(-s\hat{y} - y^{\mathrm{T}}u_x\hat{y} + yy^{\mathrm{T}}\hat{u}_x\right)\alpha + \Phi^{\mathrm{T}}D_{\mathrm{init}}\left(-s\hat{y} - y^{\mathrm{T}}u_x\hat{y}\right)\beta. \tag{A.3}$$

These formulas involve, first, the properties of matrix operations

$$\Phi^{\mathrm{T}}\Gamma v = (\Phi^{\mathrm{T}}\otimes v^{\mathrm{T}})\gamma = (v\otimes\Phi)^{\mathrm{T}}\gamma, \quad \hat{\alpha}y = -\hat{y}\alpha$$

and, second, the possibility of transferring the scalar product $y^{\mathrm{T}}u_x$ to the other part of the corresponding multiplier group: $\hat{\alpha}yy^{\mathrm{T}}u_x = -y^{\mathrm{T}}u_x\,\hat{y}\alpha$.

Let us define the matrices C'_α and C'_β :

$$C'_\alpha = D_{\mathrm{init}}\left(-s\hat{y} - y^{\mathrm{T}}u_x\hat{y} - y^{\mathrm{T}}u_x\hat{y}\right), \quad C'_\beta = D_{\mathrm{init}}\left(-s\hat{y} - y^{\mathrm{T}}u_x\hat{y}\right).$$

Then the right-hand side of (A.3) is represented as a function that linearly depends on the variables $q, \alpha, \beta, \varepsilon$, and $\delta\tilde{\nu}$:

$$\Phi^{\mathrm{T}}(y,s)z(y,s) = (v\otimes\Phi)^{\mathrm{T}}\gamma + \Phi^{\mathrm{T}}\nu_0 + \Phi^{\mathrm{T}}\delta\nu' + \Phi^{\mathrm{T}}C'_\alpha\alpha + \Phi^{\mathrm{T}}C'_\beta\beta + \varepsilon\Phi^{\mathrm{T}}D_{\mathrm{init}}y. \tag{A.4}$$

Substituting formula (A.4) into the original objective functional (19) yields

$$I(\Phi) = \sup_{(q,\alpha,\beta,\varepsilon,\delta\nu')\in\mathcal{B}'} |l(\Phi) - a^{\mathrm{T}}q|$$
$$= \sup_{(q,\alpha,\beta,\varepsilon,\delta\nu')\in\mathcal{B}'} \left|\int\left((v\otimes\Phi)^{\mathrm{T}}\gamma + \Phi^{\mathrm{T}}\nu_0 + \Phi^{\mathrm{T}}\delta\nu' + \Phi^{\mathrm{T}}C'_\alpha\alpha + \Phi^{\mathrm{T}}C'_\beta\beta + \varepsilon\Phi^{\mathrm{T}}D_{\mathrm{init}}y\right)dyds - a^{\mathrm{T}}q\right|.$$

Since $q = \mathrm{col}(\gamma, \nu_0)$, the function $l(\Phi) - a^{\mathrm{T}}q$ linear depends on q, and the multiplier at q is

$$\begin{pmatrix} \int v \otimes \Phi \, dyds \\ \int \Phi \, dyds \end{pmatrix} - a.$$

Therefore, if condition (22) is violated, we have $\sup_{q \in \mathbf{R^{12}}} |l(\Phi) - a^{\mathrm{T}}q| = +\infty$ for a fixed Φ and arbitrary admissible $\alpha, \beta, \varepsilon$, and $\delta\nu'$. Consequently,

$$\sup_{(q,\alpha,\beta,\varepsilon,\delta\nu') \in \mathcal{B}'} |l(\Phi) - a^{\mathrm{T}}q| = \sup_{(q,\alpha,\beta,\varepsilon,\delta\nu') \in \mathcal{B}'} \left| \int \left(\Phi^{\mathrm{T}} \delta\tilde{\nu} + \Phi^{\mathrm{T}} C'_\alpha \alpha + \Phi^{\mathrm{T}} C'_\beta \beta + \varepsilon \Phi^{\mathrm{T}} D_{\mathrm{init}} y \right) dyds \right|.$$

In other words, it is necessary to maximize the absolute value of a linear function where each term depends on only one variable not figuring in the other terms. This means that the maximum can be found independently in each of the variables. For a fixed Φ, the maximum is determined in an explicit form:

$$\begin{aligned} \sup_{\alpha:\, |\alpha_i| \leqslant \alpha_{\max}} \int \Phi^{\mathrm{T}} C'_\alpha \alpha dyds &= \sup_{\alpha:\, |\alpha_i| \leqslant \alpha_{\max}} \int \left(\sum_{i=1}^{3} (C'^{\mathrm{T}}_\alpha \Phi)_i \alpha_i \right) dyds \\ &= \sum_{i=1}^{3} \sup_{\alpha_i:\, |\alpha_i| \leqslant \alpha_{\max}} \int (C'^{\mathrm{T}}_\alpha \Phi)_i \alpha_i dyds \\ = \int \left(\sum_{i=1}^{3} \alpha_{\max} \operatorname{sgn}((C'^{\mathrm{T}}_\alpha \Phi)_i) \, (C'^{\mathrm{T}}_\alpha \Phi)_i \right) dyds &= \int \|C_\alpha \Phi\|_1 \, dyds. \end{aligned}$$

A similar chain of considerations applies to the other terms in the objective functional (18). Thus, the explicit calculation of the supremum of the original objective functional finally leads to the optimization problem (21)–(22).

REFERENCES

1. Lidov, M.L., On the A Priori Accuracy of Parameter Estimation by Means of the Least Squares Method, *Kosm. Issled.*, 1964, vol. 2, no. 5, pp. 713–715.
2. Krasovskii, N.N., On the Theory of Controllability and Observability of Linear Dynamic Systems, *J. Appl. Math. Mech.*, 1964, vol. 28, no. 1, pp. 1–14.
3. Lidov, M.L., Minimax Estimation Methods, *Preprint No. 71 of Keldysh Inst. of Applied Mathematics, Russ. Acad. Sci.*, Moscow, 2010.
4. Matasov, A.I., *Estimators for Uncertain Dynamic Systems*, Dordrecht–Boston–London: Springer Science+Business Media, 2013.
5. Akimov, P.A., Derevyankin, A.V., and Matasov, A.I., *Garantiruyushchii podkhod i l_1-approksimatsiya v zadachakh otsenivaniya parametrov BINS pri stendovykh ispytaniyakh* (Guaranteeing Approach and l_1-norm Approximation in the Problems of SDINS Parameter Estimation under Bench Testing), Moscow: Mosk. Gos. Univ., 2012.
6. Kozlov, A.V., Parusnikov, N.A., Vavilova, N.B., Tarygin, I.E., and Golovan, A.A., Dynamic Calibration of Assembled Strapdown Inertial Navigation Systems on a Rate Table, *Izv. SFedU. Engin. Sci.*, 2018, no. 1, pp. 241–257.
7. Vavilova, N.B., Vasineva, I.A., Golovan, A.A., Kozlov, A.V., Papusha, I.A., and Parusnikov, N.A., The Calibration Problem in Inertial Navigation, *J. Math. Sci.*, 2021, vol. 253, no. 6, pp. 818–836.
8. Veremeenko, K.K. and Galay, I.A., Calibration Algorithm Development for Inertial Navigation System with Use of Two-Axis Motion Simulator, *Trudy MAI*, 2013, no. 63, pp. 1–14.

9. Tarygin, I.E., Calibration of the Thermal Model of an Inertial Measurement Unit with Three Angular Rate Sensors, *Gyroscopy Navig.*, 2020, vol. 11, pp. 25–33.

10. Kailath, T., Sayed, A.H., and Hassibi, B., *Linear Estimation*, New Jersey: Prentice Hall, 2000.

11. Matasov, A.I., Variational Problems for Calibrating an Accelerometer Unit, *Autom. Remote Control*, 2019, vol. 80, no. 12, pp. 2135–2151.

12. Golovan, A.A. and Parusnikov, N.A., *Matematicheskie osnovy navigatsionnykh sistem, Ch. I: Matematicheskie modeli inertsial'noi navigatsii* (Mathematical Foundations of Navigation Systems, Part I: Mathematical Models of Inertial Navigation), Moscow: MAKS Press, 2011.

13. Boyd, S. and Vandenberghe, L., *Convex Optimization*, Cambridge: Cambridge University Press, 2004.

14. Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Foundations and Trends in Machine Learning*, 2010, vol. 3, no. 1, pp. 1–122.

This paper was recommended for publication by O.A. Stepanov, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 810–822.

CONTROL IN TECHNICAL SYSTEMS

Period-Time Parametric Identification Method for Solving Location and Navigation Tasks

Yu. G. Bulychev[*,a] **and A. A. Mozol**[**,b]

[*] *JSC "Concern Radioelectronic Technologies", Moscow, Russia*
[**] *JSC "VNII "Gradient", Rostov-on-Don, Russia*
e-mail: [a] *profbulychev@yandex.ru,* [b] *amozol@bk.ru*

Received June 16, 2022
Revised May 4, 2023
Accepted May 21, 2023

Abstract—With regard to location and navigation tasks for single-position passive observer, a bearing-free method for identifying parameters of a polynomial model of object motion has been developed taking into account evolution of the discrepancy between the periodic radiated and received quasi-periodic signal. The passage of a signal in an arbitrary physical environment is considered, at the same time, knowledge of the period of the emitted signal and assessing the current Doppler frequency are not required. The method is based on counting the number of periods of the received signal in a given surveillance intertissue. The issues related to the analysis of the resulting discrepancy by the observability of the method and its accuracy characteristics are considered. Useful practical recommendations and an illustrative context are given.

Keywords: radiating target, periodic signal, quasi-periodic signal, single-position passive observer, bearing-free method, time mistie, period-time method, polynomial motion, parametric identification, observability of the method, complete correlation matrix of estimation errors, adaptation

DOI: 10.25728/arcRAS.2023.75.79.001

1. INTRODUCTION

Methods of passive location and navigation of a radiating target based on a single-position passive observer are widely reflected in the well-known literature [1–20]. Among them, doppler-time bearing-free methods are quite popular, operating with periodic signals and geared towards measurement capability the continuous displacement of the doppler frequency the received signal at the observation point caused by target movement (for location tasks) motion of the observer (for navigational tasks); [6] on pp. 169–173 an exhaustive list of literature on this issue is given, and it is available in the open press. In this case, measurements can be implemented at any characteristic frequency from the spectrum of the emitted signal (for example, on the central) or modulating function; as well as by comparing the moments of the arrival of the fronts of consecutive pulses taking into account the known period. These methods are based on the idea of "base synthesis," which ultimately leads to the formation of several observation points on the guidepath and possibilities of using well-known methods of multi-position location and navigation (for example, triangulation, difference-rangefinder, trilateration and their combinations [21, 22]). In this case, as a rule, such path functions are considered, which are either known at the observation site (for example, orbital ones with known motion parameters), or are approximated with sufficient accuracy for practice by a model of straight-line uniform motion (both with known and unknown motion parameters). At the same time, the fundamental point is accounting of information given a priori about the speed of the target or observer, which is often unacceptable for practice.

In [20] the period-time method is developing (PTM), which removes the restriction, related to obtaining of information given a priori about the speed value, and also the question of parametric identification is considered in relation to the model of nonlinear motion, taking into account the possible maneuver of the target or observer. At the same time, a preliminary current estimate of the Doppler frequency is not required, which is equivalent to finding the derivative of time mistie between the periods of the radiated (periodic) and accepted (quasi-periodic) signal. However, the results obtained in [20], apply only to radio signals (spreading as an electromagnetic wave at the speed of light) with a known period, and the dependence of the resulting time mistie on the parameters of the target movement has not been investigate. This article is a further development of the well-known PTM in terms of eliminating these shortcomings in relation to signals, spreading in arbitrary physical environments.

2. PROBLEM STATEMENT

Let the moving RT form in the current t periodic signal $S_0(t)$ (periodic signal $T_S = \text{const}$ may be unknown), spreading in a given physical medium in the form of a wave at a speed of v_S (we can talk about different waves, for example, electromagnetic or acoustic). At the observation point associated with SOPO, at the surveillance intertissue $[0, T]$ a quasi-deterministic signal is received $S(t)$ with a variable period.

According to the PTM, the observation segment is represented as

$$[0, T] = \bigcup_{n=1}^{N} [t_{n-1}, t_n], \quad t_n > t_{n-1}, \quad t_0 = 0, \quad t_N \le T, \tag{2.1}$$

where $t_0 = 0$ is a fixed moment of time corresponding to the beginning of the received signal (for example, the arrival of the first pulse), t_n is a fixed time of receipt $M_n = \sum_{p=1}^{n} \Delta M_p$ periods of the received quasi-periodic signal (ΔM_p — the number of periods counted on the segment $[t_{p-1}, t_p]$), at the same time, at the moment of time t_n number M_n the whole period fits into the segment $[0, t_n]$.

Theoretical and practical issues related to the calculation of these periods, are solved using electronic digital frequency meters and are described in detail in the well-known technical literature [23, pp. 148–161].

At the observation point (where the SOPO is located) taking into account the movement of the RT, the signal becomes quasi-periodic, because there is a time mistie $\delta(t)$ between the periods of the emitted and received signals

$$\delta(t) = v_S^{-1} \Delta_R(t) = v_S^{-1} [R(t) - R_0], \quad t \in [0, T], \tag{2.2}$$

where $R(t)$ — current range to RT, $R_0 = R(0)$ — initial range.

In a rectangular cartesian reference system XYZ (in the center of which there is SOPO) the motion of the RT is described by a polynomial model (to simplify the calculations and clarity of the method instead of a generalized finite polynomial with arbitrary basis functions, we restrict ourselves to a power polynomial of the second degree with an initial condition $\mathbf{r}_0 = \mathbf{r}(0)$, $\|\mathbf{r}_0\| = R_0$)

$$\mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}_0 t + 2^{-1} \mathbf{a}_0 t^2, \quad t \in [0, T], \tag{2.3}$$

where $\mathbf{r}(t) = \mathbf{r} = [x, y, z]^{\mathrm{T}}$ — position vector ($\|\mathbf{r}(t)\| = R(t)$),

$\mathbf{v}_0 = [v_{x0}, v_{y0}, v_{z0}]^{\mathrm{T}}$ — initial velocity vector ($v_0 = \|\mathbf{v}_0\|$ — speed value),

$\mathbf{a}_0 = [a_{x0}, a_{y0}, a_{z0}]^{\mathrm{T}}$ — acceleration vector ($a_0 = \|\mathbf{a}_0\|$ — acceleration value), while the vectors $\mathbf{r}_0$, $\mathbf{v}_0$ and $\mathbf{a}_0$ are a priori unknown.

If we take the value t_n as the measured parameter, then we can use the following vector equation of observation:

$$\mathbf{h} = \mathbf{t} + \boldsymbol{\xi} = \bar{\mathbf{t}} + \boldsymbol{\delta} + \boldsymbol{\xi}, \tag{2.4}$$

where $\mathbf{h} = \left[h_n, n = \overline{1,N}\right]^{\mathrm{T}}$, $\mathbf{t} = \left[t_n, n = \overline{1,N}\right]^{\mathrm{T}}$, $\bar{\mathbf{t}} = \left[\bar{t}_n, n = \overline{1,N}\right]^{\mathrm{T}}$, $\boldsymbol{\delta} = \left[\delta_n, n = \overline{1,N}\right]^{\mathrm{T}}$, $\boldsymbol{\xi} = \left[\xi_n, n = \overline{1,N}\right]^{\mathrm{T}}$, $h_n = h(t_n)$, $\xi_n = \xi(t_n)$.

In (2.4), $\boldsymbol{\xi} = \left[\xi_n, n = \overline{1,N}\right]^{\mathrm{T}}$ is understood as the Gaussian measurement with zero mathematical expectation and the correlation matrix $\mathbf{K}_\xi$, measured parameter t_n connected with number of counted periods by the ratio

$$t_n = M_n T_S + \delta_n = \bar{t}_n + \delta_n = \bar{t}_n + v_S^{-1}\left[R_n - R_0\right], \tag{2.5}$$

where $\delta_n = \delta(t_n)$ is unknown time discrepancy, $\bar{t}_n = M_n T_S$, $R_n = R(t_n)$, $t_0 = 0$.

Formula (2.5) can be commented as follows [6, p. 154]: during the time $\bar{t}_n = M_n T_S$ the distance between the RT and SOPO the range will change by $\Delta R_n = R_n - R_0$, which corresponds to the time mistie $\delta_n = v_S^{-1}\Delta R_n$ between the periods of the emitted and received signals. If the target was stationary or moving in a circle (in the center of which there is SOPO) that range increment would be missing and $\delta_n = 0$ for all n. It is the passage of an additional section of the path length by the wave ΔR_n with speed v_S is the cause of the time mistie δ_n.

Recall that for a known period T_S as a measured parameter, it was possible to take the value $\delta_n = t_n - M_n T_S$ (this is how the observation equation was formed in [6, 20]), for an unknown period T_S only the values t_n and M_n are available for measurement.

If the distance between the RT and the SOPO decreases, then $\delta_n < 0$, otherwise $\delta_n > 0$. The appearance of the time mistie $\delta_n = \delta(t_n)$ is due to the effect of compression or stretching of the initial periodic signal at the observation point due to the movement of the RT.

It is required, taking into account (2.1)–(2.5) to develop a method of parametric identification of RT with a curved (polynomial) movement based on a period-a temporary SOPO that does not require knowledge of the period T_S the emitted signal and the calculation of the current Doppler frequency. The method should include solving the following issues:

— obtaining dependencies that allow us to assess the nature of the evolution of the received signal period (caused by the movement of RT), is fundamental for this method;

— formation of an algorithm for identification of the inclined range and a number of characteristic parameters of the RT movement based on accurate data (taking $\boldsymbol{\xi}_n = 0$, $n = \overline{1,N}$);

— determination of the conditions for the correct application of the method on accurate data (i.e., determination of the observability conditions of the method);

— accounting for random measurement errors;

— solving the identification problem on redundant data ($\mathbf{h}$) taking into account measurement noise (smoothing problem based on the least squares method (PTM)) and obtaining a ratio for calculating the correlation matrix of identification errors;

— conducting a computational experiment to demonstrate the capabilities of the method.

3. INVESTIGATION OF THE EVOLUTION OF THE SIGNAL PERIOD

The foray $\delta(t)$ is described by the expression (for the case of rectilinear uniform motion)

$$\delta(t) = v_S^{-1}\left\{\left[R_0^2 + 2tR_0v_0\cos\gamma_0 + t^2v_0^2\right]^{1/2} - R_0\right\}, \quad t \geq 0, \quad \delta(0) = 0, \tag{3.1}$$

where γ_0 — angle between vectors $\mathbf{r}_0$ and $\mathbf{v}_0$.

For $0 < \gamma_0 \le \pi/2$ the function $\delta(t)$ is non-negative, smooth and strictly convex, $\delta^{(1)}(t) = d\delta(t)/dt = 0$ at the point $t = 0$. Ror $\pi/2 < \gamma_0 < \pi$ the function $\delta(t)$ is smooth and strictly convex, has two roots ($t = 0$ and $t = -2R_0 \cos\gamma_0/v_0$), at the point $t = -R_0 \cos\gamma_0/v_0$ reaches the minimum value ($v_S^{-1} R_0 (\sin\gamma_0 - 1)$). For $\gamma_0 = 0$ we have $\delta(t) = (v_0/v_S)\, t$, the raid is a linear non-negative function independent of R_0. For $\gamma_0 = \pi$ we have $\delta(t) = -(v_0/v_S)\, t$ for $0 \le t \le R_0/v_0$, so $\delta(t)$ is a linear function and reaches its minimum $(-R_0/v_S)$ at the point $t = R_0/v_0$. Since for $\gamma_0 = 0$ and $\gamma_0 = \pi$ the foray $\delta(t)$ does not depend on R_0, then, for these incorrect cases associated with the movement of the RT along the line of sight, it is impossible to determine the range taking into account the evolution of the signal period at the observation point.

For a more detailed study of $\delta(t)$ we will find the first few occurrences in time (at the point $t = 0$):

$$\begin{cases} \delta_0^{(1)} = v_S^{-1} v_R, \\ \delta_0^{(2)} = (v_S R_0)^{-1} v_\tau^2, \\ \delta_0^{(3)} = -3 \left(v_S R_0^2\right)^{-1} v_\tau^2 v_R, \end{cases} \tag{3.2}$$

where $v_R = R_0 \cos\gamma_0$ and $v_\tau = v_0 \sin\gamma_0$ — respectively, the values of the radial and tangential velocity.

As a result, you can use the decomposition based on the Taylor series

$$\delta(t) = v_S^{-1} t \left(v_R + \frac{v_\tau^2 t}{2R_0} - \frac{v_\tau^2 v_R t^2}{2R_0^2} + \dots \right) = v_S^{-1} t \left[v_R + \frac{v_\tau^2 t}{2R_0} \left(1 - \frac{v_R t}{R_0} \right) + \dots \right], \tag{3.3}$$

from which it follows that the spectral composition of the function $\delta(t)$ significantly depends on the observation conditions, and in many practically important cases it is not possible to neglect derivatives of the second and higher orders, especially for long observation intervals and short ranges.

Formulas (3.1)–(3.3) are very useful in substantiating the possibility of practical implementation of the developed PTM in each specific case, taking into account the accepted initial data.

4. BUILDING A PARAMETRIC IDENTIFICATION ALGORITHM BASED ON ACCURATE DATA

Taking into account (2.3) we can use the following dependency

$$R^2(t) - R_0^2 = 2t \langle \mathbf{r}_0, \mathbf{v}_0 \rangle + t^2 \left(v_0^2 + \langle \mathbf{r}_0, \mathbf{a}_0 \rangle \right) + t^3 \langle \mathbf{v}_0, \mathbf{a}_0 \rangle + 4^{-1} t^4 a_0^2, \tag{4.1}$$

where $\langle \cdot, \cdot \rangle$ is the symbol of the scalar product of two vectors, $\|\cdot\|$ is the symbol of the vector norm.

Formula (4.1) represents the first basic ratio of the developed PTM.

The second basic relation follows directly from the formula (2.2):

$$R^2(t) - R_0^2 = 2 v_S R_0 \delta(t) + v_S^2 \delta^2(t). \tag{4.2}$$

Equating expressions (4.1) and (4.2), after simple transformations we obtain the equation

$$-2 v_S \delta(t) \chi_1 + 2t \chi_2 + t^2 \chi_3 + t^3 \chi_4 + 4^{-1} t^4 \chi_5 = v_S^2 \delta^2(t), \tag{4.3}$$

where

$$\begin{cases} \chi_1 = R_0, \\ \chi_2 = \langle \mathbf{r}_0, \mathbf{v}_0 \rangle, \\ \chi_3 = \left(v_0^2 + \langle \mathbf{r}_0, \mathbf{a}_0 \rangle\right), \\ \chi_4 = \langle \mathbf{v}_0, \mathbf{a}_0 \rangle, \\ \chi_5 = a_0^2 \end{cases} \tag{4.4}$$

— unknown fraction that have a clear physical meaning and are subject to identification.

Since the values of δ_n are unknown, then, taking into account (2.5) for discrete time, we write down the equation with respect to unknown quantities T_S and χ_i, $i = \overline{1,5}$:

$$-2v_S\left(t_n - M_n T_S\right)\chi_1 + 2t_n\chi_2 + t_n^2\chi_3 + t_n^3\chi_4 + 4^{-1}t_n^4\chi_5 = v_S^2\left[\left(t_n - M_n T_S\right)\right]^2. \tag{4.5}$$

After simple but cumbersome transformations, formula (4.5) can be represented as a new equation (relative to the coefficients A_i)

$$\sum_{i=1}^{6} B_{in}A_i = D_n, \tag{4.6}$$

where

$$\begin{cases} A_1 = (v_S\chi_1 - \chi_2)v_S^{-2}T_S^{-1}, \quad A_2 = -\chi_1 v_S^{-1}, \\ A_3 = \left(v_S^2 - \chi_3\right)\left(2v_S^2T_S\right)^{-1}, \quad A_4 = 2^{-1}T_S, \\ A_5 = -\chi_4\left(2T_Sv_S^2\right)^{-1}, \quad A_6 = \chi_5\left(8T_Sv_S^2\right)^{-1}, \\ B_{1n} = t_n, \quad B_{2n} = M_n, \quad B_{3n} = t_n^2, \\ B_{4n} = M_n^2, \quad B_{5n} = t_n^3, \quad B_{6n} = t_n^4, \\ D_n = M_n t_n. \end{cases} \tag{4.7}$$

The relations (4.6) and (4.7) are the basis for identifying the parameters of the curvilinear motion of the RT at an unknown period of the emitted signal. In (4.6) the unknown coefficients are A_i, $i = \overline{1,6}$, those that are uniquely related to the desired parameters of the motion of the RT and the period of the emitted signal. If Eq. (4.6) $n = \overline{1,N}$, where $N \geq 6$, then we get a system of linear algebraic equations (SLAE) (with a rectangular matrix $\mathbf{B}$)

$$\mathbf{BA} = \mathbf{D}, \tag{4.8}$$

where $\mathbf{B} = \left[B_{in}, n = \overline{1,N}, i = \overline{1,6}\right]$, $\mathbf{A} = \left[a_i, i = \overline{1,6}\right]^{\mathrm{T}}$, $\mathbf{D} = \left[D_n, n = \overline{1,N}\right]^{\mathrm{T}}$.

This SLAE allows us to solve the problem of estimating these coefficients and parameters, as well as the signal period for redundant measurements. For $N > 6$ we are talking about the problem of smoothing based on OLS using orthogonal-singular decomposition [24].

Consider a special case when the RT moves rectilinearly and uniformly, and the signal period is unknown. Now instead of (4.3) we have the equation

$$-2v_S\delta(t)\chi_1 + 2t\chi_2 + t^2\chi_3 = v_S^2\delta^2(t), \tag{4.9}$$

where

$$\begin{cases} \chi_1 = R_0, \\ \chi_2 = \langle \mathbf{r}_0, \mathbf{v}_0 \rangle, \\ \chi_3 = v_0^2. \end{cases} \tag{4.10}$$

In this case, instead of (4.6) we have

$$t_n A_1 + M_n A_2 + t_n^2 A_3 + M_n^2 A_4 = M_n t_n. \tag{4.11}$$

If we assume that the signal period is known, i.e., the values are known δ_n, then, taking into account (4.9) to find the parameters of a rectilinear uniform motion of the RT, it is sufficient to solve the SLAE (regarding $\chi_i,\ i = \overline{1,3}$)

$$-2v_S\delta_n\chi_1 + 2t_n\chi_2 + t^2\chi_3 = v_S^2\delta_n^2, \quad n = \overline{1,N}. \tag{4.12}$$

At the same time, we find the range R_0, the speed value $v_0 = \|\mathbf{v}_0\|$ and the angle γ_0 between the vectors $\mathbf{r}_0$ and $\mathbf{v}_0$ taking into account the obvious relations:

$$\begin{cases} R_0 = \chi_1, \\ v_0 = \sqrt{\chi_3}, \\ \gamma_0 = \arccos\left[\chi_2 \left(R_0 v_0\right)^{-1}\right]. \end{cases} \tag{4.13}$$

In the case of rectilinear equidistant motion of the RT (when the vectors $\mathbf{v}_0$ and $\mathbf{a}_0$ are collinear) it is necessary to solve the SLAE (regarding χ_i, $i = \overline{1,5}$)

$$-2v_S\delta_n\chi_1 + 2t_n\chi_2 + t_n^2\chi_3 + t_n^3\chi_4 + 4^{-1}t_n^4\chi_5 = v_S^2\delta_n^2. \tag{4.14}$$

Now we have

$$\begin{cases} \chi_1 = R_0, \\ \chi_2 = R_0 v_0 \cos\gamma_0, \\ \chi_3 = \left(v_0^2 + R_0 a_0 \cos\gamma_0\right), \\ \chi_4 = v_0 a_0, \\ \chi_5 = a_0^2. \end{cases} \tag{4.15}$$

Based on the found values $\chi_1, \ldots, \chi_5$ we calculate the following parameters of the movement of the RT:

$$\begin{cases} R_0 = \chi_1, \\ a_0 = \sqrt{\chi_5}, \\ v_0 = \chi_4 a_0^{-1}, \\ \gamma_0 = \arccos\left[\chi_2 \left(R_0 v_0\right)^{-1}\right]. \end{cases} \tag{4.16}$$

Expressions (4.1)–(4.16) form the mathematical basis of the developed PTM.

In the next section we will analyze the observability conditions of the developed method, i.e., we will identify situations in which it becomes incorrect from a computational point of view.

5. ANALYSIS OF THE OBSERVABILITY OF THE METHOD

The developed PTM can be implemented on any set of nodes from the set $\{t_1, \ldots, t_N\}$, which allows not only to reduce the amount of calculations, but also in some cases to increase the reliability of the generated estimates (especially in the absence of reliable a priori information about the weighting factors necessary for the implementation of LSM). To do this, we introduce vectors of temporary nodes $\mathbf{t}_{[l]} = \left[t_{[l]p}, p = \overline{1, P_{[l]}}\right]^{\mathrm{T}}$, where $l = \overline{1,L}$, $t_{[l]p} \in \{t_1, \ldots, t_N\}$, $t_{[l]p+1} > t_{[l]p}$. Here L is

the number of sets, $P_{[l]}$ — the number of nodes in the l set, $t_{[l]p}$ is the node with number $[l]\,p$ (this is a natural number belonging to the set $\{1,\dots,N\}$). Based on (4.12) we will form the following SLAE:

$$\mathbf{C}_{[l]}\boldsymbol{\chi}_{[l]} = \mathbf{Y}_{[l]}, \tag{5.1}$$

where $\mathbf{Y}_{[l]} = \left[\delta^2_{[l]p}, p = \overline{1,P_{[l]}}\right]^{\mathrm{T}}$, $\boldsymbol{\chi}_{[l]} = \left[\chi_{i[l]}, i = \overline{1,5}\right]^{T}$, and the matrix $\mathbf{C}_{[l]}$ (size $P_{[l]} \times 5$) is formed by strings $v_S^{-2}\left(-2v_S\delta_{[l]p},\, 2t_{[l]p},\, t^2_{[l]p},\, t^3_{[l]p},\, 4^{-1}t^4_{[l]p}\right)$, $p = \overline{1,P_{[l]}}$.

The introduction of $\mathbf{t}_{[l]}$ makes it possible to find such sets of nodes taking into account the observation geometry, the characteristics of the RT and the SOPN, to find such sets of nodes on which the identification issue is solved most qualitatively (this refers to the well-known problem of experiment planning [25]).

Without reducing the generality of reasoning, we will limit ourselves to the flat case (assuming $z = 0$) and a signal with a known period, and also, we will ask $P_{[l]} = 5$, what corresponds to a square matrix $\mathbf{C}_{[l]}$. It is obvious that for the correct application of the developed method, related to the SLAE solution (5.1), it is necessary and sufficient to fulfill the condition $\det\mathbf{C}_{[l]} \neq 0$, what leads to the desired result $\boldsymbol{\chi}_{[l]} = \mathbf{C}^{-1}_{[l]}\mathbf{Y}_{[l]}$. To identify cases in which this condition is violated, we write down the columns of the matrix $\mathbf{C}_{[l]}$ in the form of vectors:

$$\mathbf{C}_{[l]1} = \left[-2v_S\delta_{[l]p}, p = \overline{1,5}\right]^{\mathrm{T}}, \quad \mathbf{C}_{[l]2} = \left[2t_{[l]p}, p = \overline{1,5}\right]^{\mathrm{T}},$$
$$\mathbf{C}_{[l]3} = \left[t^2_{[l]p}, p = \overline{1,5}\right]^{\mathrm{T}}, \quad \mathbf{C}_{[l]4} = \left[t^3_{[l]p}, p = \overline{1,5}\right]^{\mathrm{T}}, \quad \mathbf{C}_{[l]5} = \left[4^{-1}t^4_{[l]p}, p = \overline{1,5}\right]^{\mathrm{T}}.$$

It is light to notice that the columns $\mathbf{C}_{[l]2}$, $\mathbf{C}_{[l]3}$ and $\mathbf{C}_{[l]4}$ are linearly independent, therefore, to check the condition $\det\mathbf{C}_{[l]} \neq 0$ it is enough to show that the column $\mathbf{C}_{[l]1}$ cannot be represented as a linear combination of these columns.

Since $R_{[l]p} = \left[x^2_{[l]p} + y^2_{[l]p}\right]^{-2}$ $\left(\text{where } R_{[l]p} = R\left(t_{[l]p}\right),\ x^2_{[l]p} = \left(x_0 + v_{x0}t_{[l]p} + 2^{-1}a_{x0}t^2_{[l]p}\right)^2 \text{ and } y^2_{[l]p} = \left(y_0 + v_{y\,0}t_{[l]p} + 2^{-1}a_{y0}t^2_{[l]p}\right)^2\right)$, that violation of the condition $\det\mathbf{C}_{[l]} \neq 0$ it is equivalent to the fact that the vectors $\boldsymbol{\mu}_{[l]} = \left[x^2_{[l]p}, p = \overline{1,5}\right]^{\mathrm{T}}$ and $\boldsymbol{\eta}_{[l]} = \left[y^2_{[l]p}, p = \overline{1,5}\right]^{\mathrm{T}}$ are not bound by the collinearity condition: $\boldsymbol{\mu}_{[l]} = k\boldsymbol{\eta}_{[l]}$, where k — where is the proportionality coefficient. Otherwise we have

$$R_{[l]p} = \left[x^2_{[l]p} + y^2_{[l]p}\right]^{-2} = \left[k^2y^2_{[l]p} + y^2_{[l]p}\right]^{-2} = q\left|y_{[l]p}\right|, \tag{5.2}$$
$$-2v_S\delta_{[l]p} = -2\left[R_{[l]p} - R_0\right] = -2\left[q\left|y_{[l]p}\right| - R_0\right], \tag{5.3}$$

where $q = \left(k^2+1\right)^{-2}$.

It follows from (5.2) and (5.3) that the coordinates of vector $\mathbf{C}_{[l]1}$ can be represented by a linear combination of the coordinates of vectors $\mathbf{C}_{[l]2}$, $\mathbf{C}_{[l]3}$ and $\mathbf{C}_{[l]4}$. The physical meaning of the condition $\boldsymbol{\mu}_{[l]} = k\boldsymbol{\eta}_{[l]}$ (the condition of computational incorrectness of the method) is that the RT moves rectilinearly along the line of sight SOPN.

Thus, for the correctness of the method, it is necessary to exclude cases when the RT moves along the specified line or in its vicinity. This imposes certain restrictions on the conditions for monitoring RT, which must be provided for in practice.

If we limit ourselves to the model of rectilinear uniform motion and a signal with a known period in this case, $\left(\text{in (5.1) we must put } p = \overline{1,3} \text{ and } \mathbf{t}_{[l]} = \left[t_{[l]1}, t_{[l]2}, t_{[l]3}\right]^{\mathrm{T}}\right)$, the solution of SLAE (5.1)

with the correct application of the method, allows us to determine the desired parameters of the motion of the RT

$$\begin{cases} R_{0[l]} = 2^{-1} v_S \left(\dfrac{\delta^2_{[l]1}\Delta^t_{[l]23} - \delta^2_{[l]2}\Delta^t_{[l]13} + \delta^2_{[l]3}\Delta^t_{[l]12}}{-\delta_{[l]1}\Delta^t_{[l]23} + \delta_{[l]2}\Delta^t_{[l]13} - \delta_{[l]3}\Delta^t_{[l]12}} \right), \\ \langle \mathbf{r}_0, \mathbf{v}_0 \rangle_{[l]} = 2^{-1} v_S^2 \left(\dfrac{t^2_{[l]1}\Delta^\delta_{[l]23} - t^2_{[l]2}\Delta^\delta_{[l]13} + t^2_{[l]3}\Delta^\delta_{[l]12}}{-\delta_{[l]1}\Delta^t_{[l]23} + \delta_{[l]2}\Delta^t_{[l]13} - \delta_{[l]3}\Delta^t_{[l]12}} \right), \\ v_{0[l]} = \left[\dfrac{t_{[l]3}\Delta^\delta_{[l]12} - t_{[l]2}\Delta^\delta_{[l]13} + t_{[l]1}\Delta^\delta_{[l]23}}{\delta_{[l]1}\Delta^t_{[l]23} - \delta_{[l]2}\Delta^t_{[l]13} + \delta_{[l]3}\Delta^t_{[l]12}} \right]^{1/2}, \\ \gamma_{0[l]} = \arccos \left[\dfrac{\langle \mathbf{r}_0, \mathbf{v}_0 \rangle_{[l]}}{R_{0[l]} v_{0[l]}} \right], \end{cases} \tag{5.4}$$

where $\Delta^t_{[l]12} = t_{[l]1} t_{[l]2} \left(t_{[l]1} - t_{[l]2} \right)$, $\Delta^\delta_{[l]12} = \delta_{[l]1}\delta_{[l]2} \left(\delta_{[l]1} - \delta_{[l]2} \right)$ and, if you do not take into account measurement and calculation errors, $R_{0[l]} = R_0$, $v_{0[l]} = v_0$, $\langle \mathbf{r}_0, \mathbf{v}_0 \rangle_{[l]} = \langle \mathbf{r}_0, \mathbf{v}_0 \rangle$, $\gamma_{0[l]} = \gamma_0$.

Therefore, it becomes possible to determine the motion parameters R_0, v_0 and γ_0 $\Big($where $R_0 = \chi_1$, $v_0 = \sqrt{\chi_3}$, $\gamma_0 = \arccos \left[\chi_2 \left(R_0 v_0 \right)^{-1} \right] \Big)$, without resorting to the numerical solution of SLAE, which is an undoubted advantage of the developed PTM.

6. ACCOUNTING FOR RANDOM MEASUREMENT ERRORS

Assuming the signal period is known, we use the traditional procedure for calculating the elements of the correlation matrix to assess the effect of random measurement errors on the accuracy characteristics of the method $\mathbf{K}_{\chi[l]}$ errors in estimating the coordinates of the vector $\boldsymbol{\chi}$ in linear approximation [26]. To do this, taking into account SLAE (5.1) (assuming for simplicity the matrix $\mathbf{C}_{[l]}$ square size 5×5) let's use the representation $\boldsymbol{\chi}_{[l]} = \mathbf{C}^{-1}_{[l]} \mathbf{Y}_{[l]} = \left[\chi_k \left(\boldsymbol{\delta}_{[l]} \right), k = \overline{1,5} \right]^{\mathrm{T}}$ (where $\boldsymbol{\delta}_{[l]} = \left[\delta_{[l]p}, p = \overline{1,5} \right]^{\mathrm{T}}$) and partial derivatives of the following form: $\partial \chi_{k[l]} \left(\boldsymbol{\delta}_{[l]} \right) / \partial \delta_{[l]p}$. The correlation matrix is found by the rule

$$\mathbf{K}_{\chi[l]} = \mathbf{F}_{\chi[l]} \mathbf{K}_\xi \mathbf{F}^{\mathrm{T}}_{\chi[l]}, \tag{6.1}$$

where $\mathbf{F}_{\chi[l]} = \left[\partial \chi_{k[l]} \left(\boldsymbol{\delta}_{[l]} \right) / \partial \delta_{[l]p}, \; k = \overline{1,5}, \; p = \overline{1,5} \right]$.

Expression (6.1) allows a priori, based on the mathematical expectations of the measured parameters, to assess the potential capabilities of the developed PTM and develop practical recommendations for its best use under specific conditions of observation of RT, and also reasonably approach the choice of the main parameters of the method (the length of the observation interval (T), the number of nodes (N) and time sets ($\mathbf{t}_{[l]}$)). So, the number $l^* \in \{1, \ldots, L\}$ of optimal set $\boldsymbol{\delta}_{[l^*]}$, ensuring the minimization of the estimation error, is found according to the following adaptive rule:

$$l^* = \arg\min_l \left\| \mathbf{K}_{\chi[l]} \right\|, \tag{6.2}$$

where $\left\| \mathbf{K}_{\chi[l]} \right\|$ — this is any of the norms of the matrix $\mathbf{K}_{\chi[l]}$, used in evaluation tasks.

In the practical implementation of the developed PTM, the factor should be taken into account that for large values of v_S (for example, when $v_S = c$, where c is the speed of light), the solution of

the square SLA (4.8) in the presence of random measurement errors can lead to incorrect results. Let us explain this fact for the case $N = 6$ by the example of calculating the velocity v_0. Because $v_0 = c\sqrt{1 - 2T_S A_3}$, that's a mistake $\Delta_3 = \hat{A}_3 - A_3$ (where $\hat{A}_3$ — calculated coefficient value A_3 by solving SLAE (4.8) taking into account measurement errors) leads to the following speed estimate: $\hat{v}_0 = \sqrt{v_0^2 + 2c^2 T_S \Delta_3}$. That is, a correct assessment of the speed is possible only if the condition is met $\Delta_3 > -v_0^2 \left(2c^2 T_S\right)^{-1}$, which imposes a very strict restriction on the magnitude of the error Δ_3. This effect also applies to all SLOUGH coefficients (4.8), except A_2 and A_4.

To overcome this incorrectness (at high v_S speeds), a two-step approach to identification is recommended. At the first stage, SLAE is solved (4.8), of which only an assessment will be required $\hat{A}_4$ for A_4. This allows you to form the desired estimate $\hat{T}_S = 2\hat{A}_4$ for period T_S, and based on it, estimates for residuals $\hat{\delta}_n = t_n - M_n \hat{T}_S$. All estimates of the parameters of the RT movement are based on the SLAE (5.1), in which the value of δ_n is substituted instead of $\hat{\delta}_n$.

7. ACCOUNTING FOR REDUNDANT MEASUREMENTS

Now consider the case of redundant measurements when the matrix $\mathbf{C}_{[l]}$ and the vector $\mathbf{Y}_{[l]}$ in (5.1) have an arbitrary number of lines $P_{[l]} \le N$, which, as a rule, significantly exceeds the number of estimated parameters. To simplify the calculations, we will consider in SLAU (5.1) the component. $\mathbf{Y}_{[l]} = \left[v_S^2 \delta^2_{[l]p}, p = \overline{1, P_{[l]}}\right]^{\mathrm{T}}$ as a vector of secondary measured parameters $h_{[l]1}, \ldots, h_{[l]P_{[l]}}$ and primary measurements (5.1) the correlation matrix of measurement errors of the coordinates of the vector $\mathbf{Y}_{[l]}$ we can imagine it like this

$$\mathbf{K}_{\mathbf{Y}[l]} = \mathbf{F}_{\boldsymbol{\delta}[l]} \mathbf{K}_{\boldsymbol{\xi}} \mathbf{F}^{\mathrm{T}}_{\boldsymbol{\delta}[l]}. \tag{7.1}$$

Assuming that the matrix $\mathbf{K}_{\xi}$ is diagonal, we have $\mathbf{K}_{Y[l]} = \operatorname{diag}\left[4\delta^2_{[l]1}, 4\delta^2_{[l]2}, \ldots, 4\delta^2_{[l]P_{[l]}}\right]$. Under the condition of sufficiently small measurement errors, the least squares method can be used to construct a smoothed estimate of the vector $\boldsymbol{\chi}$ [25]

$$\boldsymbol{\chi}^*_{[l]} = \left(\mathbf{C}^{\mathrm{T}}_{[l]} \mathbf{K}^{-1}_{Y[l]} \mathbf{C}_{[l]}\right)^{-1} \mathbf{C}^{\mathrm{T}}_{[l]} \mathbf{K}^{-1}_{\mathbf{Y}[l]} \mathbf{h}_{\mathbf{Y}[l]}, \tag{7.2}$$

where $\mathbf{h}_{\mathbf{Y}[l]} = \left[h_{\mathbf{Y}[l]p}, p = \overline{1, P_{[l]}}\right]^{\mathrm{T}}$ — the vector of secondary measurements.

The correlation matrix of estimation errors is found as follows:

$$\mathbf{K}_{\boldsymbol{\chi}^*_{[l]}} = \left(\mathbf{C}^{\mathrm{T}}_{[l]} \mathbf{K}^{-1}_{\mathbf{Y}[l]} \mathbf{C}_{[l]}\right)^{-1}. \tag{7.3}$$

To select the optimal set with a number $l^* \in \{1, \ldots, L\}$ we use an adaptive algorithm of type (6.2).

It should be noted that the approach (7.1)–(7.3) is not strictly optimal, since the elements of the matrix $\mathbf{C}_{[l]}$ depend on the results of observations. But with certain limitations on measurement errors, it gives a completely acceptable result.

For more accurate smoothing, well-known nonlinear optimal estimation procedures can be used, which in practice lead to time-consuming recurrent computational algorithms involving the setting of a sufficiently high-quality initial condition.

Another simplest and fairly reliable way to construct a smooth estimate of $\boldsymbol{\chi}^*_{[l]}$ is to pre-smooth the primary measurements $h_{[l]1}, \ldots, h_{[l]P_{[l]}}$ by the corresponding polynomial $\delta^*_{[l]}(t)$ and the application of the results obtained to the solution of SLAE (5.1). In addition, you can find a smoothed range estimate for any $t \in [0, T]$, exactly,

$$R_{[l]}(t) = R^*_{0[l]} + c\delta^*_{[l]}(t). \tag{7.4}$$

Here we take a set with a number as the optimal one $l = l^* \in \{1, \ldots, L\}$.

8. SOME GENERALIZATIONS AND PRACTICAL RECOMMENDATIONS

The case of estimating the initial range was considered above $R_0 = R(0)$ for time $t = 0$. However, if the Taylor series used to describe the curvilinear motion of the RT is written with respect not to the initial, but to any arbitrary $t = t_* \in [0, T]$, then, by analogy with the above, it is possible to solve the identification problem precisely for the moment of time t^*, in particular, to find the range $R_* = R(t_*)$.

The developed method is easy to implement in the form of the following algorithms: by sampling an increasing volume, on a "sliding grid" or in the form of a filter [25]. At the same time, the movement of RT in the observation interval can be considered as piecewise polynomial (in [20] it was considered as piecewise linear).

During the practical implementation of the method, questions arise (for example, the choice of the degree of the polynomial describing the motion of the RT or the number of counted pulses) related to the organization of the measuring experiment. [25] provides practical recommendations for solving these issues in full. It is obvious that the developed method is most effective when it comes to large distances traveled (i.e., a base of sufficient size is "synthesized"), and this sets certain restrictions on the type of RT (in particular, on his speed, maneuverability, etc.), on the adequacy of the polynomial used at a given observation interval and on the technical characteristics of the SOPN.

For cases related to the movement of RT along the line of sight, a hybrid variant of using the developed and well-known energy method can be proposed [27]. It is proved that this method, operating with the relative level of the received signal, implements its potential capabilities when moving RT along the line of sight. In a sense, the developed and energetic methods are "orthogonal" to each other in terms of accuracy. Therefore, by combining these methods, it is possible to align the working area of the hybrid method and achieve acceptable accuracy characteristics for various conditions of observation of the RT.

For a more effective application of the energy method, clustering and majority processing procedures should be used to reduce and eliminate unreliable measurements.

9. ILLUSTRATIVE EXAMPLE

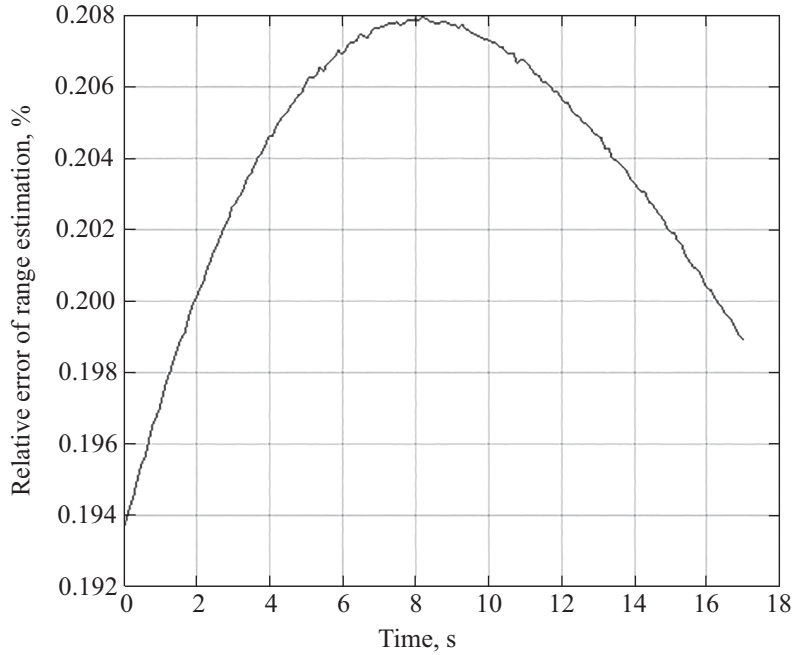
Suppose that the RT carries out a planar movement $x(t) = x_0 + v_{x0}t$, $y(t) = y_0 + v_{y0}t$, where $x_0 = y_0 = 11 \times 10^3$, $v_{x0} = -5 \times 10^2$, $v_{y0} = 6 \times 10^2$, $\gamma_0 = 85$. Here and further, the time and measurement errors of time intervals are set in seconds (s), coordinates and range — in meters (m), speed — in m/s, acceleration — in m/s^2, frequency — in hertz (Hz), angle — in degrees, relative error — as a percentage.

The RT generates a pulsed radio signal

$$S_0(t) = \sum_{k=1}^{K} \operatorname{rect}\left[(t - kT_S)\,\tau^{-1}\right] \cos(2\pi f_0 t),$$

where $T_S = 10^{-2}$, $\tau = 10^{-5}$, $f_0 = 10^{10}$. Parameters of the SOPN operation: $T = 18$, $v_S = c = 3 \times 10^8$, $L = 1$ (that is, one single set of nodes is used), $P_{[1]} = 4$ (set size), $\Delta M_p = \Delta M = 10$, $\mathbf{K}_\xi = \operatorname{diag}\left[\sigma^2, \ldots, \sigma^2\right]$, at the same time, the measurement errors of the time position of the pulse fronts were assumed to be uncorrelated and were set according to the normal distribution law with zero mathematical expectation and the value of the standard deviation $\sigma = 10^{-9}$.

The method was implemented in two stages using a random number sensor and averaging over a thousand experiments. At the first stage, SLAE (4.8) was solved with a square matrix $\mathbf{B}$ of size 4×4



Relative error of range estimation.

(since the RT with zero acceleration is considered), at the same time, a vector is used to calculate the elements of matrix $\mathbf{B}$ and column $\mathbf{D}$ $\mathbf{t}_{[1]} = \bar{\mathbf{t}}_{[1]} + \boldsymbol{\delta}_{[1]} = \left[t_{[1]p}, p = \overline{1,4}\right]^{\mathrm{T}}$ with node numbers: $[1]1 = 12$, $[1]2 = 65$, $[1]3 = 118$, $[1]4 = 171$, $\bar{\mathbf{t}}_{[1]} = \left[\bar{\mathbf{t}}_{[1]p}, p = \overline{1,4}\right]^{\mathrm{T}} = \left[[1]p \times 10^{-1}, p = \overline{1,4}\right]^{\mathrm{T}}$. From all four estimates of unknown coefficients, only the estimate of the signal period is selected $\hat{T}_{S[1]} = 9.999999731646 \times 10^{-3}$ (obtained based on the set $\mathbf{t}_{[1]} = \bar{\mathbf{t}}_{[1]} + \boldsymbol{\delta}_{[1]}$), which corresponds to the relative error $\delta T_{S[1]} = 2.683540941544882 \times 10^{-6}$.

At the second stage, taking into account $\hat{\delta}_{[1]p} = t_{[1]p} - M_{[1]p}\hat{T}_{S[1]} = t_{[1]p} - [1]\,p\Delta M\hat{T}_{S[1]}$ The SLAE (5.1) was solved with a square matrix $\mathbf{C}_{[l]}$ (3×3 in size), with $\chi_4 = \chi_5 = 0$ and a set was used $\bar{\mathbf{t}}_{[1]} = \left[\bar{t}_{[1]p}, p = \overline{1,3}\right]^{\mathrm{T}} = [1.1;\ 9.1;\ 17.1]^{\mathrm{T}}$. The matrix itself is formed by strings $c^{-2}\left(-2c\hat{\delta}_{[1]p},\ 2t_{[1]p},\ t^2_{[1]p}\right)$, $p = \overline{1,3}$. As a result of the true range $R_0 = 1.555634918 \times 10^4$ match rating $\hat{R}_{0[1]} = 1.559672203 \times 10^4$ measure of inaccuracy $\Delta R_{0[1]} = 0.259526489$, true speed $v_0 = 7.810249675 \times 10^2$ — assessed value $v_{0[1]} = 7.821417156 \times 10^2$ measure of inaccuracy $\Delta v_{0[1]} = 0.142984942$, true angle $\gamma_0 = 84.805571092$ — assessed value $\gamma_{0[1]} = 84.761511501$ measure of inaccuracy $\Delta\gamma_{0[1]} = 0.051953650$.

The figure shows a graph of the dependence of the relative error of the range estimation, obtained taking into account (7.4).

For more effective use of the method developed in the article, the question of choosing the size of the observation interval and the nodes of the time grid is, as well as their coordination with the dynamics of the RT movement and the magnitude of measurement errors should be solved in the optimization formulation. When solving SLAE, well-known regularization methods should be used. The results of the numerical experiment show that the greater the distance between the nodes of the time grid used, the less influence random measurement errors have on the resulting estimation accuracy. This distance must be consistent with the dynamics of the RT, namely: the lower the speed of movement of the RT, the greater the step of this grid and the duration of the observation interval should be.

10. CONCLUSION

The developed PTM makes it possible to identify a model of curvilinear polynomial motion of the RT based on the results of recording the time discrepancy between the periods of the emitted signal and the same periods, calculated at the observation point. The method does not require knowledge of the signal period and a preliminary estimate of the current Doppler frequency, as well as knowledge of any a priori data on the parameters of the accepted motion model of the RT. The observability and the main limitations of the method, the conditions for its most effective application are investigated. Analytical relations are obtained that allow us to estimate the evolution of the time discrepancy taking into account the characteristics of the RT and the SOPO, as well as the accuracy characteristics of the method for various observation conditions.

The method can be implemented in various ways: by a fixed sample of measurements, by a sample of measurements of increasing volume, in the form of a dynamic filtration algorithm (linear, quasi-linear or nonlinear), etc.

The method can be implemented either independently or as part of a hybrid method, combining other well-known approaches of passive single-position and multi-position location and navigation of RT. Since the developed method allows you to determine the range, it can be used in rangefinder-rangefinder systems of multi-position location when solving the well-known trilateration problem [21, 22].

If there are not only fluctuation errors in the period-time measurements, but also singular errors, it is advisable to initially subject these measurements to the procedure of generalized invariant-unbiased estimation [28], compensation for these errors, achieving the smoothing effect and optimal estimation of various numerical characteristics (linear functionals, e.g., derivatives, integrals, spectral coefficients, etc.), useful not only for improving the computational stability of the method, but also for evaluating its effectiveness. To solve the SLAE using the regularization procedure, a well-known approach can be applied [29].

REFERENCES

1. *Osnovy manevrirovaniya korablei* (Fundamentals of Ship Maneuvering), Skvortsov, M., Ed., Moscow: Voenizdat, 1966.

2. Shebshaevich, V.S., *Vvedenie v teoriyu kosmicheskoi navigatsii* (Introduction to the Theory of Space Navigation), Moscow: Sovetskoe Radio, 1971.

3. Gromov, G.N., *Differentsial'no-geometricheskii metod navigatsii* (Differential Geometric Method of Navigation), Moscow: Radio i Svyaz', 1986.

4. Khvoshch, V.A., *Taktika podvodnyh lodok* (Tactics of Submarines), Moscow: Voenizdat, 1989.

5. Solov'ev, Yu.A., *Sputnikovaya navigatsiya i ee prilozheniya* (Satellite Navigation and Its Applications), Moscow: Ekotrends, 2003.

6. Mel'nikov, Yu.P. and Popov, S.V., *Radiotekhnicheskaya razvedka* (Radio Engineering Intelligence), Moscow: Radiotekhnika, 2008.

7. Yarlykov, M.S., *Statisticheskaya teoriya radionavigatsii* (Statistical Theory of Radio Navigation), Moscow: Radio i svyaz', 1985.

8. Sosulin, Yu.G., Kostrov, V.V., and Parshin, Yu.N., *Otsenochno-korrelyatsionnaya obrabotka signalov i kompensatsiya pomekh* (Evaluation and Correlation Signal Processing and Interference Compensation), Moscow: Radiotekhnika, 2014.

9. Bulychev, Yu.G. and Manin, A.P., *Matematicheskie aspekty opredeleniya dvizheniya letatel'nykh apparatov* (Mathematical Aspects of Determining the Motion of Aircraft), Moscow: Mashinostroenie, 2000.

10. Bulychev, Yu.G., Vasil'ev, V.V., Dzhugan, R.V., et al., *Informatsionno–izmeritel'noe obespechenie naturnykh ispytanii slozhnykh tekhnicheskikh kompleksov* (Information and Measurement for Live Testing of Complex Technical Systems), Manin, A.P. and Vasil'ev, V.V., Eds., Moscow: Mashinostroenie–Polet, 2016.

11. Gel'tser, A.A, A One-Position Method for Determining the Location of a Radio Emission Source Using Signal Reflections from a Variety of Relief Elements and Local Objects, *Cand. Sci. Dissertation*, Tom. Gos. Univ. Sistem Upravlen. Radioelektroniki, 2012.

12. Sirenko, I.L., Donets, I.V., Reisenkind, Ya.A., et al., Single-Position Determination of Coordinates and Velocity Vector of Radio-Emitting Objects, *Radiotekhnika*, 2019, no.10 (16), pp. 28–32.

13. Bulychev, Yu.G., Bulychev, V.Yu., Ivakina, S.S., and Nasenkov, I.G., Passive Location of a Group of Moving Targets with One Stationary Bearing with Prior Information, *Autom. Remote Control*, 2017, vol. 78, no. 1, pp. 125–137.

14. Bulychev, Yu.G., Bulychev, V.Yu., Ivakina, S.S., and Nicholas, P.I, Estimation of the Inclined Range to the Target with the Polynomial Law of Motion, *Vestn. Kazan. Gos. Univ.*, 2013, no. 1, pp. 67–74.

15. Bulychev, Yu.G., Some Aspects of Identification of Dynamic Objects under Incorrect Observation Conditions, *Autom. Remote Control*, 2020, vol. 81, no. 6, pp. 1073–1090.

16. Dyatlov, A.P. and Dyatlov, P.A., Doppler Detectors of Moving Objects Using an "Extraneous" Radiation Source, *Spets. tekhnika*, 2010, no. 5, pp. 16–22.

17. Aidala, V.J. and Nardone, S.C., Biased Estimation Properties of the Pseudolinear Tracking Filter, *IEEE Transact. Aerospas. Electron. Syst.*, 1982, vol. 18, no. 4, pp. 432–441.

18. Amelin, K.S. and Miller, A.B., An Algorithm for Refinement of the Position of a Light UAV on the Basis of Kalman Filtering of Bearing Measurements, *J. Commun. Techn. Electron.*, 2014, vol. 59, no. 6, pp. 622–631.

19. Miller, A.B., Development of the Motion Control on the Basis of Kalman Filtering of Bearing-Only Measurements, *Autom. Remote Control*, 2015, vol. 76, no. 6, pp. 1018–1035.

20. Bulychev, Yu.G. and Mozol, A.A., Single-Position Passive Location of the Radiation Source with Curvilinear Motion and Taking into Account the Evolution of the Signal Period at the Receiving Point, *Radiotekh. Elektron.*, 2023, vol. 68, no. 2, pp. 131–137.

21. Kondratiev, V.S., Kotov, A.F., and Markov, L.N., *Mnogopozitsionnye radiotekhnicheskie sistemy* (Multi-Position Radio Engineering Systems), Moscow: Radio i Svyaz', 1986.

22. Chernyak, V.S., *Mnogopozitsionnaya radiolokatsiya* (Multi-Position Radar), Moscow: Radio i Svyaz', 1993.

23. Nefedov, V.I., Sigov, A.S., Bityukov, V.K., and Samokhina, E.V., *Elektroradioizmereniya*, Moscow: Forum: Infra-M, 2018.

24. Lawson, Ch. and Henson, R., *Chislennoe reshenie zadach metoda naimen'shikh kvadratov* (Numerical Solution of Least Squares Method Problems), Moscow: Nauka, 1986.

25. Zhdanyuk, B.F., *Osnovy statisticheskoi obrabotki traektornykh izmerenii* (Fundamentals of Statistical Processing of Trajectory Measurements), Moscow: Sovetskoe Radio, 1978.

26. Wentzel, E.S., *Teoriya Veroyatnostei* (Probability Theory), Moscow: Vysshaya Shkola, 1999.

27. Bulychev, Yu.G., Ivakina, S.S., and Nasenkov, I.G., Method of Passive-Energy Location and Navigation in Stationary and Non-Stationary Productions, *Radiotekhnika*, 2015, no. 6, pp. 107–115.

28. Bulychev, Yu.G. and Eliseev, A.V., Computational Scheme of Invariant Unbiased Estimation of Values of Linear Operators of a Given Class, *Zh. Vychisl. Mat. Mat. Fiz.*, 2008, vol. 48, no. 4, pp. 580–592.

29. Bulychev, Yu.G. and Burlai, I.V., A Parametric Identification Method of Control Systems under Inexact Input Data, *Avtom. Telemekh.*, 1997, no. 11, pp. 56–65.

This paper was recommended for publication by O.N. Granichin, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 823–830.

Machine Learning for Diagnosis of Diseases with Complete Gene Expression Profile

A. M. Mikhailov[*,a], **M. F. Karavai**[*,b], **V. A. Sivtsov**[*,c], **and M. A. Kurnikova**[**,d]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
[**] *Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia*
e-mail: [a] *alxmikh@gmail.com,* [b] *mkaravay@yandex.ru,* [c] *TheDeGe@yandex.ru,* [d] *mish2109@yandex.ru*

Received July 19, 2022
Revised March 21, 2023
Accepted March 30, 2023

Abstract—This paper considers the use of machine learning for diagnosis of diseases that is based on the analysis of a complete gene expression profile. This distinguishes our study from other approaches that require a preliminary step of finding a limited number of relevant genes (tens or hundreds of genes). We conducted experiments with complete genetic expression profiles (20 531 genes) that we obtained after processing transcriptomes of 801 patients with known oncologic diagnoses (oncology of the lung, kidneys, breast, prostate, and colon). Using the indextron (instant learning index system) for a new purpose, i.e., for complete expression profile processing, provided diagnostic accuracy that is 99.75% in agreement with the results of histological verification.

Keywords: pattern recognition, machine learning, inverse patterns, gene expression profiles, diagnosis of diseases

DOI: 10.25728/arcRAS.2023.92.25.001

1. INTRODUCTION

The common denominator of existing approaches to the diagnosis of oncologic diseases based on gene expression profiles is the use of a limited set of genes; see, for example, methodological and review papers [1–6]. With this approach, it is also necessary to perform a preliminary analysis and identify individual genes or combinations of genes whose expression activity is most characteristic in the case of specific diseases. However, the diagnostic accuracy achieved does not exceed 95% on sets that include up to 90 genes [4], which is presumably due to the limited number of genes used. For example, in [5], this number is reduced to just 10 genes. Thus, the data analysis process consists of two stages. For example, in [6], in the first stage, researchers use the principal component analysis, selecting 103 genes. They carry out the final diagnosis in the second stage, where a Bayesian neural network provided 93.66% accuracy, a deep neural network provided 93.41%, and logistic regression — 92.82%. Furthermore, the third stage can take place. It would reject questionable results based on a decision threshold that allows to almost completely eliminate wrong diagnoses [6] at the cost of not diagnosing some cases, which is considered to be more acceptable than a wrong diagnosis.

This study examines the use of machine learning for disease diagnosis based on the analysis of the complete gene expression profile, i.e., the activity of all 20 531 known genes, which simplifies the process by eliminating the need for preselection of relevant genes. This paper discusses the details of the operation of the machine learning system used.

Fig. 1. HiSeq 2500 system.

In biology, the expression level estimates the transcriptional activity of a gene as the amount of messenger RNA (mRNA) it produces. Obtaining gene expression profiles is a minimally invasive or completely non-invasive procedure, such as saliva sampling, which determines the comfort of such a procedure for the patient. The quality of diagnosis depends on the completeness of the profile, i.e., the number of genes considered. However, the use of complete gene expression profiles is more of a future of medical diagnosis, as currently there is a lack of inexpensive equipment for mass application to obtain such profiles. Currently, we can obtain profiles using the following technologies, among others [7]:

— PCR tests;

— DNA microarrays;

— sequencing.

If the analysis involves the expression level of a relatively small number of genes, we can use the available and relatively inexpensive real-time quantitative PCR (qPCR) method. The second technology, DNA microarrays, provides a more accurate assessment of gene expression. However, when it comes to the expression profile of all genes, sequencing is currently required. One of the most popular platforms for high-throughput sequencing is Illumina equipment [8]. The HiSeq 2500 system utilizes next-generation SBS technology, which supports massively parallel sequencing using fluorescently labeled nucleotides, allowing the reading of individual bases as they are incorporated into growing DNA strands. Figure 1 shows an example of equipment that utilizes this method.

Illumina HiSeq performs two functions, such as genome sequencing and determination of expression level. The latter is achieved by reading not DNA but mRNA code. The expression level is inferred from the number of mRNA copies. Considering the decreasing cost of RNA sequencing, it is possible in the future to create simpler and less expensive equipment for installation in widely spread laboratories such as INVITRO. With the widespread availability of such equipment, prospective diagnostics could be discussed as the use of the complete expression profile automatically takes into account all possible **active** gene combinations, which are difficult to predict and that influence the onset of various diseases. Additionally, gene expression profiles serve as valuable materials not only in diagnostics but also in various scientific and clinical research.

The second component of the considered methodology is machine learning, which allows automatic learning from complete biological profiles without analyzing genetic combinations for subsequent classification and diagnosis. Artificial neural networks [9] are traditionally used for machine learning and have become a popular method applied to solve a variety of prediction and pattern recognition problems. However, training such networks takes a long time and requires significant computational resources, as it involves the calculation of a large number of coefficients to adapt the multilayer network architecture to a specific task. Nevertheless, we can significantly accelerate training and make it borderline instantaneous if we use an image indexing system [10, 11], similar to search engines like Google [12], where incremental learning is reduced to indexing new documents. In 1998, the term indextron was introduced [13] — a term used only as the name of the image indexing device, but not the name of the image indexing method. The difference between index-

tron and search engines lies in the fact that it inverts numerical data instead of textual documents. Section 3 discusses the specifics of inverting numerical images.

It should be noted that the approach used in [10] has similarities with the later proposed TF-IDF method [14], developed for document retrieval by computing the inverse document word frequencies. The TF-IDF method is utilized in search engines, where document name frequencies are calculated in fully inverted files. However, textual search engines do not work with numerical data as the document frequency of a word can change depending on the current set of documents, and the keyword itself might completely disappear due to even slight noise.

It should be mentioned that the indexing approach to pattern recognition is hardly employed in machine learning systems, where most methods utilize iterative learning, gradient descent, and a significant number of adaptive coefficients, which ultimately leads to slow learning. At the same time, the human brain can memorize new visual patterns at a glance. This study employ the method of numerical data indexing instead of iterative learning to achieve almost instantaneous learning when dealing with large volumes of data. Previous work on indexing non-textual numerical data can be found in [15], where the authors use such fast methods for image recognition in movies. However, the study [15] reduces the recognition of noisy numerical patterns representing images to converting the numerical patterns to textual form, followed by the use of existing methods for text information retrieval. The approach we use in this paper is applying the index recognition method [10, 11], which enables the indexing of noisy numerical patterns without intermediate conversion into textual form. It is also worth noting that the objective of this article is not to develop a new machine learning method, but to utilize the indexing recognition method for a new purpose, namely, for nearly instantaneous learning when working with large databases of biological data.

2. SOURCE DATA

In the experiment, we used a dataset called Gene Expression Cancer RNA-Seq [16] to compare the effectiveness of training and classifying data using a neural network and an indextron. This data set consists of 801 rows, each containing 20 531 floating point numbers (see Table 1). Each number with profile/gene coordinates represents the level of activity of the corresponding gene, measured in arbitrary units ranging from 0 (no activity or absence of the gene) to a maximum activity of 15. The numbers in each nth row represent the activity levels of the corresponding genes of the nth patient. The diagnoses of all patients in the dataset [15] are known and were determined using other clinical methods. In the experiment, we used the 401 odd rows for training, during which we provided the system with the diagnosis associated with the corresponding row. During testing, we used the 400 even rows, which were classified by assigning them to one of the five classes, and the class found was compared with the known diagnosis of the corresponding patient.

Table 1. Gene activity of patients

Profile	Gene 0	Gene 1	Gene 2	⋯	Gene 20 529	Gene 20 530	Class (diagnosis)
0	0.0	2.017	3.266	...	5.287	0.0	3
1	0.0	0.592	1.588	...	2.094	0.0	1
...	...	...	...	...	...	...	...
800	0.0	2.325	3.806	...	4.551	0.08	5

3. RECOGNITION PROBLEM STATEMENT AND SOLUTION METHOD

In this study, we use the index recognition method, proposed and considered in [10] and improved in [11] and other works, for a new purpose, namely diagnosing diseases by classifying gene expression profiles. We discuss the specifics of this method application in this problem below.

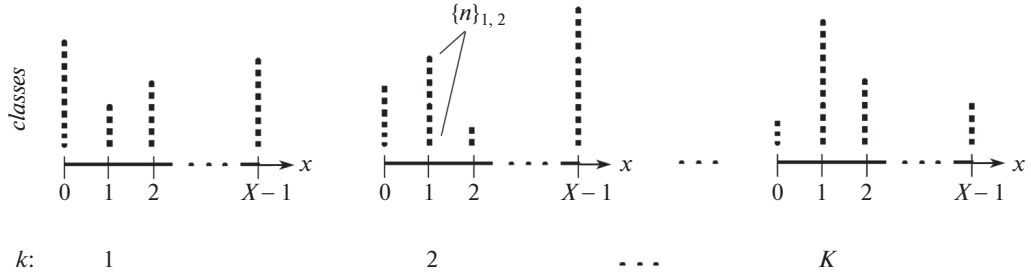


Fig. 2. Inverse patterns shown as K column groups.

Let all variables be integers, and let there be N patterns, where each pattern is represented by a K-dimensional feature vector

$$\mathbf{x_n} = (x_{n,1}, \ldots, x_{n,k}, \ldots, x_{n,K}), \quad n = 1, \ldots, N. \tag{1}$$

Here, Chebyshev distance between any two patterns $\mathbf{x_p}$ and $\mathbf{x_q}$ $(p, q \leqslant N)$ is greater than a certain predetermined number R, i.e. $|\mathbf{x_p} - \mathbf{x_q}| > R$, and variable $x_{n,k}$ $(0 \leqslant x_{n,k} < X)$ is a value of kth feature of the vector that represents nth pattern. Inequality $|\mathbf{x_p} - \mathbf{x_q}| > R$ means that for vectors $\mathbf{x_p}$, $\mathbf{x_q}$ there exists at least one dimension k such that $|x_{p,k} - x_{q,k}| > R$. In this case, every nth vector represents the nth class, which includes all vectors $\mathbf{x}$ such that $|\mathbf{x} - \mathbf{x_n}| \leqslant R$. Value R determines the size of the class and is called a radius of generalization.

Classification problem. For given vector $\mathbf{x} = (x_1, \ldots, x_k)$, find class n such that $|\mathbf{x} - \mathbf{x_n}| \leqslant R$. If such a class does not exist then add a new vector $\mathbf{x_{N+1}} = \mathbf{x}$ to the list (1) and increase the number of classes by 1.

Obviously, we can solve this problem by comparing the given vector $\mathbf{x}$ with vectors that represent classes, i.e., by exhaustive enumeration of classes. However, solving the classification problem with the inverse pattern method allows a significantly accelerated search, especially in the case of a large number of classes.

Consider solving the classification problem with inverse patterns. When using this method an unknown pattern $\mathbf{x}$ belongs to class m such that

$$m : H_R(m|\mathbf{x}) = \max_{n=1}^{N} H_R(n|\mathbf{x}).$$

Here, $H_R(m|\mathbf{x})$ is a histogram of class names contained in inverse patterns of features of vector $\mathbf{x}$. Thus, $H_R(m|\mathbf{x})$ is a conditional histogram of classes, as it depends on vector $\mathbf{x}$.

Note that if radius of generalization R is equal to zero, then the number of classes would be equal to number N of vectors in the list (1). If the number of actual externally determined classes is less than N then we use table function $class(n)$, $n = 1, \ldots, N$ that settles the mapping between the classes in the list (1) and external classes which are always known during supervised learning. Studies [10, 11] determine inverse patterns of features such as certain sets of class names. We index such sets with two-dimensional indices. Each two-dimensional index is determined by a pair of numbers (x, k), where x is a feature value, and k is a measurement number. The set of names of $\{n\}$ classes in the left part of equality (2) is an inverse pattern of feature x in measurement k.

$$\{n\}_{x,k} = \{n : x_{n,k} = x\}, \quad k = 1, \ldots, K, \quad x = 0, 1, \ldots, X - 1. \tag{2}$$

Figure 2 graphically illustrates the concept of inverse patterns. There, they are represented by columns of points, heights of which equal energies of the corresponding inverse patterns. As mentioned above, the elements of columns are class names. Due to the fact that inverse sets have

two-dimensional indices, we split all columns into K groups where each group k can contain up to X columns and each column up to N classes. Here, X is a feature value range. What remains to be done to solve the classification problem with inverse patterns is to find the class histogram $H_R(n|\mathbf{x})$. Let the input pattern be presented by vector $\mathbf{x} = (x_1, \ldots, x_k)$. Then we can find the histogram of classes contained in inverse patterns using the following algorithm:

$$\forall n \in \{n\}_{x(k)+r,k}, \quad H_R(n|\mathbf{x}) = H_R(n|\mathbf{x}) + 1. \tag{3}$$

Here $r = -R, \ldots, -1, 0, 1, \ldots, R,\ k = 1, \ldots, K,\ x(k+r), k = x_{k+r,k}$. Therefore, the classification criterion presented above is a class histogram, the maximum position of which we have to find. To understand the algorithm we need to use the concept of "inverse patterns." Expression (2) defines it and is equipped with the necessary commentaries. Inverse patterns that contain sets of pattern classes act as input information for loop (3). This loop histograms classes and yields the required histogram as its output. The position of its maximum corresponds to the desired class. In terms of programming, this is a loop on macrocolumns r, measurements k, and classes n that are contained in inverse patterns.

During the indextron training, if the histogram maximum for input $\mathbf{x}$ is lower than a certain chosen threshold then the number of classes is increased by 1, $N = N + 1$, and this new number gets stored in corresponding inverse patterns

$$\{n\}_{x(k),k} = \{n\}_{x(k),k} \bigcup N, \quad k = 1, \ldots, K.$$

The initial condition is always $N = 0$. However, we do not calculate the class histogram (3) if the training uses a zero-valued radius of generalization $R = 0$. At the same time, every iteration the number of classes increases by 1. The study [17] presents a graphical illustration of this algorithm.

During recognition, if the histogram maximum for input $\mathbf{x}$ is lower than a certain threshold, then the input patterns remain unrecognized.

Note that the histogram maximum equals the dimension K of pattern $\mathbf{x}$. It is possible to show that it is derived from the following properties of inverse patterns: the columns of each group

- contain strictly N different class names: $\sum\limits_{x=0}^{X-1} |\{n\}_{x,k}| = N$,
- do not intersect: $\{n\}_{x,k} \cap \{n\}_{y,k} = \varnothing$.

During learning and recognizing a pattern, we use the value of its features x_k, $k = 1, \ldots, K$ as an address of a column located in the kth dimension. At the same time, we coerce the real feature values x into an integer range of [0–255]:

$$x = 255(x - x_{\min})/(x_{\max} - x_{\min}).$$

Experience suggests that such continuous value sampling usually does not reduce the accuracy of classification. Table 2 presents an example of normalization of Table 1 data.

Therefore, the features of each received K-dimensional pattern isolate K columns — one column in each group. The issue is that, usually, for numerical features, the column intersection appears

Table 2. Normalized data

Profile	Gene 0	Gene 1	Gene 2	⋯	Gene 20 529	Gene 20 530	Class (diagnosis)
0	0	82	137	...	112	0	3
1	0	24	66	...	36	0	1
...	...	...	...	...	...	...	...
800	0	95	160	...	95	0	5

to be empty due to measurement errors and pattern deformations alter the column addresses. For this reason, we introduce macrocolumns, i.e., along with the column with address x we consider its neighbors with addresses $x \in [-R, R]$. The possibility of using macrocolumns is the result of inverse pattern properties mentioned above. As noted previously, we find the column intersections by calculating a histogram that determines the frequencies of class appearances. At the same time, the input pattern belongs to the most frequently appearing class.

4. SOFTWARE AND HARDWARE IMPLEMENTATION OF INDEXTRON

We implemented the parallel version of the algorithm in Python and ran it on the Nvidia GeForce GTX 1660 Super GPU. This graphics card has 44 multiprocessors and 1408 cores, allowing for parallelization of one process into 1408 parallel subprocesses. In the problem considered, training can be parallelized into 20 531 processes, corresponding to the number of all genes in the profile. The maximum possible parallelization is $20\,531 * (2R + 1)$ processes, as for each gene, all values within the generalization radius can be checked in parallel.

We achieved parallelization using the *cuda* module from the *numba* library. To process threads by the GPU multiprocessors, the threads were divided into blocks using the following formula:

$$blocks\_per\_grid = (number\_of\_iterations| + (threads\_per\_block - 1))//threads\_per\_block,$$

where *number_of_iterations* is the number of parallel iterations.

Before the execution of the program, we copied the arrays of input data from the main memory to the GPU memory using the *cuda.to_device*() function, and upon completion of the program, we copied the execution results back to the main memory using the *device_array.copy_to_host*() function.

We used a decorator to parallelize the write and read functions. Within the parallelized functions, we call a function returning the index of the thread on which the corresponding parallel iteration should be executed.

5. RESULTS

Table 3 presents the accuracy and the number of operations for the indextron and the neural network training for the disease diagnosis problem based on data from 801 patients.

Table 3. Results comparison.

	Neural network	Indextron
Accuracy (%)	99.5	99.75
Number of operations	46 bln addition/ multiplication operations	8 mil memory write operations

The indextron training time on a single-core 1.6 GHz laptop is 0.43 s. Training time of the four-layer neural network on the same hardware increases, according to the increase in the number of operations, $46 * 109/8 * 106 = 5750$ times. Note that the architecture of the indextron is ideal for parallel implementation. It allowed us to achieve almost instant learning in 75 ms for the problem considered. Generalization radii R in the training and recognition modes are 0 and 84 correspondingly, i.e., 33% of the feature variation range (0–255).

6. CONCLUSION

1. The simplicity of the indextron algorithm, which only requires writing K integers to memory for training each pattern and has classification complexity of $O(hK)$ operations for reading and

summation, allows the creation of large gene expression databases to diagnose various diseases (h is the average column height, where $h = N/X$).

2. With the wide availability of equipment for finding gene expression profiles, it becomes possible to create a search diagnostic system similar to Google for mass use, where queries are in numerical form rather than text.

3. The experiments carried out confirm the existence of such a possibility, as fast training with the addition of new data to the database, even at the software level, takes only $0.43/800 = 0.00054$ s, and the accuracy of diagnosing five types of cancer was 99.75%. However, diagnosing many other diseases will require the creation of a large number of different databases.

REFERENCES

1. Khan, J., Wei, J., Ringner, M., et al. Classification and Diagnostic Prediction of Cancers Using Gene Expression Profiling and Artificial Neural Networks, *Nat Med.*, 2001, June 7(6): 673–9. https://doi.org/10.1038/89044

2. Kumar, A. and Halder, A., Greedy Fussy Vaguely Quantified Rough Approach for Cancer Relevant Gene Selection from Gene Expression Data, *Soft Comput.*, 2022, vol. 26, pp. 13567–13581. https://doi.org/10.1007/s00500-022-07312-4

3. Houssein, E., Hassan, H., Mustafa, al-sayed, et. al., Gene Selection for Microarray Cancer Classification based on Manta Rays Foraging Optimization and Support Vector Machines, *Arabian Journal for Science and Engineering*, 2022, vol. 47, pp. 2555–2572. https://doi.org/10/1007/s13369-021-06101-8

4. Zheng, Y., Sun, Y., Kuai, Y., et al., Gene Expression Profiling for the Diagnosis of Multiple Primary Malignant Tumors, *Cancer Cell Int.*, 2021, vol. 21, Article no. 47. https://doi.org/10.1186/s12935-021-01748-8

5. Ye, Q., Wang, Q., Qi, P., et. al., Development and Validation of a 90-Gene Real-Time PCR Assay for Tumor Origin Identification, *Symposium MXW*, 2018.

6. Joshi, P. and Dhar, R., EpICC: A Bayesian Neural Network Model with Uncertainty Correction for a More Accurate Classification of Cancer, *Sci. Rep*, 12, 2022, Article no. 14628. https://doi.org/10.1038/s41598-022-18874-6

7. Steiling, K. and Christenson, S., Tools for Genetics and Genomics: Gene Expression Profiling, *UpToDate*, 2021. Retrieved from https://www.uptodate.com/contents/ tools-for-genetics-and-genomics-gene-expression-profiling

8. St. Petersburg University Research Park. High-throughput complete genome sequencing system, 2023 https://researchpark.spbu.ru/equipment-biobank-rus/ equipment-biobank-genom-rus/equipment-biobank-ngsseq-rus/1762-biobank-hiseq-2500-sequencing-system-rus

9. IBM. What are neural networks? // Retrieved from https://www.ibm.com/cloud/learn/neural-networks

10. Mikhailov, A. and Pok, Y.M., Artificial Neural Cortex, *Smart Engineer. Syst. Design.*, 2001, vol. 11, ASME PRESS, N.Y. P. 113–120.

11. Mikhailov, A. and Karavay, M., Pattern Inversion as a Pattern Recognition Method for Machine Learning, *Cornell University*, 2021. Retrieved from https://arxiv.org/abs/2108.10242

12. Brin, S. and Page, L., The Anatomy of a Large-Scale Hypertextual Web Search Engine, *Comput. Networks ISDN Syst.*, 1998, vol. 30, nos. 1–7. Stanford University, Stanford, CA, 94305, USA. Retrieved from https://doi.org/10.1016/S069-7552(98)00110-X

13. Mikhailov, A., Indextron, *Artificial Neural Networks in Engineering Conf. (ANNIE 1998)*, St. Louis, Missouri, Nov. 4–7, 1998. Proceedings Vol. 8: ANNIE 1998, Publisher: ASME Press, ISBN: 0791800822

14. Jones, K., A Statistical Interpretation of Term Specificity and Its Application in Retrieval, *J. Document.*, MCB Univer.: MCB Univer. Press, 2004, vol. 60, no. 5. pp. 493–502. ISSN 0022-0418

15. Sivic, J. and Zisserman, A., Efficient Visual Search of Videos Cast as Text Retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, vol. 31, no. 4. https://doi.org/10.1109/TPAMI.2008.111

16. UCI. Machine learning repository // Retrieved from https://archive.ics.uci.edu/ml/datasets/gene+expression+cancer+RNA-Seq

17. Mikhailov, A. and Karavay, M., Indextron, *Proceedings of the 10th International Conference on Pattern Recognition Application and Methods*, 4–6 Feb 2021, Vienna, V.1-978-989-758-486-2, pp. 143–149. https://doi.org/10.5220/0010180301430149

This paper was recommended for publication by O.N. Granichin, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 831–851.

Method for Unmanned Vehicles Automatic Positioning Based on Signal Radially Symmetric Markers Recognition of Underwater Targets

R. M. Shakirzyanov[*,a], **M. P. Shleymovich**[*,b], **and S. V. Novikova**[*,c]

[*] *Tupolev National Research Technical University, Kazan, Russia*

e-mail: [a] *RMShakirzyanov@kai.ru,* [b] *MPShleymovich@kai.ru,* [c] *SVNovikova@kai.ru*

Received November 19, 2022

Revised March 16, 2023

Accepted April 28, 2023

Abstract—The article describes a method for automatically recognizing the target points of the trajectories of unmanned vehicles moving underwater, such as autonomous submarines and flying underwater vehicles of aircraft-like structures. The coordinate of the center of an object with radial symmetry properties is considered a terminal control point. A method for constructing a multiscale weighted image model based on the developed fast radial symmetry transformation and the Hough method is proposed, which ensures noise stability and high speed of calculating the coordinates of the desired point. When the object of interest has a contour of a specific color, a model is based on our proposed chromatic and weight components. As an example of detection, we have given an algorithm for detecting a base underwater station with light markers as a signal luminous ring.

Keywords: automatic transport systems, unmanned underwater vehicle, unmanned flying underwater vehicle, computer vision, image processing, object detection in images, Hough method, fast radial symmetry transformation, image weight model, multiscale image weight model

DOI: 10.25728/arcRAS.2023.24.60.001

1. INTRODUCTION

When building automatic control systems for unmanned vehicles, a necessary condition is to set the final position of the device in space. In most works on automatic control of unmanned vehicles, the researcher sets the final position in the form of a set of specific coordinates [1, 2]. However, in actual conditions, problems arise when the device must independently detect an object of interest, the position of which will set the final position of the device during control. Transport systems, where the device independently determines the end point of its movement and calculates the trajectory of movement to it, are usually called intelligent transport systems [3]. The method for determining the terminal point and constructing the optimal trajectory does not necessarily have to include elements of artificial intelligence (AI), such as machine learning models, neural networks, etc. However, if the calculation method does not involve AI, the system is more often defined as "automatic" or "automated."

The most common unmanned vehicles are unmanned aerial vehicles (UAVs) [4]. The most common unmanned vehicles are unmanned aerial vehicles (UAVs) [5, 6]. For unmanned vehicles moving underwater, the so-called autonomous uninhabited underwater vehicles (AUV) [7], positioning in most cases is carried out using ultrasonic signals [8]. The main disadvantage of this approach is the mandatory presence of special navigation equipment on board the vehicle and the need for a response signal from the target positioning point, which makes the vehicle vulnerable to external detection.

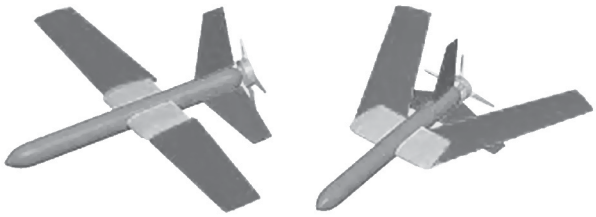


Fig. 1. The design of a diving aircraft with a folding wing (source—see [16]).

Fig. 2. The mechanics of moving guillemots in the air and underwater, applied to simulate the UAV dynamics (source—see [18]).

Positioning systems based on computer vision do not have these disadvantages. They can provide high recognition accuracy and, at the same time, do not give out their presence by external signals. In addition, visual positioning is only possible during underwater target setting for such modern unmanned systems as hybrid unmanned aerial underwater vehicles (HUAUVs) [9]. HUAUVs can overcome distances first by air, dive into the water, and perform underwater maneuvers. Due to signal propagation and dispersion characteristics at the junction of two air-water media, positioning using ultrasonic signals is challenging to implement.

For UAVs and unmanned ground vehicles, the problem of object detection based on computer vision has been well studied [10–12]. There are works on visual positioning of the UAV landing point [13], determining the position of an aircraft based on previously made satellite images of the terrain and neural network recognizers [14] etc. A common disadvantage of the proposed approaches is the need to have additional information for positioning and high requirements for computing power equipment on board the UAV.

For unmanned vehicles moving underwater, computer vision algorithms must meet additional requirements due to the peculiarities of light propagation in water and additional interference due to the turbidity of natural water bodies [15]. In the case of HUAUVs, the speed of the algorithm is also of great importance due to the relatively high speeds of movement of the vehicle and the number of calculations necessary for detecting an object since flying underwater vehicles are not capable of carrying high-power computers and their energy consumption is high relative to AUVs. Therefore, developing a visual positioning algorithm for a target set underwater with minimal energy and computational costs is especially relevant for hybrid UAUVs.

The case when the final target position of the device is underwater is more typical for flying underwater vehicles of the “diving aircraft” type [16], than, for example, for floating quadrocopters [17]. Aircraft-like structures are mainly adapted only to the direct air-to-water transition and experience significant difficulties or cannot carry out the reverse water-to-air transition at all.

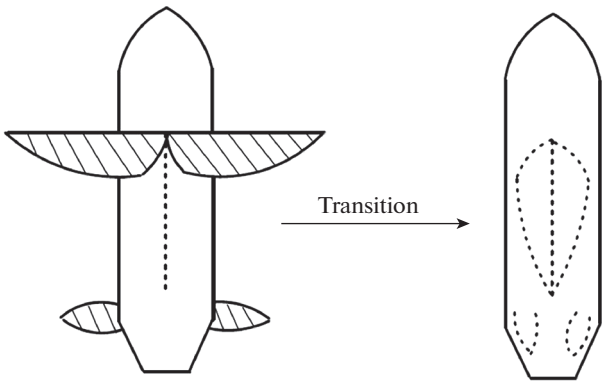


Fig. 3. UAV tail configuration in various environments (source—see [19]).

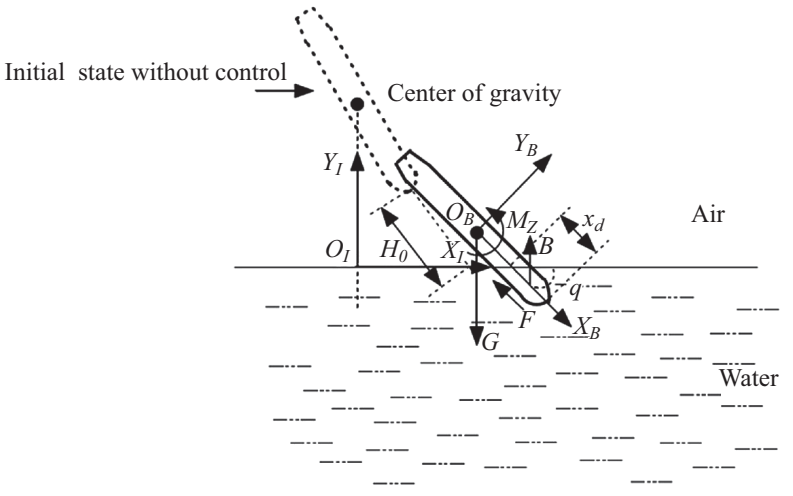


Fig. 4. The authors propose a control system for reaching the endpoint underwater by the vehicle (source—see [19]).

In [16] a design of a diving aircraft with wings deviating by 65 degrees during diving was proposed (see Fig. 1).

The design includes a variable-sweep wing and a compressed carbon dioxide engine. In [18] a design of a diving HUAUV simulating the flight and diving of a guillemot was proposed (see Fig. 2).

In [19] the automatic control system of a diving aircraft with a folding tail is described (see Fig. 3).

In [19] the authors propose a control system for reaching the endpoint underwater by the vehicle (see Fig. 4).

The control optimality criterion is the accuracy of reaching a given position (x_{target}, y_{target}) at the final time t_f.

$$J = \sqrt{(x_h(t_f) - x_{target})^2 - (y_h(t_f) - y_{target})^2}. \tag{1}$$

The control parameters are the initial pitch angle θ_0 and the initial velocity along the axis OX v_{x_0}, the trajectory is calculated without the possibility of maneuvering. As the target coordinates (x_{target}, y_{target}) the authors define a specific point underwater. However, in the case of adaptive control, the device can calculate these coordinates independently in an automatic mode based

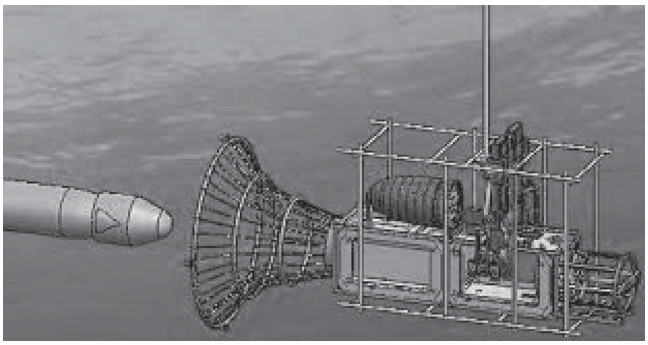


Fig. 5. An example of docking an autonomous uninhabited underwater vehicle with a docking station (source—see [20]).

on machine vision. Thus, the problem arises of recognizing the coordinates of a specific target object underwater, the mathematical center of gravity of which will be the target control point (x_{target}, y_{target}). This problem is one of the possible typical examples of a recognition problem to solve the optimal control problem.

For underwater unmanned vehicles, setting objects of interest underwater is the only possible one. Such an object of interest can mainly be a particular docking station [20], sed to recharge the AUV battery and/or transmit data (see Fig. 5).

In this case, the positioning features of the AUV and HUAUV will be the same since the target is underwater in both cases. The fact that the target is under water is vital for developing a target object detection method since water turbidity, the presence of suspension, and image distortion due to light refraction impose stringent requirements on the detection method for resistance to external noise. For example, optical positioning methods for underwater targets are described in [21, 22]. One [23] and two [24] cameras can be used in this case. The main limitation of the proposed methods is their high consumption of computing and energy resources since almost all of them are designed for AUVs. This circumstance makes the proposed methods poorly applicable to HUAUVs. So, in [20] a convolutional neural network CNN is used as a mathematical recognition model. The efficiency of such networks is high if they are specially trained for a relatively narrow class of objects. For example, it has been shown that CNNs can recognize handwritten text with an accuracy of up to 99% [25]. However, their accuracy drops to 70% or less when recognizing arbitrary natural objects. In addition, CNNs require powerful computing devices for training, which can last from several hours to several days [26]. Modern modifications can give either higher accuracy or higher speed but cannot yet combine both of these qualities. So, in [27] the so-called two-pass algorithms of the FASTER RCNN type are described, reaching an accuracy of 73.2% (VOC07[1] dataset). The network training time is not described, but the network has a relatively low recognition rate of 17 images per second. Single-pass networks like YOLO achieve a recognition rate of 155 images per second, but their accuracy is only 52.7% on the same VOC07 dataset.

Thus, developing an algorithm for detecting objects underwater with minimal computational and time costs is relevant, especially for HUAUVs.

2. FORMULATION OF THE PROBLEM

Consider the problem of recognizing an underwater object in automatic mode from an unmanned vehicle by detecting a base docking station with light markers in the form of a signal luminous ring. The center of the ring defines the target point of the optimal control trajectory without maneuvering

[1] VOC07 is a free access data set containing classified images (Visual Object Classes-VOC) for 20 classes, edition 2007

(x_{target}, y_{target}) according to the example of the problem (1). In the general case, the task can be solved simultaneously for several possible N_{target} targets.

Given:

(a) color image $I = (I_{ij})_{N_x \times N_y}$ in RGB color model;

(b) the set of radii of circles $\mathbf{N}$ to be found;

(c) the set of admissible colors $\overline{\mathbf{C}}$, in shades of which the desired circles can be colored;

(d) allowable mismatch between the positions of the centers of the circles Δ.

Find:

(e) the set of centers of circles with their radii defining the terminal point of the optimal control $\tilde{A} = \left\{ \left(x^k_{target}, y^k_{target} \right), r^k \right\}$, $k = \overline{1, N_{target}}$.

This study aims to develop a computational algorithm that takes positions (a)–(d) as input and generates positions (e) as output. An additional requirement for the algorithm is resistance to external noise and minimization of computational costs and computation time.

When the actual size of a circle (for example, a docking ring) is known, several decreasing values from the real to the minimum possible radius should be chosen as the desired image radii. Subsequently, based on the found value of the radius on the image, it will be possible to determine the distance to the object using additional measuring instruments.

The method should provide recognition under slight deformations of the circle into an ellipse with a ratio of the minor and significant semiaxes of at least 0.95, corresponding to a shift of the direct projection of the view by no more than 20° [28].

3. METHOD FOR SOLVING THE PROBLEM OF RECOGNITION OF COLORED ROUND OBJECTS

An important task solved using computer vision is processing images from an unmanned vehicle to detect and localize given objects — light markers. These objects are characterized by shape and color. Therefore, methods for analyzing color and shape features are used to detect and recognize them.

Recently popular algorithms based on deep neural networks for this problem statement will be redundant since a simple geometric figure is subject to recognition rather than a complex multi-color object. In addition, the generalization complexity and the training stage's duration remain topical problems when using neural networks. Therefore, classical methods in which image analysis is performed remain relevant.

Many practical approaches to the analysis of the shape of objects in images are based on the methods of contour analysis [29]. Light markers for docking underwater vehicles are colored objects with a radially symmetrical shape. For their detection and localization, methods based on analyzing the properties of radial symmetry can be applied. The proposed method is based on the joint use of the well-known Hough method [30] and the transformation of fast radial symmetry using the construction of the Gaussian pyramid.

3.1. Fast Radial Symmetry Transformation

This paper proposes an efficient approach to solving the problem under consideration based on the fast radial symmetry transformation FRST (Fast Radial Symmetry Transform). The FRST transformation makes constructing a weighted image model possible, which can effectively localize the center of a radially symmetric object [31, 32].

To obtain a weighted image model, you must perform the following steps:

1) convert the image to grayscale;

2) calculate the gradients of the brightness function on the image;

3) calculate the values of the weight matrix elements;

4) normalize the values of weight matrix elements;

5) calculate the values of the elements of the matrices of generalized weights;

6) perform low-pass filtering of generalized weight matrices;

7) calculate the values of the matrix elements of averaged weights.

The first step is to convert the image to grayscale, in which each pixel is assigned a brightness value. In the simplest case, the transformation calculates the average values of the intensities of the red, green, and blue color components. Thus, as a result of the first step, the image is represented as a brightness function, the arguments of which are the values of the pixel coordinates.

In the second step, the gradients of the brightness function on the image are calculated using the following operators:

$$g_x(\mathbf{p}) := \hat{\mathbf{I}}(i+1,\, j) - \hat{\mathbf{I}}(i,\, j); \tag{2}$$

$$g_y(\mathbf{p}) := \hat{\mathbf{I}}(i,\, j+1) - \hat{\mathbf{I}}(i,\, j); \tag{3}$$

$$|\mathbf{g}(\mathbf{p})| = \sqrt{g_x^2(\mathbf{p}) + g_y^2(\mathbf{p})}, \tag{4}$$

where $\mathbf{g}(\mathbf{p}) = (g_x(\mathbf{p}),\, g_y(\mathbf{p}))$ is the gradient at pixel $\mathbf{p}$ with coordinates $(i,\, j)$; $g_x(\mathbf{p})$, $g_y(\mathbf{p})$ are the components of the gradient for the horizontal and vertical directions at pixel $\mathbf{p}$ respectively; $\hat{\mathbf{I}}(i,\, j)$, $\hat{\mathbf{I}}(i+1,\, j)$, $\hat{\mathbf{I}}(i,\, j+1)$ are the brightness values in pixels of the grayscale image with coordinates $(i,\, j)$, $(i{+}1,\, j)$ and $(i,\, j{+}1)$ respectively. In expressions (2)–(4), the sign := denotes the assignment operator, which will be used in the future.

In the third step, the values of the elements of the weight matrices are calculated. For this, the following procedure is applied:

1) the set of integer values $\mathbf{N}$ is determined, where $\mathbf{N}$ is the set of radii of the objects to be found;

2) for each value n from the set $\mathbf{N}$, the initial values of the elements of two weight matrices are formed:

$$\mathbf{M}^n(\mathbf{p}) := 0; \tag{5}$$

$$\mathbf{O}^n(\mathbf{p}) := 0, \tag{6}$$

where p is the vector of coordinates $(i,\, j)$.

3) for all elements of the weight matrices determined by the values n from the set $\mathbf{N}$, the following operators are applied:

$$\mathbf{p}_+ := \mathbf{p} + \left\lceil \frac{g_x(\mathbf{p})}{|\mathbf{g}(\mathbf{p})|}\, n \right\rceil; \tag{7}$$

$$\mathbf{p}_- := \mathbf{p} - \left\lceil \frac{g_x(\mathbf{p})}{|\mathbf{g}(\mathbf{p})|}\, n \right\rceil; \tag{8}$$

$$\mathbf{M}^n\left(\mathbf{p}_+\right) := \mathbf{M}^n\left(\mathbf{p}_+\right) + |\mathbf{g}(\mathbf{p})|; \tag{9}$$

$$\mathbf{M}^n\left(\mathbf{p}_-\right) := \mathbf{M}^n\left(\mathbf{p}_-\right) - |\mathbf{g}(\mathbf{p})|; \tag{10}$$

$$\mathbf{O}^n\left(\mathbf{p}_+\right) := \mathbf{O}^n\left(\mathbf{p}_+\right) + 1; \tag{11}$$

$$\mathbf{O}^n\left(\mathbf{p}_-\right) := \mathbf{O}^n\left(\mathbf{p}_-\right) - 11, \tag{12}$$

where $\lceil \cdot \rceil$ is the operator of rounding an actual number to the nearest no less than integer value.

In the fourth step, the elements of the weight matrices are normalized:

$$\mathbf{M}^n(\mathbf{p}) := \frac{|\mathbf{M}^n(\mathbf{p})|}{\max\limits_q \{|\mathbf{M}^n(\mathbf{q})\}|}; \tag{13}$$

$$\mathbf{O}^n(\mathbf{p}) := \frac{|\mathbf{O}^n(\mathbf{p})|}{\max\limits_q \{|\mathbf{O}^n(\mathbf{q})\}|}. \tag{14}$$

In the fifth step, the values of the elements of the matrices of generalized weights are calculated:

$$\mathbf{F}^n(\mathbf{p}) := (\mathbf{O}^n(\mathbf{p}))^\alpha \, \mathbf{M}^n(\mathbf{p}), \tag{15}$$

where α is the radial stiffness parameter.

In the sixth step, low-pass filtering of the generalized weight matrices is performed. In this case, a Gaussian low-pass filter is used as a rule. Filtering is expressed using the convolution operator

$$\mathbf{S}^n := \mathbf{F}^n * \mathbf{G}^n, \tag{16}$$

where $\mathbf{G}^n$ is matrix of coefficients of the Gaussian low-pass filter, defined for the value n from the set $\mathbf{N}$.

In the last, seventh step, the values of the elements of the matrix of averaged weights are calculated:

$$\mathbf{S}(\mathbf{p}) := \frac{1}{|\mathbf{N}|} \sum_{n \in \mathbf{N}} \mathbf{S}^n(\mathbf{p}). \tag{17}$$

The matrix of average weights obtained as a result of the FRST transformation is a weight model of the image, the analysis of which makes it possible to determine the shape parameters of objects, for example, the center coordinates and the radius of round objects.

3.2. Image Weight Model

When building a weighted image model using a fast radial symmetry transformation, it is necessary to perform a relatively large number of calculations, which is determined by the power of the set of integer values $\mathbf{N}$. The elements of this set are used to analyze significant changes in brightness at the corresponding distances from the current pixel, i.e., have the meaning of radii for radially symmetrical objects centered in image pixels.

Objects in actual images have a multiscale nature since they have different sizes. For the problem under consideration, this means using a set $\mathbf{N}$ with many elements, which are determined by the range of changes in the radii of objects of interest. Therefore, analyzing the original image presented in a multiscale form is advisable. Such a multiscale analysis reduces the number of calculations and improves object detection accuracy in the image.

To build a multi-scale image weight model, you must perform the following steps:

1) convert the image to grayscale;
2) construct a Gaussian pyramid;
3) calculate the values of the average weight matrices' elements for all Gaussian pyramid levels;
4) bring the matrices of averaged weights to the size of the original image;
5) calculate the values of the elements of the integral matrix of averaged weights.

The first step is to convert the image to grayscale, in which each pixel is assigned a brightness value.

In the second step, the Gaussian pyramid is built [33]. It is a set of images $\mathbf{P}=\{\mathbf{P}_l \,|\, l=\overline{0, L-1}\}$, where L is the number of levels in the pyramid. The original grayscale image $\hat{\mathbf{I}}$ is considered as the zero level of the pyramid $\mathbf{P}_0$. The remaining levels of the pyramid are formed as follows:

$$\mathbf{P}_l := \mathbf{G} * (2 \downarrow [\mathbf{P}_{l-1}]), \tag{18}$$

where $\mathbf{G}$ is the coefficient matrix of the Gaussian low-pass filter; $2 \downarrow [\cdot]$ is the image downsampling operator, for example, by removing every second pixel in a row and column. An element of the Gaussian pyramid of level l will have dimensions four times smaller compared to the dimensions of an element of level $l-1$.

In the third step, with the help of operators (5)–(17), weight models of images of all levels of the Gaussian pyramid are formed, i.e., a set of matrices of averaged weights $\left\{\mathbf{S}^{(0)}, \mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(L-1)}\right\}$ is built.

In the fourth step, the set $\left\{\mathbf{S}^{(0)}, \mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(L-1)}\right\}$ is transformed into the set $\left\{\mathbf{R}^{(0)}, \mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(L-1)}\right\}$, each element which $\mathbf{R}^{(l)}$ is the result of reducing the matrix $\mathbf{S}^{(l)}$ to the dimensions of the original image, i.e., to the dimensions $\mathbf{S}^{(0)}$. The specified transformation is performed as follows:

$$\mathbf{R}^{(l)} := \mathbf{G} * \left(2 \uparrow \left[\mathbf{S}^{(l)}\right]\right), \tag{19}$$

where $\mathbf{G}$ is the coefficient matrix of the Gaussian low-pass filter; $2 \uparrow [\cdot]$ is the image upsampling operator (the matrix $\mathbf{S}^{(l)}$ is considered as an image with each pixel which is associated with a weight value), for example, by duplicating each pixel in a row and column.

In the last, fifth step, the resulting matrices $\mathbf{R}^{(0)}$, $\mathbf{R}^{(1)}$, $\mathbf{R}^{(L-1)}$ are added to construct an integral matrix of averaged weights:

$$\mathbf{S}(\mathbf{p}) := \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{R}^{(l)}(\mathbf{p}), \tag{20}$$

where $\mathbf{p}$ is a pixel of the original image corresponding to the coordinates of the weight matrix elements.

As a result of the above steps, a multiscale weight model is formed, the analysis of which makes it possible to determine the shape parameters of objects in the image. The use of this model makes it possible to increase the accuracy of work and reduce computational complexity due to the smaller image area at the upper level of the pyramid.

3.3. Chromatic Image Model

When analyzing images of light markers from an unmanned underwater vehicle, it is advisable to use a model containing chromatic and weight components. The first allows you to consider the color characteristics of objects of interest, and the second is the geometric features of their shape.

Then the image model can be represented as $\langle \mathbf{C}, \mathbf{S} \rangle$, where $\mathbf{C}$ is the chromatic component, and $\mathbf{S}$ is the weight component. The weight component of the model is formed using the procedures described in the previous sections. The chromatic component is formed due to color transformation and color segmentation.

To describe the chromatic component of the model, the advantages and disadvantages of several color spaces were investigated: RGB, Lab, and HSV. These spaces were compared in terms of description methods to highlight areas of the image, the color of which lies in a specific range [34]. As a result, it was noted that the main factor limiting and complicating the use of the RGB color

model for color segmentation is a single description of both the illumination and the color component of the object, which makes it difficult to clearly define the ranges of the desired colors [35]. The indisputable advantage of the L*a*b* color model is a separate L* channel responsible for illumination and planes (a*, b*), responsible for color. However, this model describes color ranges (chromatic characteristics) as geometric figures on a plane, possibly having a complex structure. This fact complicates constructing a color distribution model for segmentation [36]. The HSV color model, in contrast to L*a*b*, allows you to set the range of desired colors by specifying the range of values of the H channel (in most cases, it is enough to specify the threshold value of the saturation set in the S channel above the value whose color purity becomes acceptable), which significantly simplifies the implementation of the method [37].

Thus, the HSV color space (model) was chosen as the color space for analyzing color images. This model is based on the characteristics of information perception by the human visual system and is represented as a set of color channels H, S, and V, which determine the color's tone, saturation, and value (brightness).

The chromatic component $\mathbf{C}$ of the image model is represented as follows:

$$\begin{cases} (h(\mathbf{p}),\, s(\mathbf{p}),\, v(\mathbf{p})), & \exists\, k,\, l,\, m:\ h(\mathbf{p}) \in \mathbf{H}^k \wedge s(\mathbf{p}) \in \mathbf{S}^l \wedge v(\mathbf{p}) \in \mathbf{V}^m; \\ (0,\, 0,\, 0), & \forall\, k,\, l,\, m:\ h(\mathbf{p}) \notin \mathbf{H}^k \vee s(\mathbf{p}) \notin \mathbf{S}^l \vee v(\mathbf{p}) \notin \mathbf{V}^m, \end{cases} \tag{21}$$

where $\mathbf{p}$ is the pixel of the original image; $\mathbf{H}^k$, $\mathbf{S}^l$, $\mathbf{V}^m$ are the kth, lth, and mth intervals of the H, S and V channels, respectively, from the set of color intervals of the given objects.

3.4. Algorithm for Detecting Round Objects on Images of Light Markers

In general, the procedure for detecting round objects of a given color is as follows:

1) convert the original image from the RGB to the HSV color model;
2) build the chromatic component of the image model;
3) convert the original image to grayscale, considering the chromatic component;
4) find circles using the Hough method on the image by the chromatic component;
5) build the weight component of the image model (weight image);
6) threshold the weight image;
7) find contours on the weight image;
8) calculate the centers of mass for each contour;
9) find circles on the image by the weight component;
10) remove falsely found circles on the halftone image by the Hough method, the centers of which do not coincide with the permissible error with any center of mass of the contours on the weighted image;
11) generate a set of coordinates of target control points with radii $\tilde{A} = \left\{ \left(x^k_{target},\, y^k_{target} \right),\, r^k \right\}$.

4. COMPUTATIONAL EXPERIMENTS

The presented algorithm was tested on the problem of recognizing circular luminous markers of an underwater docking station at different angles in turbid water conditions. For computational experiments, images of the docking station were obtained during the experiments of Liu S., Ozay M., Okatani T., published in [20]. Examples of images of the docking station docking block in the form of illuminated circles, which were recognized as a result of the experiments, are shown in Fig. 6.

The images were taken with a 620 TVL (0.8 Mpx) NanoSeaCam monocular color camera with viewing angles of $59^\circ \times 44^\circ \times 72^\circ$ (up-down-sideways).

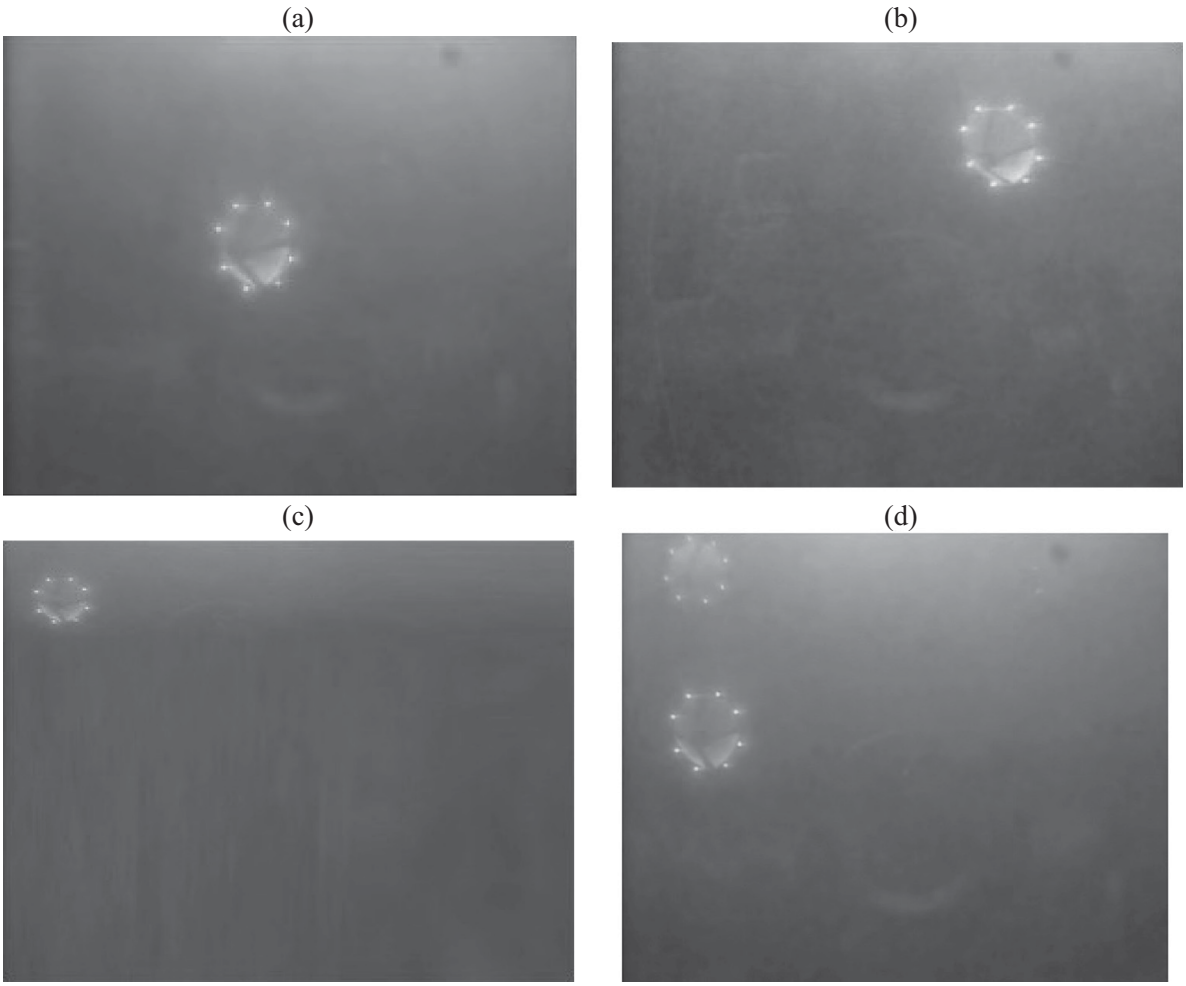


Fig. 6. Variants of images of circles — the docking luminescent ring of the underwater docking station: (a) a single object in the central part of the image, the area of the object of interest relative to the image is more than 5%; (b) a single object displaced relative to the center of the image, the area of the object of interest relative to the image is more than 5%; (c) a single object displaced relative to the center of the image, the area of the object of interest relative to the image is less than 5%; (d) multiple objects of interest in one image. (source: freely available dataset http://vision.is.tohoku.ac.jp/liushuang/a-vision-based-underwater-docking-system/dataset).

4.1. The Criterion for Assessing the Accuracy of the Algorithm

The effectiveness of the developed algorithm was evaluated based on the so-called pixel-oriented technique described in [38].

This is a statistical evaluation measure based on the count of misclassified pixels. To obtain such an estimate, we must calculate the probabilities that a randomly selected pixel in an image segmented using the algorithm belongs either to the desired object or, respectively, to the background.

Unlike generally accepted metrics such as Accuracy, Precision, or Recall, followed by ROC analysis, the pixel-based metric evaluates not the accuracy of object classification (object detected or not detected) but the detection quality. Evaluation of the detection quality includes assessing the accuracy of object localization (determining the coordinates of the center) and determining its geometric characteristics, i.e., the radius. Therefore, for the problem posed, the pixel-oriented estimate is preferable [39].

Table 1. Initial data and preliminary results of experiments on the recognition of test images

Experiment number	Prior data				Data obtained after applying the algorithm	
	$p(o)$	$p(b)$	Data obtained after applying the algorithm, %	Estimation of the object's relative area	Percentage of erroneous recognitions algorithm by criterion $p(b\|o)$	Percentage of erroneous recognitions algorithm by criterion $p(o\|b)$
1	0.01	0.99	1%	$\leq 5\%$	1	1
2	0.01	0.99	1%	$\leq 5\%$	1	5
3	0.01	0.99	1%	$\leq 5\%$	1	10
4	0.01	0.99	1%	$\leq 5\%$	1	25
5	0.05	0.95	5%	$\leq 5\%$	1	1
6	0.05	0.95	5%	$\leq 5\%$	1	5
7	0.05	0.95	5%	$\leq 5\%$	1	10
8	0.05	0.95	5%	$\leq 5\%$	1	25
9	0.1	0.9	10%	$> 5\%$	1	1
10	0.1	0.9	10%	$> 5\%$	1	5
11	0.1	0.9	10%	$> 5\%$	1	10
12	0.1	0.9	10%	$> 5\%$	1	25
13	0.25	0.75	25%	$> 5\%$	1	1
14	0.25	0.75	25%	$> 5\%$	1	5
15	0.25	0.75	25%	$> 5\%$	1	10
16	0.25	0.75	25%	$> 5\%$	1	25
17	0.5	0.5	50%	$> 5\%$	1	1
18	0.5	0.5	50%	$> 5\%$	1	5
19	0.5	0.5	50%	$> 5\%$	1	10
20	0.5	0.5	50%	$> 5\%$	1	25

According to the chosen method, the segmentation error probability of the entire image is determined as follows:

$$p_{err} = p(b|o)p(o) + p(o|b)p(b). \tag{22}$$

Here: $p(o)$ is the a priori probability that a randomly selected pixel of the original image belongs to the object (the ratio of the area of the object to the area of the entire image); $p(b)$ is the a priori probability that a randomly selected pixel of the original image belongs to the background (the ratio of the background area to the area of the entire image); $p(o|b)$ is the probability that a pixel belonging to the background is erroneously assigned to the object during segmentation (the ratio of the part of the background area erroneously assigned by the algorithm to the object to the entire background area). In the theory of statistical hypotheses, such an error is called an error of the first kind; $p(b|o)$ is the probability that a pixel belonging to an object will be erroneously assigned to the background (the ratio of the part of the object's area erroneously assigned by the algorithm to the background to the area of the object) — an error of the second kind.

So $p(o|b)$ and $p(b|o)$ are cumulative segmentation errors.

In this case, the area is understood as the number of pixels of a given area.

For cases when the desired objects are small in size relative to the entire image (occupying less than 5% of the image area), i.e.,

$$\#(o) \ll \#(\mathcal{D}),$$

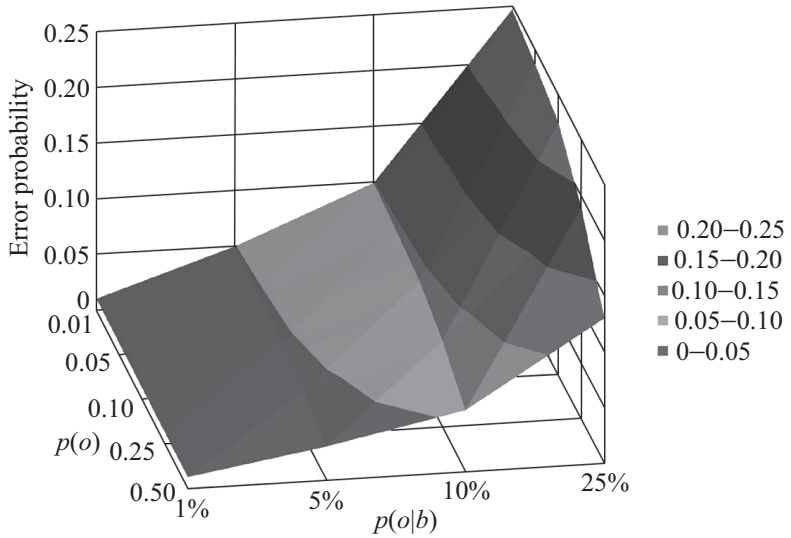


Fig. 7. Probability of the error of the algorithm calculated by the method (22) depending on the a priori probability (22) depending on the a priori probability $p(o)$ and a posteriori probability $p(b|o)$.

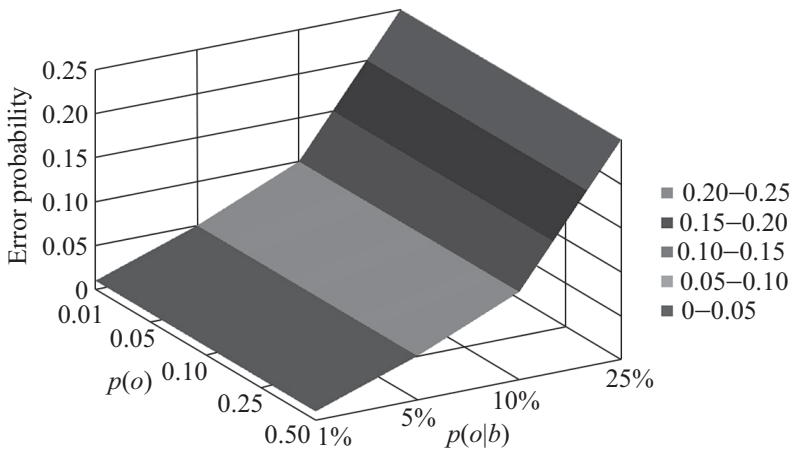


Fig. 8. Probability of an algorithm error calculated by method (23) depending on a priori probability $p(o)$ and a posteriori probability $p(b|o)$.

where $\#(o)$ is the power of the set of pixels belonging to the desired object, and $\#(\mathcal{D})$ is the power of the set of pixels of the entire image, it is advisable to use estimate (22) instead of formula (23):

$$p_{err} := \frac{\#(o_s) - \#(o)}{\#(\mathcal{D})}, \tag{23}$$

where o_s is the set of found objects as a result of segmentation.

To determine the optimal way to calculate the algorithm's accuracy, we conducted 20 independent experiments on recognizing circles in images. Among them, in 8 images, the object of interest occupied 5% or less of the area, and in 12 images, it occupied more than 5%. After image processing by the developed algorithm, posterior recognition errors $p(o|b)$ and $p(b|o)$ were obtained. A priori and a posteriori experimental data are presented in Table 1.

Then, based on the obtained data, the accuracy was calculated according to the methods (22) and (23). The results are presented in Figs. 7–10.

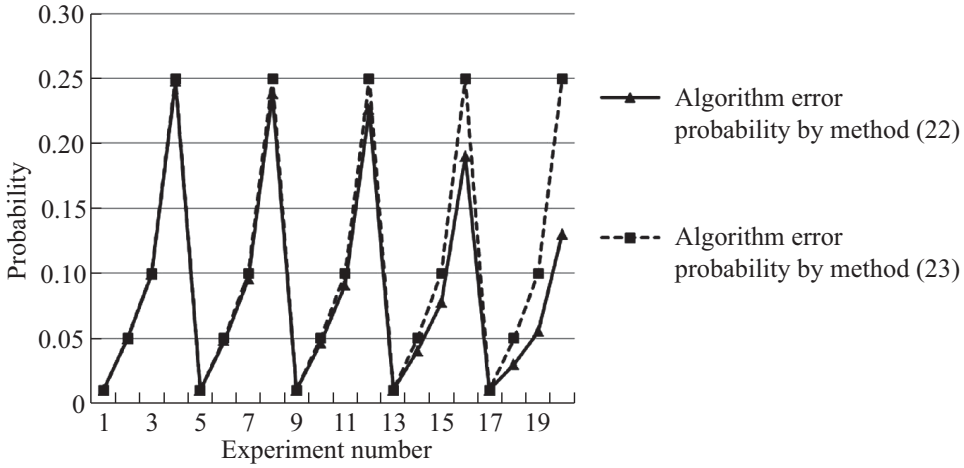


Fig. 9. Comparative efficiency of methods for estimating the algorithm's accuracy.

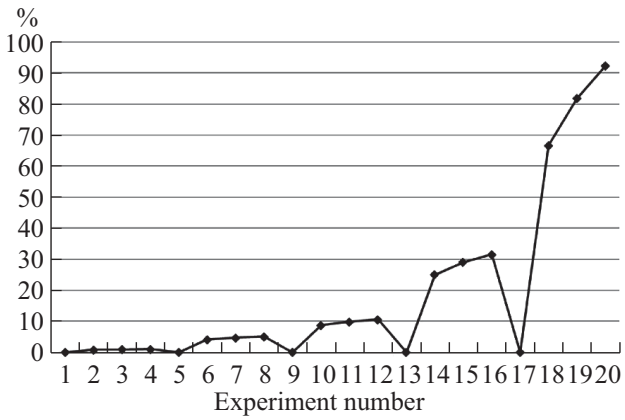


Fig. 10. Difference in accuracy of method (22) relative to method (23).

The results show that for objects with an area of up to 5% of the area of the entire image, inclusive, the correctness estimates calculated by both methods with a fixed type 2 error level of 0.01, generally have minor differences. However, if the desired object occupies an area of more than 5% of the area of the entire image, it is advisable to apply the estimate according to the Eq. (22). Thus, method (22) is preferred.

For all further studies, the accuracy was calculated according to (22).

4.2. Influence of Radial Symmetry Transformation Parameters

The dependence of the circle detection accuracy on the radial stiffness parameter α is studied. We generated and examined 100 images of circles with radii from 1 to 100 pixels on the 2400 × 8400-pixel original image. ext, we conducted experiments with a change in the stiffness parameter from $\alpha = -3$ to $\alpha = +7$. The results are presented in Fig. 11a. It can be seen that with an increase in the radial stiffness parameter α to a value of +2, the detection accuracy reaches 75% (0.75). With a further increase in α an accuracy of 95% is achieved at a value of +4 and then asymptotically tends to 100%. On Fig. 11b hows the influence of the standard deviation on the accuracy of detecting circles at the value of the radial stiffness parameter α equal to 2. It can be seen that this parameter has an insignificant effect on the result under these conditions.

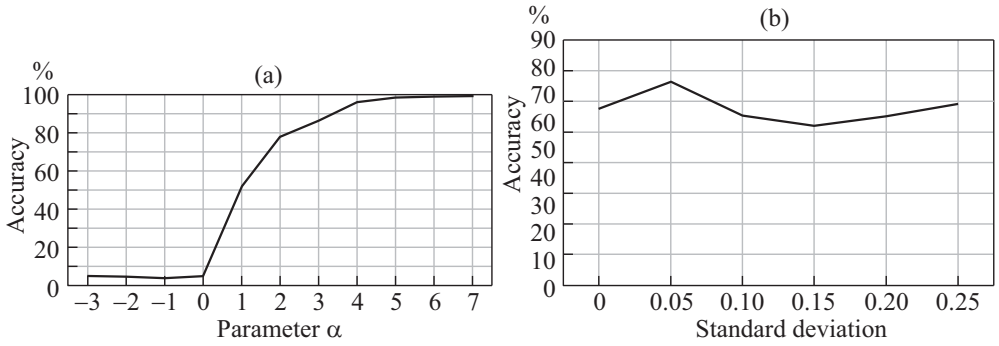


Fig. 11. Influence of algorithm parameters on the accuracy of circle detection: (a) radial stiffness parameter; (b) standard deviation.

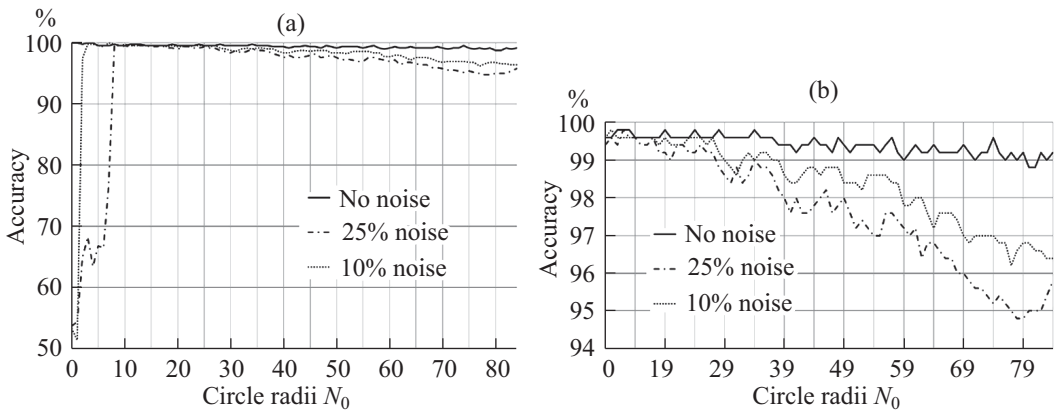


Fig. 12. Influence of impulse noise on the accuracy of circle recognition for different radii: (a) for all circles; (b) for circles with a radius greater than 10 pixels.

4.3. Influence of Noise on Recognition Accuracy

The turbidity of the water, the presence of slight organic and inorganic inclusions, and distortions caused by the refraction of light between the water and the camera lens are mathematically described as additive noise. In particular, a refined solid suspension in water is modeled as an impulse noise of the "salt" and "pepper" types. The effects of distortion and turbidity were modeled using masks as Gaussian and "blurring" noises. The impact on the recognition of each type of noise is discussed below.

4.3.1. Influence of Impulse Noise

To determine the effect of impulse noise on the algorithm's operation, 84 images of circles with radii from 1 to 84 pixels were examined. The original image was noisy, with noise at 25% intensity and 10% of the image area. The results of the algorithm are shown in Fig. 12.

It can be seen that impulse noise has a significant effect on objects of small sizes. Its influence on the result of the recognition of larger objects is less significant.

4.3.2. Influence of Gaussian Noise

To study the effect of Gaussian noise on the operation of the algorithm, we imposed noise on the original image by filling image pixels with normally distributed random numbers with mathematical expectation $\mu = 0$ and standard deviation $\sigma = 5$. Then the image was processed by

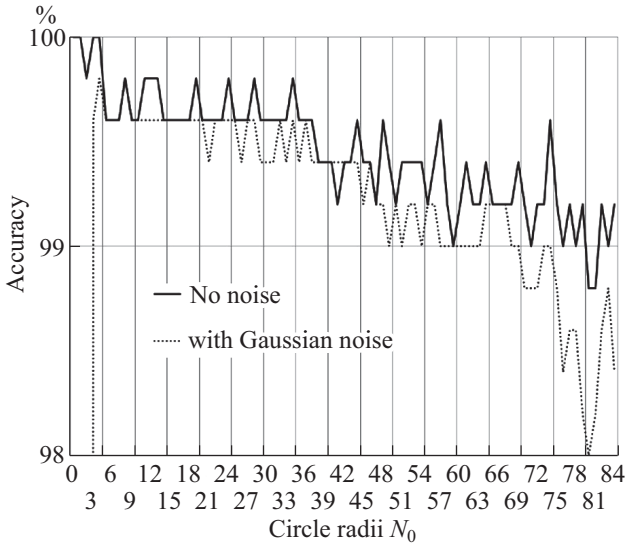


Fig. 13. Influence of Gaussian noise on recognition accuracy circles for various radii.

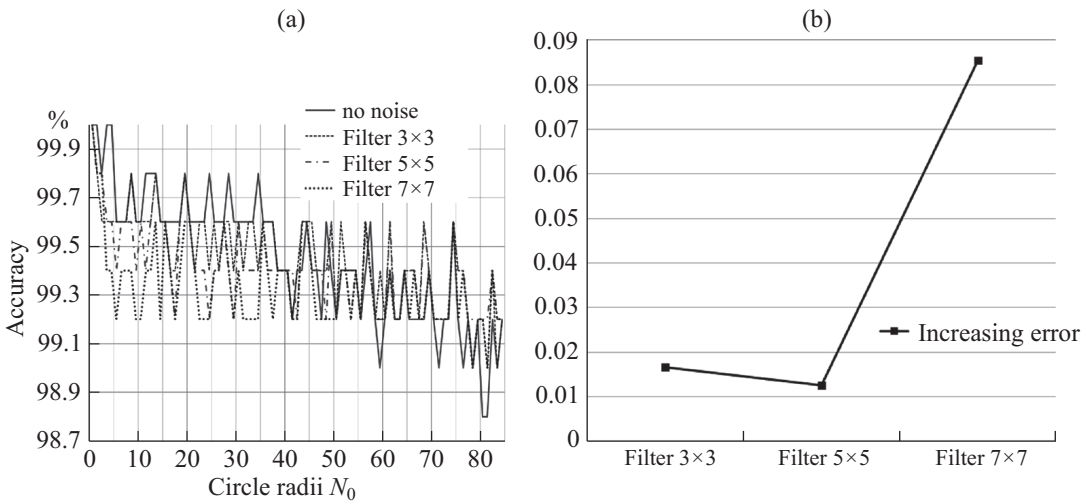


Fig. 14. Influence of image blur by filters on circle recognition accuracy: (a) comparative accuracy for various blur filters; (b) the relative average increase in the recognition error for various filters compared to the undistorted image.

a radial symmetry detector with a radial stiffness parameter α, equal to 2. The results are presented in Fig. 13.

Gaussian noise has a significant effect on small objects. For large objects, the effect of this kind of noise is insignificant.

4.3.3. Influence of Blur-Type Noise

To study the influence of noise of the "blur" type, we subjected the original image to blurring with filters of dimensions 3×3, 5×5 and 7×7. The results of changing the accuracy of circle recognition are shown in Fig. 14.

The blur-type noise practically does not affect the result of object detection. The loss of accuracy is no more than 0.09%.

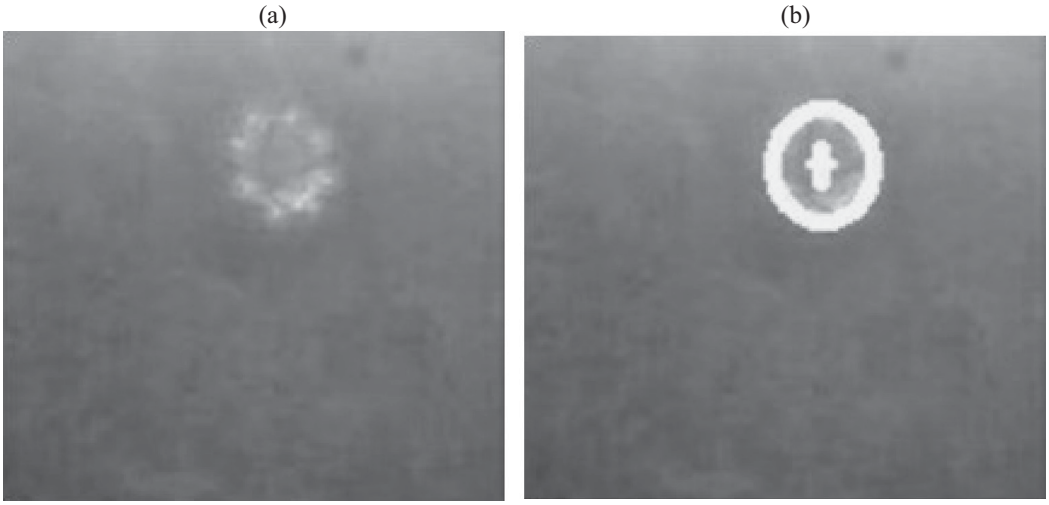


Fig. 15. Light Marker Detection Example: (a) the original image of the marked ring of the docking station at an angle; (b) a recognized circle with a calculated center.

An interesting fact is the increase in recognition accuracy in the presence of "blur" type noise compared to a non-noisy image in several experiments. This is due to the problem of false detection when the algorithm erroneously refers to the found circles of a given radius and also closely spaced circles of other, close in value, radii. With blur noise applied, the probability of such a false positive detection is reduced in some cases.

The obtained results of studying the influence of various types of noise on the accuracy of recognition of radial images show that it is not required to apply additional sharpening algorithms since the developed method is sufficiently resistant to noise within the framework of the problem being solved.

4.4. Image Detection

The image of a light docking ring in turbid water refers to images with a high degree of blurring distortion and, to a lesser extent, Gaussian and impulse noise. Thus, for objects with a radius of more than 10 pixels (large objects), the developed algorithm provides high resistance to all types of noise, and according to the main distorting criterion, blur, it provides stability for recognizing an underwater docking luminescent ring of all sizes.

An example of the image of a light marker in turbid water at an angle of 16° o the camera with a recognized docking center is shown in Fig. 15.

In the course of numerical experiments, both the circles themselves were recognized, and their centers were found, the coordinates of which can potentially serve as the coordinates of the final optimal position of the unmanned vehicle during automatic control. For the image presented in Fig. 15, in particular, we have:

Original image parameters:

(a) a color image with a dimension of 448 × 448 pixels in *.jpeg format;

(b) the number of desired circles is 1, the desired radius is 85 pixels;

(c) the set of acceptable colors — shades of yellow;

(d) allowable mismatch between the positions of the centers of the circles — 5 pixels;

found a circle centered at a point:

- $x_{target} = 282$;
- $y_{target} = 298$.

Table 2. Comparison of recognition accuracy for radial objects with a uniform increase in radii

Minimum distance between circle centers, pix.	The actual number of circles in the image	Found by the Hough method	Found by the developed algorithm
5	7	21	8
10	7	7	7
15	7	7	7
20	7	7	7
25	7	7	7
30	7	7	7
35	7	7	7
40	6	7	6
45	7	8	8
50	6	6	6

Table 3. Comparison of recognition accuracy for radial objects with a random change in radii

Minimum distance between circle centers, pix.	The actual number of circles in the image	Found by the Hough method	Found by the developed algorithm
5	8	152	12
10	8	53	8
15	8	30	8
20	8	22	8
25	8	14	8
30	8	10	8
35	8	10	8
40	8	10	8
45	8	8	8
50	8	6	8

4.5. Comparison of the Efficiency of Algorithms

In this section, we will be comparing the accuracy and speed of the developed algorithm with the Hough and fast radial symmetry algorithms that were used separately. It's worth noting that methods such as neural network recognition require additional computational costs and memory costs of the computing device at the training stage. This is mainly because of the need to have a relevant training dataset that includes all possible combinations of input parameters to obtain an adequate model. Due to this, the use of such machine-learning algorithms is unsuitable for low-performance computing devices that are installed on board HUAUVs or similar unmanned vehicles.

4.5.1. Comparison of Accuracy of Radial Objects Detection

To illustrate the method's effectiveness for detecting radial objects in images based on the proposed algorithm, we conducted two computational experiments: images of circles with uniform and random changes in radii.

Tables 2 and 3 show the results of comparing the accuracy of the algorithms when varying the minimum distance between the centers of the circles. An image is considered correctly recognized if the deviation of the found centers from the real ones does not exceed 15 pixels.

Table 4. Characteristics of image datasets for experiments comparing the speed of algorithms

Dataset name	Short description
LISA Traffic Light Dataset	100 images from a set of size 1280×960 in JPG format
Bosch Small Traffic Lights Dataset	PNG-format images from a set of various sizes
Selective set from Kaggle	200 PNG-format images from a set of various sizes

Table 5. Comparison of the speed of recognition of radial objects of different ranges of circles

Range of circles for recognition	FRST algorithm running time, ms	Proposed algorithm running time, ms	Speed gain of the developed algorithm, times
1–10	89 988	50 569	1.8
1–20	171 902	79 753	2.1
1–30	253 435	111 020	2.3
1–50	465 673	128 970	3.6

The comparison results show that the accuracy of the developed algorithm is, on average, 35% higher than the accuracy of classical methods used separately.

4.5.2. Comparison of the Speed of Radial Objects Detection

For a relative comparison of the speed of the developed algorithm, we used a computer with the following characteristics:

- processor: Intel Core i5-3230M 2.60 GHz;
- RAM: 8 GB;
- operating system: Windows 10, 64-bit.

The experiments were carried out for three test sets of images: 2 complete datasets, "LISA Traffic Light Dataset" and "Bosch Small Traffic Lights Dataset" of the Kaggle open type dataset, as well as 200 sample images from various Kaggle datasets relevant to the study. A description of the test set of images from these databases is given in Table 4.

We conducted computational experiments on circle recognition in four ranges: from 1 to 10 pixels, from 1 to 20 pixels, from 1 to 30 pixels, and from 1 to 50 pixels. Table 5 shows the results of estimating the speed of the algorithm proposed by us in comparison with the primary method based on the radial symmetry transformation FRST.

It is evident that with the increase in the range of desired circles, the advantage of the proposed method's speed increases compared to the basic one.

It should be noted that the computer used for the experiments had excess power for the practical implementation of the proposed algorithm and was used solely to compare the author's algorithm with the classical one. In natural conditions, the power of HUAUV onboard computers is sufficient for the algorithm due to the simplicity of the implemented calculations.

5. CONCLUSION

We have proposed a method that can be used to detect and localize objects of a given shape and color in images of light markers. For example, our method can be used to detect an object of interest in an unmanned underwater or aerial underwater vehicle. The described method is easy to implement and resistant to interference. It can be supplemented with various image processing operations at all stages to improve performance, vital for detecting an object in an aquatic environment with reduced visibility.

In addition, it should be noted that the procedure for generating a multiscale image weight model has the properties of natural parallelism. This determines the possibility of increasing the processing speed through hardware and software tools for high-performance parallel computing, essential for effective real-time control of an unmanned vehicle and reducing power consumption.

The limitation of the algorithm's applicability in water turbidity conditions is the size of the detected radial object. In particular, objects with a radius of less than 10 pixels are difficult to detect by the developed algorithm. Larger objects are recognized with a high accuracy of over 95%. In this case, the radial stiffness parameter α should equal +4. For effective detection of objects, the level of impulse noise in images should not exceed 25%, while Gaussian noise has little effect on detecting objects whose radius exceeds 5 pixels. The developed algorithm is resistant to blurring noise with averaging filters. Because of the preceding, the algorithm does not require additional methods to improve image clarity, which is its undeniable advantage over analogs.

This advantage is obtained through the joint application of several basic algorithms, such as the Hough method, the fast radial symmetry transformation, and the construction of the Gaussian pyramid. The resulting unified model outperforms each of the basic algorithms individually and, when applied in pairs, in performance and computing resource requirements.

REFERENCES

1. Liguo, T., Shenmin, S., Xiaoyan, Y., and Jianwen, S., An Overview of Marine Recovery Methods of UAV for Small Ships, *J. Harbin Institute Technol.*, 2019, vol. 51, no. 10, pp. 1–10.
2. Chamola, V., Kotesh, P., Agarwal, A., Naren, Gupta, N., and Guizani, M., A Comprehensive Review of Unmanned Aerial Vehicle Attacks and Neutralization Techniques, *Ad Hoc Networks*, 2021, vol. 111, p. 102324. https://doi.org/10.1016/j.adhoc.2020.102324
3. State Standard 56829-2015. Intelligent transport systems. Terms and Definitions. Moscow, Standartinform Publ. 10 december 2015. 10 p. (In Russian)
4. Albeaino, G., Gheisari, M., and Franz, B., A Systematic Review of Unmanned Aerial Vehicle Application Areas and Technologies in the AEC Domain, *J. of Inform. Technol. Construct.*, 2019, vol. 24, pp. 381–405.
5. Hajiyev, C., Soken, H.E., and Vural, S.V., Navigation Systems for Unmanned Aerial Vehicles, *State Estimation and Control for Low-cost Unmanned Aerial Vehicles*, 2015, pp. 25–49. https://doi.org/10.1007/978-3-319-16417-5_3
6. Moiseev, V.S., Shafigullin, R.R., and Gushchina, D.S., Rational Placement and Required Number of Information Unmanned Aerial Systems for On-Line Monitoring of Large Territories, *Russian Aeronautics*, 2012, vol. 55, pp. 223–229. https://doi.org/10.3103/S1068799812030014
7. Neira, J., Sequeiros, C., Huaman, R., Machaca, E., Fonseca, P., and Nina, W., Review on Unmanned Underwater Robotics, Structure Designs, Materials, Sensors, Actuators, and Navigation Control, *J. Robot.*, 2021, pp. 1–26. https://doi.org/10.1155/2021/5542920
8. Kolesnikov, M.P., Martynowa, L.A., Pashkevich, I.V., and Shelest, P.S., Method of AUV Positioning While Homing to the Docking Station, *Izvestiya Tula State University. Technical science*, 2015, no. 11–2, pp. 38–49. (In Russian)
9. Qiu, S. and Cui, W., An Overview on Aquatic Unmanned Aerial Vehicles, *Ann. Rev. Res.*, 2019, vol. 5, no. 3, p. 555663. https://doi.org/10.19080/ARR.2019.05.555663
10. Popescu, D. and Ichim, L., Image Recognition in UAV Application Based on Texture Analysis, *Advanced Concepts for Intelligent Vision Systems. ACIVS 2015. Lecture Notes in Computer Science*, 2015, vol. 9386. https://doi.org/10.1007/978-3-319-25903-1_60
11. Samadzadegan, F., Dadrass Javan, F., Ashtari Mahini, F., and Gholamshahi, M., Detection and Recognition of Drones Based on a Deep Convolutional Neural Network Using Visible Imagery, *Aerospace*, 2022, vol. 9, no. 1, p. 31. https://doi.org/10.3390/aerospace9010031

12. Fujiyoshi, H., Hirakawa, T., and Yamashita, T., Deep Learning-Based Image Recognition for Autonomous Driving, *IATSS Res.*, 2019, vol. 43, no. 1, pp. 244–252. https://doi.org/10.1016/j.iatssr.2019.11.008

13. Sevostyanov, I.E. and Devitt, D.V., Sual Positioning System for Multi-Rotor Drones for High-Precision Autonomous Landing, *Science Time*, 2021, no. 90, pp. 38–42. (In Russian)

14. Stepanov, D.N., Methods and Algorithms for Determining the Position and Orientation UAV Using on-Board Videocameras, *Software Products and Systems*, 2014, no. 1, pp. 150–157. (In Russian)

15. Deltheil, C., Didier, L., Hospital, E., and Brutzman, D.P., Simulating an Optical Guidance System for the Recovery of an Unmanned Underwater Vehicle, *IEEE J. Ocean. Engineer.*, 2000, vol. 25, no. 4, pp. 568–574. https://doi.org/10.1109/48.895364

16. Guo, D., Bacciaglia, A., Simpson, M., Bil, C., and Marzocca, P., Design and Development a Bimodal Unmanned System, *AIAA Scitech 2019 Forum*, 2019, pp. 1–7. https://doi.org/10.2514/6.2019-2096

17. Pinheiro, P.M., Neto, A.A., Grando, R.B., Silva, C.B. da, Aoki Vivian, M., Cardoso, D.S., Horn, A.C., and Drews, P.L.J., Trajectory Planning for Hybrid Unmanned Aerial Underwater Vehicles with Smooth Media Transition, *J. Intelligent Robot. Syst.*, 2022, vol. 104, no. 46. https://doi.org/10.1007/s10846-021-01567-z

18. Lock, R.J., Vaidyanathan, R., Burgess, S.C., and Loveless, J., Development of a Biologically Inspired Multi-Modal Wing Model for Aerial-Aquatic Robotic Vehicles through Empirical and Numerical Modelling of the Common Guillemot, Uria Aalge, *Bioinspirat. Biomimetics*, 2010, vol. 5, no. 4, pp. 1–15. https://doi.org/10.1088/1748-3182/5/4/046001

19. Wu, Y., Li, L., Su, X., and Gao, B., Dynamics Modeling and Trajectory Optimization for Unmanned Aerial-Aquatic Vehicle Diving into the Water, *Aerospace Sci. Technol.*, 2019, vol. 89, pp. 220–229. https://doi.org/10.1016/j.ast.2019.04.004

20. Liu, S., Ozay, M., Okatani, T., Xu, H., Sun, K., and Lin, Y., Detection and Pose Estimation for Short-Range Vision-Based Underwater Docking, *IEEE Access.*, 2019, vol. 30, no. 7, pp. 2720–2749. https://doi.org/ 10.1109/ACCESS.2018.2885537

21. Cowen, S., Briest, S., and Dombrowski, J., Underwater Docking of Autonomous Undersea Vehicles Using Optical Terminal Guidance, *Oceans '97. MTS/IEEE Conference Proceedings*, Halifax, NS, Canada, 1997, vol. 2, pp. 1143–1147. https://doi.org/10.1109/OCEANS.1997.624153

22. Negre, A., Pradalier, C., and Dunbabin, M., Robust Vision-Based Underwater Homing Using Self-Similar Landmarks, *J. Field Robot.*, 2008, vol. 25, no. 6–7, pp. 360–377. https://doi.org/10.1002/rob.20246

23. Ghosh, S., Ray, R., Vadali, S.R.K., Shome, S.N., and Nandy, S., Reliable Pose Estimation of Underwater Dock Using Single Camera: A Scene Invariant Approach, *Machine Vision Appl.*, 2016, vol. 27, no. 2, pp. 221–236. https://doi.org/10.1007/s00138-015-0736-4

24. Li, Y., Jiang, Y., Cao, J., Wang, B., and Li, Y., AUV Docking Experiments Based on Vision Positioning Using Two Cameras, *Ocean Engineer.*, 2015, vol. 110, pp. 163–173.

25. Chichkarev, E., Sergienko, A., and Balalaeva, E., Using Machine Learning Models and Deep Learning Networks to Recognize Handwritten Numbers and Letters of the Russian and Latin Alphabets, *Inter-Conf.*, Prague, Czech Republic, 2021, pp. 363–380. https://doi.org/10.51582/interconf.21-22.11.2021.044 (In Russian)

26. Sirota, A.A., Mitrofanova, E.Yu., and Milovanova, A.I., Analysis of Algorithms for Searching Objects in Images Using Various Modifications of Convolutional Neural Network, *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*, 2019, no. 3, pp. 123–137. https://doi.org/10.17308/sait.2019.3/1313 (In Russian)

27. Zou, Z., Shi, Z., Guo, Y., and Ye, J., Object Detection in 20 Years: A Survey, *Proceedings of the IEEE*, 2019, no. 111, pp. 257–276.

28. Kletenik, D.V., *Sbornik zadach po analiticheskoi geometrii* (Collection of Problems in Analytic Geometry), Moscow: Nauka, 1998.

29. Gonzalez, R.C. and Woods, R.E., *Digital Image Processing*, 3rd ed., London: Pearson, 2007.

30. Hough, P.V.C., Machine Analysis of Bubble Chamber Pictures, *2nd International Conference on High-Energy Accelerators and Instrumentation, HEACC 1959*, CERN, Geneva, Switzerland, 1959, pp. 554–558.

31. Lyasheva, S., Shleymovich, M., and Shakirzyanov, R., The Image Analysis Using Fast Radial Symmetry Transform in Control Systems Base on the Computer Vision, *International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon*, 2019, pp. 1–6. https://doi.org/10.1109/FarEastCon.2019.8934298

32. Shakirzyanov, R.M., Detection of Traffic Signals Using Color Segmentation and a Radial Symmetry Detector, *Bulletin of the Voronezh State Technical University*, 2020, vol. 16, no. 6, pp. 25–33.

33. Adelson, E., Burt, P., Anderson, C., Ogden, J.M., and Bergen, J., Pyramid Methods in Image Processing, *RCA Engineer.*, 1984, vol. 29, no. 6, pp. 33–41.

34. O'Malley, R., Jones, E., and Glavin, M., Rear-Lamp Vehicle Detection and Tracking in Low-Exposure Color Video for Night Conditions, *IEEE Transactions on Intelligent Transportation Systems*, 2010, vol. 11, no. 2, pp. 453–462. https://doi.org/10.1109/TITS.2010.2045375

35. Druki, A.A., Algorithms for Selecting Faces on Static RGB Images and in a Video Stream, *Izvestiya TPU*, 2012, no. 5, pp. 65–69. (In Russian)

36. Darge, A., Rajendran, R.S., Zerihum, D., and Chung, P.Y.K., Multi Color Image Segmentation using L*A*B* Color Space, *Int. J. Advanced Engineer., Management Sci.*, 2019, vol. 5, pp. 346–352. https://doi.org/10.22161/ijaems.5.5.8

37. Forsyth, D. and Ponce, J., *Computer Vision: A Modern Approach*, London: Pearson, 2012.

38. Droogenbroeck, V.M. and Barnich, O., Design of Statistical Measures for the Assessment of Image Segmentation Schemes, *Proceedings of 11th International Conference on Computer Analysis of Images and Patterns (CAIP2005), Lecture Notes in Computer Science*, Rocancourt, France, 2005, vol. 3691, pp. 280–287.

39. Koltsov, P.P., Osipov, A.S., Kutsaev, A.S., Kravchenko, A.A., Kotovich, N.V., and Zakharov, A.V., On the Quantitative Performance Evaluation of Image Analysis Algorithms, *Computer Optics.*, 2015, vol. 39, no. 4, pp. 542–556. (In Russian)

This paper was recommended for publication by O.A. Stepanov, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 852–871.

Statistical Complexity as a Criterion for the Useful Signal Detection Problem

A. A. Galyaev[*,a], **P. V. Lysenko**[*,b], **and L. M. Berlin**[*,c]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a] *galaev@ipu.ru,* [b] *pavellysen@ipu.ru,* [c] *berlin.lm@phystech.edu*

Received March 21, 2023
Revised April 19, 2023
Accepted May 11, 2023

Abstract—Three variants of the statistical complexity function, which is used as a criterion in the problem of detection of a useful signal in the signal-noise mixture, are considered. The probability distributions maximizing the considered variants of statistical complexity are obtained analytically and conclusions about the efficiency of using one or another variant for detection problem are made. The comparison of considered information characteristics is shown and analytical results are illustrated on an example of synthesized signals. A method is proposed for selecting the threshold of the information criterion, which can be used in decision rule for useful signal detection in the signal-noise mixture. The choice of the threshold depends a priori on the analytically obtained maximum values. As a result, the complexity based on the total variation demonstrates the best ability of useful signal detection.

Keywords: statistical complexity, signal detection, information divergence

DOI: 10.25728/arcRAS.2023.27.32.001

1. INTRODUCTION

The concept of information entropy was firstly introduced in Claude Shannon's article [1] in 1948. This work marked the beginning of a new field of science called information theory [2]. The development of information theory made possible an analytical and practical research in many applied fields of science and technology. Such terms as Gibbs and von Neumann entropies, Kullback–Leibler distance, Jensen–Shannon divergence, information divergences and some others were introduced and interpreted and later began to serve as criteria for various optimization problems of recognition [3], classification [4] and filtering.

By the end of the last century various information criteria, mainly Shannon information entropy, had began to be actively applied in the tasks of digital signal processing, in particular in the problem of detection of a useful signal in a noise environment [5]. The concept of spectral entropy [6], associated with the Fourier spectrum of the considered signal, has appeared and proved to be especially relevant in the analysis of acoustic signals [7]. In addition, the entropic approach has been successfully applied in the analysis of time series in the medical field, such as ECG or EEG [8]. Later, a statistical complexity function was proposed as a development of the entropy concept [9–11]. However, the articles mostly do not provide an analytical study of the properties of these functions, which turns out to be especially important when solving the problem of hypothesis testing.

It should be noted that there are several classical ways of solving the detection problem. The first of them is based on solving the problem of optimal filtering and requires knowledge of the properties of the signal: periodicity, bandwidth, etc. [12]. The second way is based on the Neumann–Pearson lemma, solves the problem of hypothesis testing, and determines the fact of exceeding the optimal threshold at a given false alarm probability and requires estimation of statistical properties of sample distributions of noise and mixture of signal and noise [13]. The third way is equivalent to solving the changepoint detection problem when the unknown statistical characteristics of the signal distributions change. The anomaly detection problem [14] has a similar formulation. All these methods demonstrate qualitative and reliable performance when the signal exceeds the noise, but for small signal-to-noise ratios often give the wrong answer.

The article is devoted to the problem of detection of useful signal in the signal-noise mixture and combines all three previously listed ways of solving the detection problem. We propose to use a variant of the Neumann–Pearson lemma for the problem of hypothesis testing [13], which is indeed valid when the error probability is close to one and depends on the total variation of the measure of two distributions of the null and alternative hypotheses. Based on the analytical expression of this error function, the criterion of the signal detection problem is formalized as one of the variants of the statistical complexity [15], which takes into account the deterministic nature of the signal mixed in with the noise. The peculiarity of the statistical complexity is that it is multiplicative and consists of two multipliers, one of which is zero on deterministic sinusoidal signals of the same frequency (in physics these are objects of a given structure, such as crystals [9]) and the other is zero on uniform distribution functions [10], corresponding, for example, to white noise. Then the introduced criterion is compared with the already known two variants of statistical complexity based on Euclidean distance square and Jensen–Shannon divergence, their properties are established, and optimization as a function of many variables on a set of discrete distributions is performed. As a result, families of optimal distributions are identified and maxima of statistical complexity functions are calculated.

The article has the following structure. Section 1 provides a literature review and highlights the current state of research on the topic of the article. Section 2 is devoted to the connection of the considered information criteria with the classical criterion of the signal detection problem. Section 3 investigates the properties of the three types of statistical complexity. In 4 the analytical results of the previous section are supported by numerical simulations for synthesized signals. Section 5 summarizes the results obtained in the paper and lists plans for the future.

2. NEYMAN-PEARSON LEMMA AND STATISTICAL COMPLEXITY

The problem of signal detection $s(n)$ is traditionally reduced to the problem of hypothesis testing

$$\begin{cases} \Gamma_0 : x(n) = w(n), \\ \Gamma_1 : x(n) = s(n) + w(n), \ n = 1, \ldots, N. \end{cases}$$

Hypothesis Γ_0 corresponds to the decision of receiving only noise, and hypothesis Γ_1 — of receiving a mixture of useful signal and noise, where the sequences $\{x(n)\},\ n = 1, \ldots, N$ are time series of the received data, $\{s(n)\}$ — useful signal, $\{w(n)\}$ — additive white Gaussian noise, N — the length of the time series of data.

The random variables of the time series $(x(1), \ldots, x(n), \ldots, x(N))$ take values $(x_1, \ldots, x_n, \ldots \ldots, x_N) \in \mathbb{R}^N$.

In order to obtain an analytical expression for estimating the error probability in hypothesis testing, we can apply a variant of the Neyman–Pearson lemma [13, 16].

Lemma 1 Neyman–Pearson. *Let there be an arbitrary, called a decision rule or test, measurable function of many variables* $(x_1, \ldots, x_N) \in \mathbb{R}^N$ *such that*

$$d(x_1, \ldots, x_N) = \begin{cases} 1, & \textit{hypothesis } \Gamma_0 \textit{ is true}, \\ 0, & \textit{hypothesis } \Gamma_1 \textit{ is true}, \end{cases}$$

by which the following probabilities can be determined:

$$\alpha(d) = \textit{Probability } (\textit{accept } \Gamma_0 | \Gamma_1 \textit{ is true}),$$
$$\beta(d) = \textit{Probability } (\textit{accept } \Gamma_1 | \Gamma_0 \textit{ is true}).$$

Then the decision rule d^* *is optimal if*

$$\alpha(d^*) + \beta(d^*) = \inf_d [\alpha(d) + \beta(d)] = \mathcal{E}r(N; \Gamma_0, \Gamma_1) \text{ — } \textit{error function}, \tag{1}$$

where the infinum is taken for all tests.

Here $\alpha(\cdot)$ is the probability of a false alarm, and $\beta(\cdot)$ is the probability of a useful signal missing.

The exact formula for the error function is as follows:

$$\mathcal{E}r(N; \Gamma_0, \Gamma_1) = 1 - \frac{1}{2}\|P_0^{(N)} - P_1^{(N)}\| = 1 - TV(P_0, P_1), \tag{2}$$

where $P_0^{(N)}$ is the multivariate distribution function of the observation statistics by hypothesis Γ_0, $P_1^{(N)}$ is the multivariate distribution function of the observation statistics by hypothesis Γ_1, and $TV(P_0, P_1)$ is the total variation of the signed measure, $\|Q\| = 2\sup_A |Q(A)|$. Thus, if the supports of measures P_0, P_1 do not overlap, then error-free distinguishing of hypotheses is possible. If the measures $P_0^{(N)}$ and $P_1^{(N)}$ are close, then $\|P_0^{(N)} - P_1^{(N)}\| \approx 0$, leading to $\mathcal{E}r(N; \Gamma_0, \Gamma_1) \approx 1$.

For the problem of detecting a deterministic useful signal, for example, at a small signal-to-noise ratio, the case $\|P_0^{(N)} - P_1^{(N)}\| = 2TV(P_0, P_1) \approx 0$ is of interest and the possibility of reasonable estimation of this value. Therefore, when the probability of the total error of distinguishing two hypotheses is close to one, it becomes possible to use the analytical expression $TV(P_0, P_1)$ to design a criterion in the problem of detecting a useful signal in a mixture. But first let us turn to already known criteria and establish their properties.

Most often, for the convenience of mathematical investigation, both of the useful signal and noise are modeled by Gaussian random processes with different parameters. In that case the problem of finding the moment of appearance of the signal $s(n)$ in the received sequence of samples is called the the problem of changepoint detection [13].

Here and below, we consider discrete probability distributions $p = (p_1, \ldots, p_i, \ldots, p_N)$, that by definition have the following properties:

$$\forall\ p_i \in [0, 1], \quad \sum_{i=1}^{N} p_i = 1. \tag{3}$$

To formalize criterion that takes into account the deterministic component of the signal as well as the random one, let us explore the concepts of disequilibrium function D and statistical complexity C of the distribution. The simplest example of the disequilibrium function is the square of Euclidean distance in the space of discrete probability distributions [10].

Definition 1. The disequilibrium D_{SQ} has the meaning of the variance of a distribution relative to a uniform distribution

$$D_{SQ}(p) = \sum_{i=1}^{N} \left(p_i - \frac{1}{N}\right)^2 = \sum_{i=1}^{N} p_i^2 - \frac{1}{N}. \tag{4}$$

Definition 2. The statistical complexity, defined through the expression of disequilibrium by the Definition 1, is equal to

$$C_{SQ}(p) = H(p) \times D_{SQ}(p), \tag{5}$$

where

$$H(p) = \frac{1}{\log N}\left(-\sum_{i}^{N} p_i \log p_i\right) \tag{6}$$

— Shannon's normalized entropy [1].

In evaluating the sum (6), it is assumed that $\frac{0}{\log 0} = 0$ by continuity, and this assumption holds for all subsequent equations.

It follows from the Definition 1 that disequilibrium of the form (4) and complexity of the form (5) are convenient to apply in estimation and comparison of signals having spectral distribution close to uniform. In general, instead of a uniform distribution $q_i = 1/N$ at $i = 1, \ldots, N$, the formula (4) may include an arbitrary discrete distribution.

The formula (4) is proposed in [10] for computing the disequilibrium with respect to a uniform distribution, but most studies use the Jensen–Shannon divergence $JSD(p||q)$ [17] instead.

Definition 3. The Jensen–Shannon disequilibrium equals

$$D_{JSD}(p) = JSD(p||q), \tag{7}$$

where $q = (1/N, \ldots, 1/N)$ is the uniform distribution.

Definition 4. Statistical complexity defined through the expression of disequilibrium from Definition 3, is expressed as

$$C_{JSD}(p) = H(p) \times D_{JSD}(p). \tag{8}$$

Remark 1. It was noted above that $\sqrt{D_{SQ}}$ is a Euclidean metric on the space of discrete distributions. At the same time $\sqrt{D_{JSD}}$ is also a metric which is proportional to the Fisher metric.

Since the error function of distinguishing between two hypotheses depends on the total variation $TV(p, q)$, which is obtained in the Neyman–Pearson lemma 1, we introduce another notion of disequilibrium.

Definition 5. The disequilibrium based on the total variation of signed measure is equal to

$$D_{TV}(p) = TV^2(p, q), \tag{9}$$

where $q = (1/N, \ldots, 1/N)$.

Definition 6. The statistical complexity, defined through the disequilibrium expression according to the Definition 5, is equal to

$$C_{TV}(p) = H(p) \times D_{TV}(p). \tag{10}$$

The information divergence functions presented above, which define different variants of the disequilibrium function, can be unified by the general concept of *f-divergence* [18]:

$$D_f(p||q) = \sum_{x \in \mathbb{R}^N} q(x) f\left(\frac{p(x)}{q(x)}\right). \tag{11}$$

The choice of function f gives rise to a whole family of different divergences:

- The Kulback–Leibler divergence $D_{KL}(p,q)$ is obtained from (11) by choosing $f(x) = x\log(x)$, $x > 0$.
- The Jensen–Shannon divergence is obtained from (11) by choosing

$$f(x) = x\log\frac{2x}{x+1} + \log\frac{2}{x+1}, \quad x > 0. \tag{12}$$

- The total variation is obtained when $f(x) = \frac{1}{2}|1-x|$:

$$TV(p,q) = \frac{1}{2}\sum_{x\in\mathbb{R}^N}|p(x) - q(x)|; \tag{13}$$

$TV(p,q)$ is also a metric on the space of probability distributions. The total variation is related to the Jensen–Shannon divergence by the following relation:

$$JSD(p||q) \leqslant TV(p,q). \tag{14}$$

It follows from the inequality (14) that the total variation is the upper bound of Jensen–Shannon divergence.

Next, let us investigate the possibility of using each variant of statistical complexity as a criterion for indicating the appearance of a signal, but at first their properties must be established.

3. STATISTICAL COMPLEXITY OPTIMIZATION

3.1. Optimization of C_{SQ}

Let us formulate the problem of maximizing the statistical complexity function on the set of discrete distributions $p = (p_1, \ldots, p_N)$

$$C_{SQ}(p) = \frac{1}{\log N}\left(-\sum_{i=1}^{N} p_i\log p_i\right)\left(\sum_{i=1}^{N}\left(p_i - \frac{1}{N}\right)^2\right) \longrightarrow \max_p \tag{15}$$

with the condition

$$\sum_{i=1}^{N} p_i = 1. \tag{16}$$

An auxiliary result will be needed to formulate the lemma about the maximum value of statistical complexity.

Lemma 2. *Let $0 < x \leqslant y \leqslant z \leqslant 1$, then $f(x,y,z) = x^y y^{-x} z^x x^{-z} y^z z^{-y} \geqslant 1$, with equality possible only when either $x = y$ or $y = z$.*

Proof. Let us introduce a new function $g(x,y,z) = \ln f(x,y,z)$,

$$g(x,y,z) = y\ln x - x\ln y + x\ln z - z\ln x + z\ln y - y\ln z.$$

Then it is required to prove that $g(x,y,z) \geqslant 0$, $\quad 0 < x \leqslant y \leqslant z \leqslant 1$.

By the Kuhn–Tucker theorem, the solution of the conditional optimization problem of a function of three variables is either at the interior point of the constraint manifold or at its boundary. The

necessary conditions for the unconditional extremum of the function $g(x,y,z)$ take the following form

$$\begin{aligned}\frac{\partial g}{\partial x} &= \ln z - \ln y + \frac{y-z}{x} = 0,\\ \frac{\partial g}{\partial y} &= \ln x - \ln z + \frac{z-x}{y} = 0,\\ \frac{\partial g}{\partial z} &= \ln y - \ln x + \frac{x-y}{z} = 0.\end{aligned} \tag{17}$$

Let us summarize all the equations of the last system:

$$\frac{y-z}{x} + \frac{z-x}{y} + \frac{x-y}{z} = 0,$$

which can be rewritten as

$$\frac{(y-z)(x-y)(z-x)}{xyz} = 0.$$

This means that when one of the equalities either $x=y$ or $y=z$ is satisfied, the function $g(x,y,z)$ possibly has a minimum. Let $x=y$, then the third equation from (17) is fulfilled identically, and the first and second equations are identical and can be written as

$$\ln\eta = \eta - 1, \quad \eta = \frac{z}{y}.$$

The last equation has only one root $\eta = 1$, i.e., $y=z$.

Let us calculate the second derivatives and write the Hesse matrix:

$$G(x,y,z) = \begin{pmatrix} \dfrac{z-y}{x^2} & \dfrac{1}{x}-\dfrac{1}{y} & \dfrac{1}{z}-\dfrac{1}{x} \\ \dfrac{1}{x}-\dfrac{1}{y} & \dfrac{x-z}{y^2} & \dfrac{1}{y}-\dfrac{1}{z} \\ \dfrac{1}{z}-\dfrac{1}{x} & \dfrac{1}{y}-\dfrac{1}{z} & \dfrac{y-x}{z^2} \end{pmatrix}. \tag{18}$$

Minors of the Hesse matrix are equal to

$$\begin{aligned} M_1(x,y,z) &= \frac{z-y}{x^2},\\ M_2(x,y,z) &= -\frac{(x-y)^2+(z-x)^2+(y-z)^2}{2x^2y^2},\\ M_3(x,y,z) &= 0.\end{aligned} \tag{19}$$

The Hesse matrix is not sign-defined, moreover, its determinant equals zero. Therefore, let us consider a small vicinity of the extremum point.

In a small vicinity $x=y=z$, provided that $\delta x \leqslant \delta y \leqslant \delta z$, the variation δg of the function $g(x,y,z)$ is written in the form of

$$\begin{aligned}\delta g &= (x+\delta y)\ln(x+\delta x) - (x+\delta x)\ln(x+\delta y) + (x+\delta x)\ln(x+\delta z)\\ &\quad - (x+\delta z)\ln(x+\delta x) + (x+\delta z)\ln(x+\delta y) - (x+\delta y)\ln(x+\delta z)\\ &= (\delta z-\delta x)(\delta y-\delta x)(\delta z-\delta y) + o(((\delta x)^2+(\delta y)^2+(\delta z)^2)^{3/2}) \geqslant 0,\end{aligned}$$

where values in the cubes of variations of the independent variables are nonzero, and the variation of the function $g(x, y, z)$ itself is positive by virtue of the lemma conditions. In the case when, for example, $\delta y = \delta x$, we have $g(x, y, z) \equiv 0$, and $f(x, y, z) \equiv 1$. Therefore, the extremum of the function is its non-strict minimum. ■

Lemma 3. *The maximum statistical complexity* (15) *is achieved on the distribution of the form*

$$\begin{cases} p_i = \dfrac{1 - p_{\max}}{N - 1}, \quad i = \overline{1, N} \setminus k, \\ p_k = p_{\max}, \end{cases} \tag{20}$$

where $p_{\max} = \text{const}$*, i.e., at the appearance of a single component of an arbitrary index* k *over the uniform distribution.*

Proof. Without loss of generality, let us assume $k = N$. From Eq. (16) one variable p_N from the set p_i can be expressed through all the others:

$$p_N = 1 - \sum_{i=1}^{N-1} p_i. \tag{21}$$

Let us rewrite the Eq. (15) in the form

$$C_{SQ} = -\frac{1}{\log N}\left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N\right)\left(\sum_{i=1}^{N-1}\left(p_i - \frac{1}{N}\right)^2 + \left(p_N - \frac{1}{N}\right)^2\right). \tag{22}$$

A necessary condition for the extremum of a function at an interior point of the domain (simplex 3) is that all partial derivatives of p_i are equal to zero:

$$\frac{\partial C_{SQ}}{\partial p_i} = 0, \quad i = 1, \ldots, N-1. \tag{23}$$

Substituting the function (22) into (23) gives (provided that $\frac{\partial p_N}{\partial p_i} = -1$):

$$\begin{aligned} \frac{\partial C_{SQ}}{\partial p_i} &= -\frac{1}{\log N}\left(\log p_i - \log p_N\right) \times \left(\sum_{i=1}^{N-1}\left(p_i - \frac{1}{N}\right)^2 + \left(p_N - \frac{1}{N}\right)^2\right) \\ &- \frac{2}{\log N}\left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N\right) \times (p_i - p_N) = 0, \quad i = 1, \ldots, N-1. \end{aligned} \tag{24}$$

In a more convenient form the equations can be rewritten as

$$\frac{\partial C_{SQ}}{\partial p_i} = \frac{1}{\log N}\left(-\log p_i + \log p_N\right) \times D + 2H \times (p_i - p_N) = 0, \quad i = 1, \ldots, N-1. \tag{25}$$

Let us write the difference of any two equations from the system above for indices i and j:

$$\frac{\partial C_{SQ}}{\partial p_i} - \frac{\partial C_{SQ}}{\partial p_j} = \frac{1}{\log N}\left(-\log p_i + \log p_j\right) \times D + 2H \times (p_i - p_j) = 0. \tag{26}$$

Given that the values of D and H are positive, the following equations can be constructed from the Eqs. (25) and (26), provided that the considered probabilities p_j, $j = 1, \ldots, N-1$ are not equal to p_N:

$$\frac{\log p_i - \log p_j}{\log p_N - \log p_j} - \frac{p_i - p_j}{p_N - p_j} = 0, \tag{27}$$

$$(p_N - p_j)\log p_i + (p_i - p_N)\log p_j + (p_j - p_i)\log p_N = 0, \tag{28}$$

$$p_i^{p_N - p_j} \times p_j^{p_i - p_N} \times p_N^{p_j - p_i} = 1. \tag{29}$$

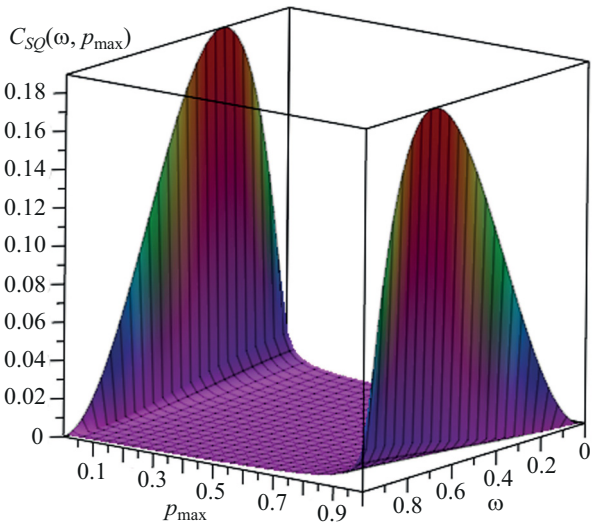


Fig. 1. Level surfaces of statistical complexity $C_{SQ}(\omega, p_{\max})$.

After applying the Lemma 2 we conclude that the last equation can be satisfied when $p_i = p_j$.

Thus, it is obtained that each of the probabilities p_i can take one of two different values, which define a distribution of the form

$$\begin{cases} p_i = \dfrac{1 - p_{\max}}{K}, & \forall\ i = 1, \ldots, K, \\ p_i = p_N = \dfrac{p_{\max}}{N - K}, & \forall\ i = K + 1, \ldots, N. \end{cases} \tag{30}$$

Now we need to show that the maximum complexity corresponds to values $K = 1$ and $K = N - 1$. For this purpose, let us calculate the value of the disequilibrium (4) on the distribution (30), which we denote by $D^{(K)}(\omega, p_{\max})$:

$$D^{(K)}(\omega, p_{\max}) = \frac{1}{N} \frac{(p_{\max} + \omega - 1)^2}{\omega(1 - \omega)}, \quad \omega = \frac{K}{N}. \tag{31}$$

In turn, entropy is equal to

$$H^{(K)}(\omega, p_{\max}) = 1 - \frac{1}{\log N} \left((1 - p_{\max}) \log \frac{1 - p_{\max}}{\omega} + p_{\max} \log \frac{p_{\max}}{1 - \omega} \right). \tag{32}$$

The maximum of $C_{SQ}(\omega, p_{\max})$ at $N \leqslant 100$ was investigated numerically, and it was reached at $K = 1$. From the expression for $D^{(K)}(\omega, p_{\max})$ (31), it can be seen that at $N \geqslant 101$ and when changing from $K = 1$ to $K = 2$ or from $K = N - 1$ to $K = N - 2$, its value changes by almost a factor of two, while the entropy (32) changes only slightly. Thus, the probability distribution (30) that delivers the complexity function to the extremum value at $K = 1$ or $K = N - 1$ is of the form (20). ■

For clarity, Fig. 1 shows the graph $C_{SQ} = C_{SQ}(\omega, p_{\max})$ at $N = 1024$, where ω is changing continuously (although K is changing discretely).

Corollary 1. *Let us substitute the values of p_i and $p_N = p_{\max}$ from* (20) *into* (22) *and consider the complexity C_{SQ} as a function of $p_{\max}$. For sufficiently large values of N, it will take the following form*

$$C_{SQ} \approx (1 - p_{\max}) \times p_{\max}^2.$$

Whence it follows that this function takes the maximum value $C_{SQ}^* \approx 4/27$ when $p_{\max} = 2/3$.

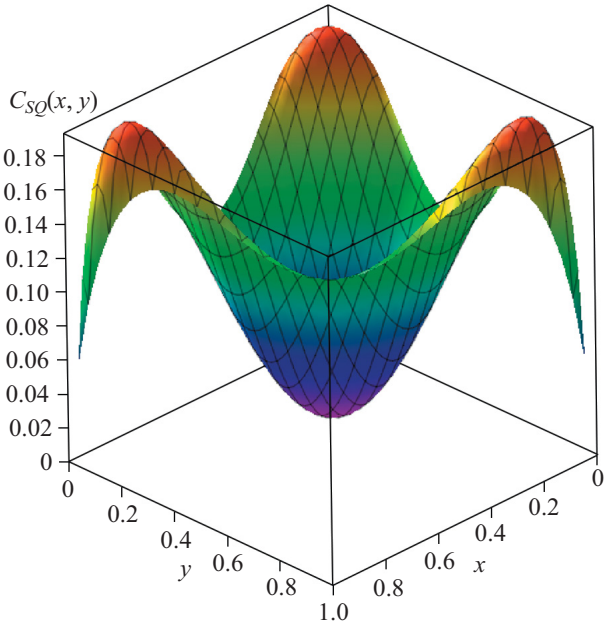


Fig. 2. Level surfaces of statistical complexity $C_{SQ}(x, y)$ for $p = \{p_1 = x,\ p_2 = y,\ p_3 = 1 - x - y\}$.

Corollary 2. *The minimum value of* $C_{SQ} = 0$ *is achieved on a uniform distribution* $p = (1/N, \ldots, 1/N)$.

The validity of the Lemma 3 for the case when the discrete distribution $p = \{p_1, p_2, p_3\}$ consists of three samples is demonstrated in Fig. 2. The complexity depends on two variables, since one of the probabilities can be expressed through the others. Here C_{SQ} has three identical pronounced maxima and three identical local minima pertaining to the cases $p_1 = p_2$, $p_2 = p_3$, $p_1 = p_3$ when the necessary extremum conditions are met, and a global minimum when $p_1 = p_2 = p_3$.

Table 1 shows the change of optimal parameters $C_{SQ}(w, p_{\max})$ with increasing N.

Table 1. Optimal parameters $C_{SQ}(\omega, p_{\max})$ for different values of N

N	$C_{SQ}(\omega^*, p^*_{\max})$	$p^*_{\max}$	ω^*	$N - K^*$
3	0.1932	0.8315	0.6666	1
256	0.1994	0.7044	0.9960	1
512	0.1942	0.7008	0.9980	1
1024	0.1898	0.6979	0.9990	1
2048	0.1861	0.6955	0.9995	1

The necessary extremum conditions C_{SQ} for the discrete distribution $p = \{p_1 = x,\ p_2 = y,\ p_3 = 1 - x - y\}$ are written out according to (25) as follows:

$$\begin{cases} (-\log x + \log(1-x-y))\left(\left(x-\frac{1}{3}\right)^2 + \left(y-\frac{1}{3}\right)^2 + \left(1-x-y-\frac{1}{3}\right)^2\right) \\ \quad - 2(x\log x + y\log y + (1-x-y)\log(1-x-y))\,(-1+y)) = 0, \\ (-\log y + \log(1-x-y))\left(\left(x-\frac{1}{3}\right)^2 + \left(y-\frac{1}{3}\right)^2 + \left(1-x-y-\frac{1}{3}\right)^2\right) \\ \quad - 2(x\log x + y\log y + (1-x-y)\log(1-x-y))\,(-1+x) = 0. \end{cases} \tag{33}$$

The implicit equations of the system (33) describe the curves shown in Fig. 3.

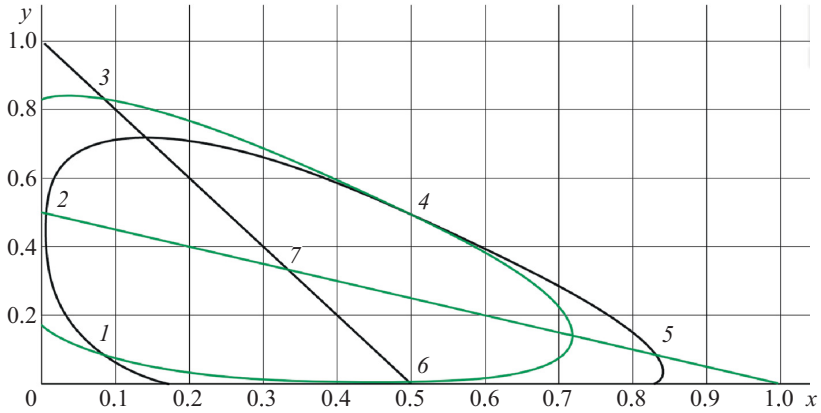


Fig. 3. Curves corresponding to the necessary conditions of extremum C_{SQ} for $p = \{p_1 = x, p_2 = y, p_3 = 1 - x - y\}$.

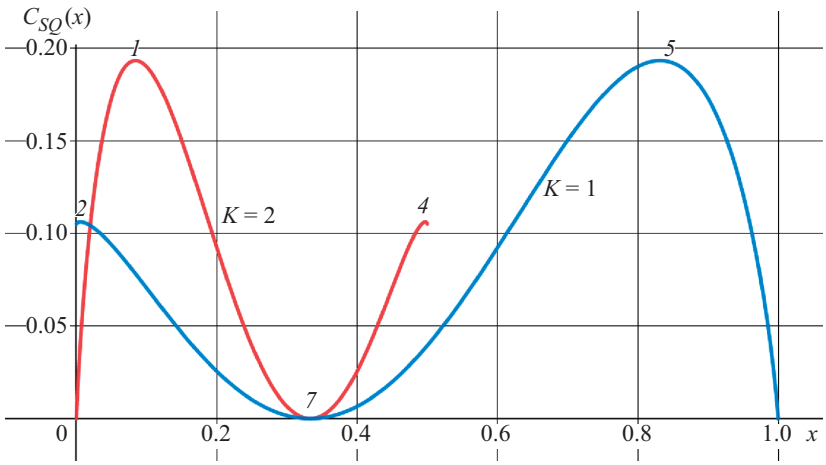


Fig. 4. Statistical complexity C_{SQ} for $p = \{p_1 = x,\ p_2 = x,\ p_3 = 1 - 2x\}$.

The black and green curves correspond to the first and second implicit equations of the (33) system, respectively. In Fig. 3 seven extremum points are marked, for which the value of statistical complexity is calculated. All the obtained data is summarized in Table 2.

The first, third and fifth points of maximum correspond to the same value of the maximum of the function. The second, fourth and sixth minimum points correspond to the same value of the minimum of the function. It is worth noting that the extremum points correspond to the case $K = N - 1 = 2$ except for the global minimum, where all probabilities are equal to each other, and thus describe three local minima, one global minimum, and three equal maxima of statistical complexity in Fig. 2. We can separately plot the statistical complexity in the case $p = \{p_1 = x, p_2 = x, p_3 = 1 - 2x\}$.

In Fig. 4 the extremum points are marked according to Table 2, which cover all cases $p = \{p_1 = x,\ p_2 = y,\ p_3 = 1 - x - y\}$.

Table 2. Extremum points of statistical complexity at $N = 3$

p	1	2	3	4	5	6	7
p_1	0.08425	0.006	0.08425	0.497	0.8315	0.497	0, (3)
p_2	0.08425	0.497	0.8315	0.497	0.08425	0.006	0, (3)
p_3	0.8315	0.497	0.08425	0.006	0.08425	0.497	0, (3)
C_{SQ}	0.1932	0.1062	0.1932	0.1062	0.1932	0.1062	0

3.2. Optimization of C_{JSD}

Let us apply a similar approach using the Jensen–Shannon divergence as the disequilibrium to the statistical complexity C_{JSD}

$$C_{JSD}(p) = H(p) \times JSD(p||q), \qquad q_j = 1/N, \quad j = 1, \ldots, N, \tag{34}$$

which can be written by considering the expression $JSD(p||q)$ through entropy

$$C_{JSD}(p) = H(p) \times \left(H(m) - \frac{1}{2}(H(p) + H(q)) \right) \times \log N, \quad m = \frac{p+q}{2}. \tag{35}$$

It is possible to write out the necessary conditions of extremum for the statistical complexity of the form (35), but at the same time a lemma similar to the Lemma 3 cannot be proved.

Equation (35) is written in variables p_i and p_N in accordance with the approach of the Lemma 3

$$\begin{gathered} C_{JSD}(p) = -\left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) \\ \times \left(H(m) - \frac{1}{2} \left(-\frac{1}{\log N} \left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N \right) + 1 \right) \right), \end{gathered} \tag{36}$$

where

$$H(m) = -\frac{1}{\log N} \left(\sum_{i=1}^{N-1} \frac{p_i + \frac{1}{N}}{2} \log \frac{p_i + \frac{1}{N}}{2} + \frac{p_N + \frac{1}{N}}{2} \log \frac{p_N + \frac{1}{N}}{2} \right). \tag{37}$$

Then taking into account (21)

$$\frac{\partial H(m)}{\partial p_i} = -\frac{1}{\log N} \left(\frac{1}{2} \log \frac{p_i + \frac{1}{N}}{2} - \frac{1}{2} \log \frac{p_N + \frac{1}{N}}{2} \right), \quad i = 1, \ldots, N-1. \tag{38}$$

The necessary conditions of extremum are obtained in the following form after combining all partial derivatives:

$$\begin{gathered} \frac{\partial C_{JSD}(p)}{\partial p_i} = \frac{H(p)}{2} \times \left(-\left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_N + \frac{1}{N}}{2} \right) + (\log p_i - \log p_N) \right) \\ - (\log p_i - \log p_N) \times JSD(p||q) = 0, \quad i = 1, \ldots, N-1. \end{gathered} \tag{39}$$

The equations after simplification are given:

$$\begin{gathered} H(p) \times \left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_N + \frac{1}{N}}{2} \right) \\ + (\log p_i - \log p_N) \times (2JSD(p||q) - H(p)) = 0, \quad i = 1, \ldots, N-1. \end{gathered} \tag{40}$$

Then the difference of Eqs. (40) for indices i, j takes the following form

$$(\log p_i - \log p_j) \times (2JSD(p||q) - H(p)) + H(p) \times \left(\log \frac{p_i + \frac{1}{N}}{2} - \log \frac{p_j + \frac{1}{N}}{2} \right) = 0. \tag{41}$$

Remark 2. It follows from the form of the system of Eqs. (41) that the system is satisfied if $p_i = p_j$, which is one of the necessary conditions for the extremum of the function (35). Due to the nonlinearity of the system consisting of Eqs. (41), it may have other roots.

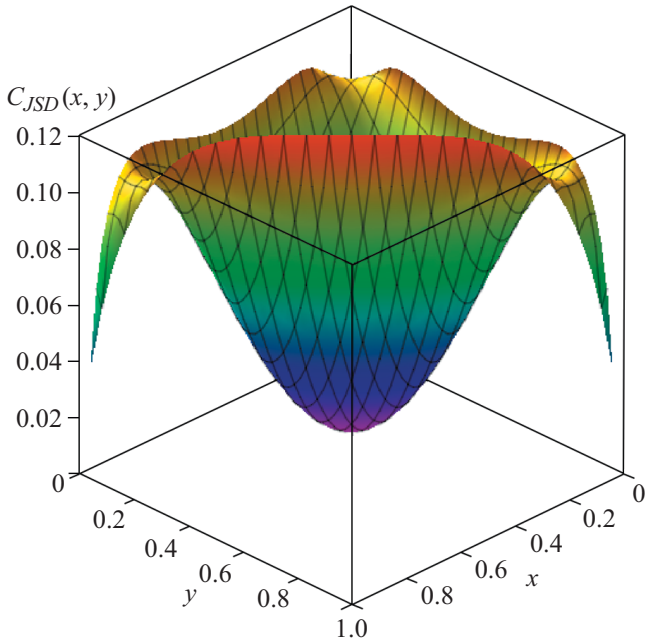


Fig. 5. Level surfaces of statistical complexity C_{JSD} for $p = \{p_1 = x, p_2 = y,\ p_3 = 1 - x - y\}$.

Figure 5 shows a surface plot of the statistical complexity level of the form (35) when the discrete distribution $p = \{p_1, p_2, p_3\}$ consists of three samples to illustrate the Remark 2.

It can be seen from Fig. 5 that the points satisfying $p_1 = p_2$, $p_2 = p_3$ and $p_2 = p_3$ are the saddle points of the surface if the necessary extremum conditions are satisfied.

It was previously established that the distribution (30) delivers an extremum to C_{SQ} at $K = N - 1$. Next, it will be shown that it also delivers the extremum to the complexity based on the total variation of the measure $TV(p, q)$. Therefore, we propose to find the maximum of C_{JSD} on this distribution and compare the obtained optimal distribution parameters at fixed N. Let us write out the complexity explicitly and obtain

$$C_{JSD}^{(K)} = H^{(K)} \times \left(H^{(K)}(m) - \frac{1}{2}(H^{(K)} + 1) \right) \times \log N, \tag{42}$$

where $H^{(K)}$ is corresponding to (32), and $H^{(K)}(m)$ is given by the following formula:

$$H^{(K)}(m) = 1 - \frac{1}{\log N} \left(\frac{(1 - p_{\max} + \omega)}{2} \log \frac{(1 - p_{\max} + \omega)}{2\omega} + \frac{(1 + p_{\max} - \omega)}{2} \log \frac{(1 + p_{\max} - \omega)}{2 - 2\omega} \right). \tag{43}$$

Table 3 shows the change of optimal parameters C_{JSD} with the growth of N.

Table 3. Optimal parameters $C_{JSD}(\omega, p_{\max})$ for different values of N

N	$C_{JSD}(\omega^*, p^*_{\max})$	$p^*_{\max}$	ω^*	$N - K^*$
3	0.1266	1	0.4083	1 or 2
256	0.4482	1	0.8703	33
512	0.4790	1	0.8897	56
1024	0.5065	1	0.9051	97
2048	0.5312	1	0.9171	170

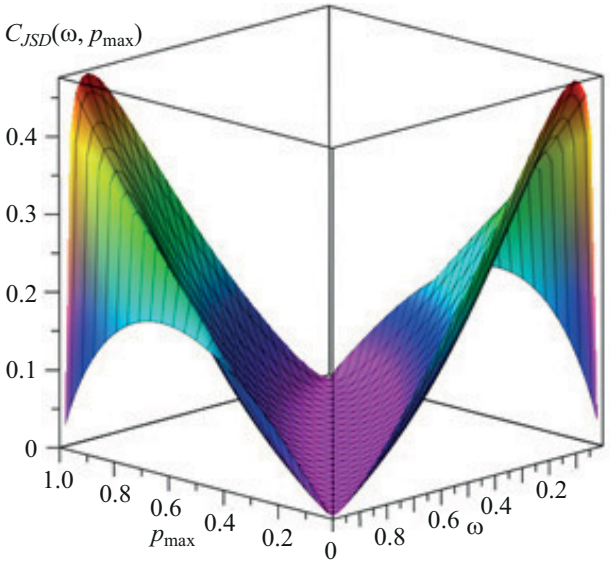


Fig. 6. Level surfaces of statistical complexity $C_{JSD}(\omega, p_{\max})$.

For clarity, Fig. 6 shows the graph $C_{JSD} = C_{JSD}(\omega, p_{\max})$ at $N = 1024$, where ω is changing continuously (although K is changing discretely).

The results shown in Table 3 demonstrate that for the chosen class of distributions (30) the set of N components, where K is equal to each other and the rest are zero, is optimal. It is worth noting that for the resulting distribution C_{JSD} is not zero, as well due to the summand $H^{(K)}(m)$, which corresponds to the already "shifted" distribution consisting of K elements equal to $\frac{\frac{1}{K}+\frac{1}{N}}{2}$, and $N - K$ samples of $\frac{1}{2N}$ each.

3.3. Optimization C_{TV}

Let us proceed to analyze statistical complexity based on total variation

$$C_{TV}(p) = -\frac{1}{4\log N}\left(\sum_{i=1}^{N} p_i \log p_i\right) \times \left(\sum_{i=1}^{N}\left|p_i - \frac{1}{N}\right|\right)^2. \tag{44}$$

Proposition 1. *According to the expression for the error function* (2) *from the Neyman–Pearson Lemma 1 and the definition of* (44)*, we propose to use C_{TV} as a criterion to solve the problem of hypothesis testing and indicating the appearance of a deterministic component of a useful signal in noise.*

The following lemma is valid.

Lemma 4. *The maximum statistical complexity* (44) *is achieved on the family of distributions* (30)*.*

Proof. Given the symmetry of the function (44) and the simplex (3), without restriction of generality, we find an integer $K \in \{1, \ldots, N-1\}$ for which the maximum of this lemma is achieved on the part of the simplex (3) defined by the constraints $p_i \leqslant 1/N$ for $i = 1, \ldots, K$ and $p_i \geqslant 1/N$ for $i = K+1, \ldots, N$. Let us rewrite the Eq. (44) in the form of

$$C_{TV} = -\frac{1}{4\log N}\left(\sum_{i=1}^{N-1} p_i \log p_i + p_N \log p_N\right) \times \left(\sum_{i=1}^{K}\left(-p_i + \frac{1}{N}\right) + \sum_{i=K+1}^{N}\left(p_i - \frac{1}{N}\right)\right)^2. \tag{45}$$

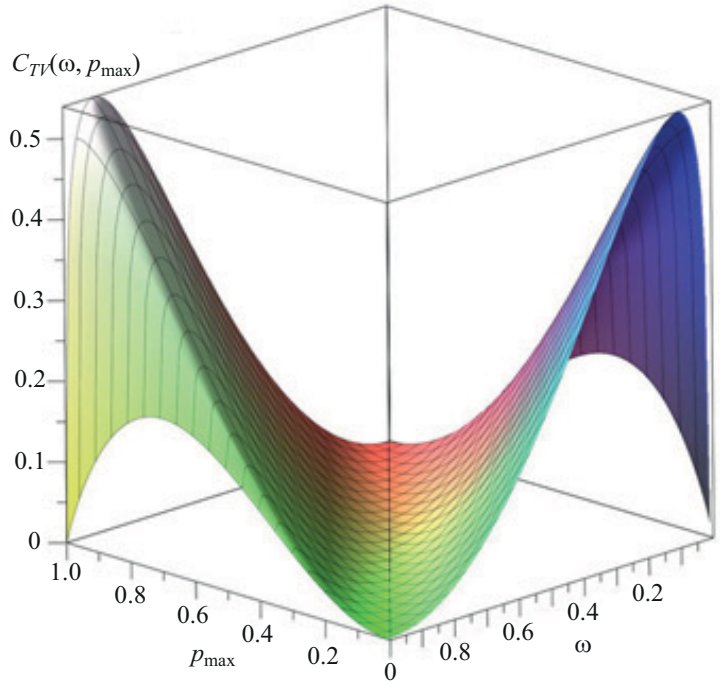


Fig. 7. Level surfaces of statistical complexity $C_{TV}(\omega, p_{\max})$.

Then for $i = 1, \ldots, K$ the necessary conditions of extremum take the form

$$\frac{\partial C_{TV}}{\partial p_i} = -\frac{1}{\log N}\left(\log p_i - \log p_N\right) \times D_{TV} - 2H(p)\sqrt{D_{TV}} = 0, \quad i = 1, \ldots, K, \tag{46}$$

and for $i = K+1, \ldots, N$ the following is true

$$\frac{\partial C_{TV}}{\partial p_i} = -\frac{1}{\log N}\left(\log p_i - \log p_N\right) \times D_{TV} = 0, \quad i = K+1, \ldots, N. \tag{47}$$

Let us compose the difference of two equations from (46) for indices i and j. Whence it follows that if $D_{TV} \neq 0$, then $p_i = p_j$ at $i = 1, \ldots, K$. Whereas it follows from (47) that $p_i = p_N$ at $i = K+1, \ldots, N$. Again we obtain that the family of distributions (30) delivers the maximum of the complexity function, now C_{TV}. ■

Next, we need to determine the optimal values of K and $p_{\max}$. For this purpose, let us calculate the disequilibrium value $D_{TV}^{(K)}$ on the distribution (30):

$$D_{TV}^{(K)} = (p_{\max} + \omega - 1)^2, \quad \omega = \frac{K}{N}. \tag{48}$$

In turn, entropy is equal to

$$H^{(K)} = 1 - \frac{1}{\log N}\left((1 - p_{\max})\log\frac{1 - p_{\max}}{\omega} + p_{\max}\log\frac{p_{\max}}{1 - \omega}\right). \tag{49}$$

For clarity, Fig. 7 shows the graph $C_{TV} = C_{TV}(\omega,\ p_{\max})$ at $N = 1024$, where ω is changing continuously (although K is changing discretely).

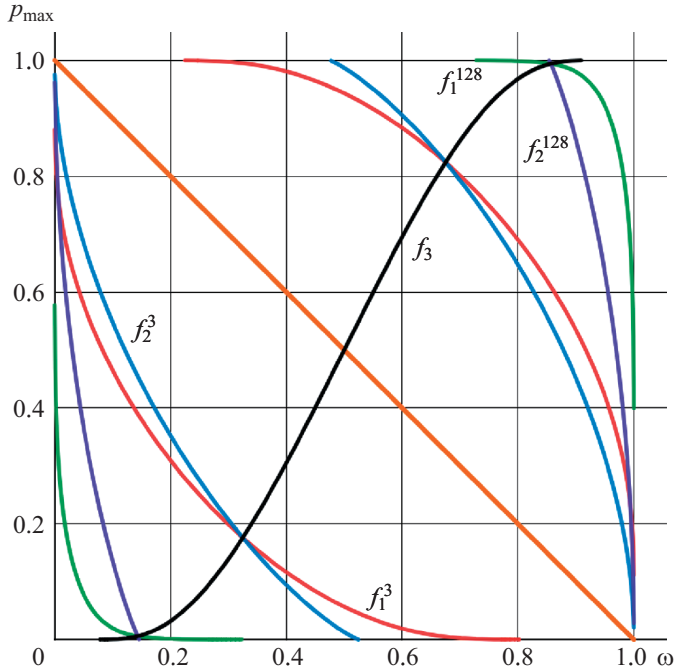


Fig. 8. Curves of necessary conditions of extremum C_{TV} for $N = 3$ and $N = 128$. Equations $f_1^N(p_{\max}, \omega)$ and $f_2^N(p_{\max}, \omega)$ have common solution $(p_{\max} + \omega - 1 = 0)$ (orange curve).

Let us compose the necessary conditions (46) and (47) of the extremum of the statistical complexity C_{TV} written out through the variables $p_{\max},\ \omega$.

$$\begin{cases} f_1^N(p_{\max}, \omega) := 2(p_{\max} + \omega - 1)\left(\left(1 - \frac{(1-p_{\max})\log\frac{1-p_{\max}}{\omega} + p_{\max}\log\frac{p_{\max}}{1-\omega}}{\log N}\right)\right. \\ \left. - \frac{(p_{\max} + \omega - 1)}{2\log N}\left(\log\frac{p_{\max}}{1-\omega} - \log\frac{1-p_{\max}}{\omega}\right)\right) = 0, \\ f_2^N(p_{\max}, \omega) := 2(p_{\max} + \omega - 1)\left(\left(1 - \frac{(1-p_{\max})\log\frac{1-p_{\max}}{\omega} + p_{\max}\log\frac{p_{\max}}{1-\omega}}{\log N}\right)\right. \\ \left. - \frac{(p_{\max} + \omega - 1)}{2\log N}\left(\frac{p_{\max}}{1-\omega} - \frac{1-p_{\max}}{\omega}\right)\right) = 0. \end{cases} \tag{50}$$

The intersections of the curves corresponding to the implicit Eqs. (50) are related to to the extremum points of C_{TV}. Let us compose the difference of two necessary conditions of the extremum

$$\begin{aligned} f_3^N(p_{\max}, \omega) &:= f_1^N(p_{\max}, \omega) - f_2^N(p_{\max}, \omega) \\ &= \frac{(p_{\max} + \omega - 1)^2}{\log N}\left(-\log\frac{p_{\max}}{1-\omega} + \log\frac{1-p_{\max}}{\omega} + \frac{p_{\max}}{1-\omega} - \frac{1-p_{\max}}{\omega}\right) = 0. \end{aligned} \tag{51}$$

Let us construct the implicit curves of equations $f_1^N(p_{\max}, \omega),\ f_2^N(p_{\max}, \omega),\ f_3^N(p_{\max}, \omega)$ for some values of N. For convenience, the index N of $f_3^N(p_{\max}, \omega)$ can be omitted since the implicit curve of the Eq. (51) is independent of N.

According to Fig. 8 the statistical complexity has, in addition to the minimum points $(p_{\max} + \omega - 1 = 0 = 0)$, where $C_{TV} = 0$, also two maximum points for each value of N: $(p^*_{\max}, \omega^*)$ and $(1 - p^*_{\max}, 1 - \omega^*)$, which lie on the curve $f_3(p_{\max}, \omega)$.

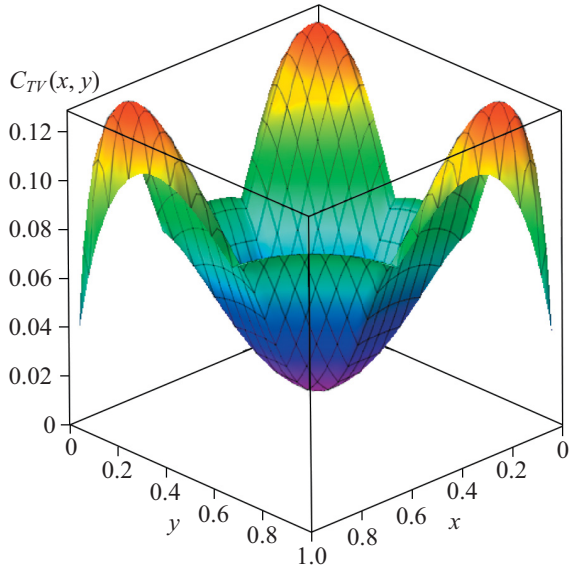


Fig. 9. Level surfaces of statistical complexity $C_{TV}(x, y)$ for $p = \{p_1 = x,\ p_2 = y,\ p_3 = 1 - x - y\}$.

In Table 4 the optimal values of the parameters of the formulas (48) and (49) that maximize the statistical complexity of C_{TV} are given.

Table 4. Optimal parameters $C_{TV}(\omega, p_{\max})$ for different values of N

N	$C_{TV}(\omega^*, p^*_{\max})$	$p^*_{\max}$	ω^*	$N - K^*$
3	0.1289	0.8241	0.6751	1 or 2
256	0.4789	0.9976	0.8752	32
512	0.5120	0.9991	0.8901	56
1024	0.5410	0.9997	0.9022	100
2048	0.5667	0.9999	0.9122	180

Additionally, the case $N = 3$ is shown in Fig. 9, which shows a graph of the level surface C_{TV} when the discrete distribution $p = \{p_1, p_2, p_3\}$ consists of three samples.

4. STATISTICAL COMPLEXITY MODELING AND COMPARISON

We analyze the optimal parameters that maximize different types of statistical complexity and compare the values in Tables 1, 3, and 4. Of main interest are the maximum complexity values and the optimal values of K. The maximum values of $C_{TV}(\omega^*, p^*_{\max}) \in [0, 1]$, $C_{JSD}(\omega^*, p^*_{\max}) \in [0, 1]$ are close to each other and grow with increasing N. The optimal values of K for these two types of complexity are also close.

To demonstrate the analytical results obtained in the previous sections, the application of three variants of statistical complexity in the problem of useful signal indication in a noise mixture for synthesized signals is shown. An algorithm from [15] based on the computation of discrete distributions p from the spectral representation of time series is applied.

The synthesized 10-second signal is the sum of a finite number of cosine oscillations mixed with white noise:

$$x(t) = I(t) \sum_{i=1}^{K} A_i \cos(2\pi f_i t + \Delta\phi_i) + w(t), \quad t \in [0, 10], \tag{52}$$

where $A_i,\ f_i,\ \Delta\phi_i$ are the amplitudes, frequencies, and random phases of the harmonic oscillations, respectively, $w(t)$ is the white noise, and $I(t)$ is the indicator function for the presence of the useful signal in the signal-noise mixture.

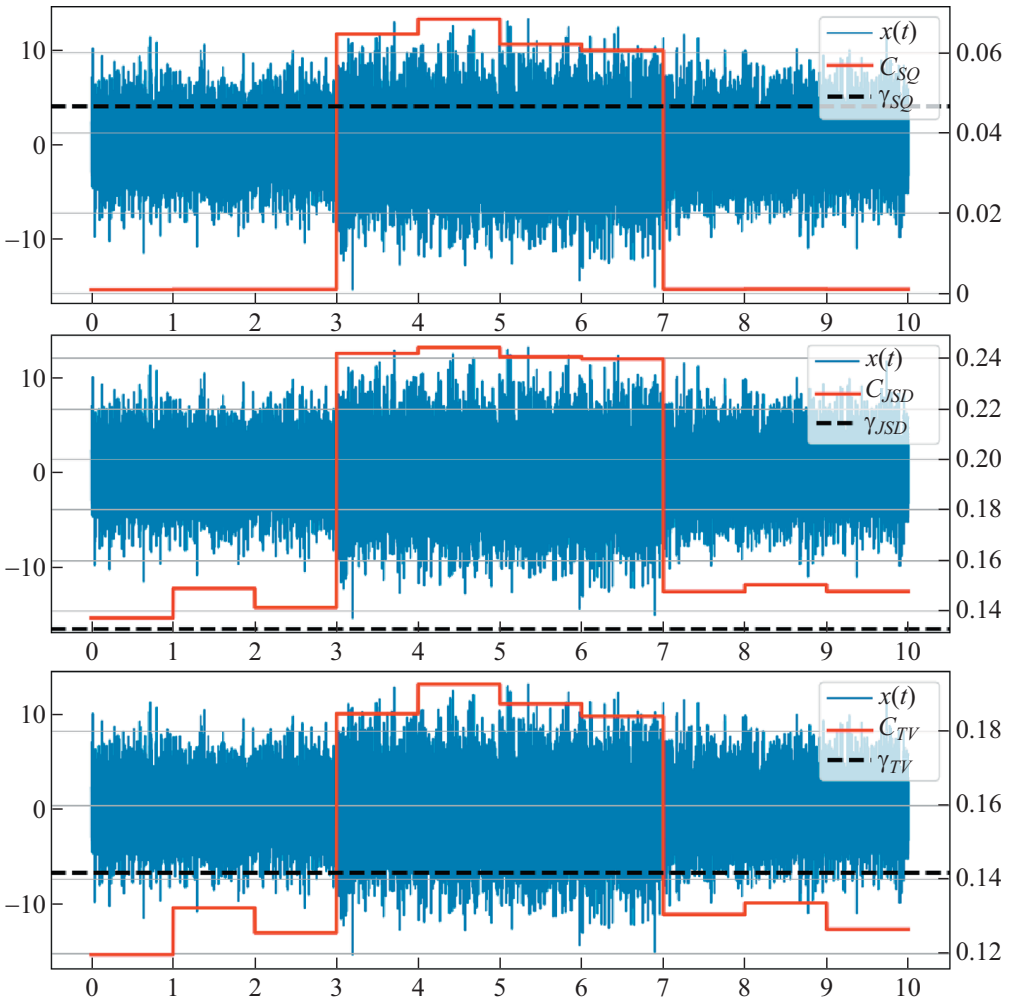


Fig. 10. Three components, $K = 3$.

$I(t)$ is chosen so that the harmonic signals are present in the middle of the final sequence $x(t)$.

$$I(t) = \begin{cases} 0, & t \in [0, 3), \\ 1, & t \in [3, 7], \\ 0, & t \in (7, 10]. \end{cases} \tag{53}$$

The algorithm has the following structure:

1. The signal synthesized with sampling frequency f_s is divided into short windows containing $N = 2048$ samples each.

2. Next, the spectrum for each window is calculated using the FFT algorithm.

3. Based on the spectrum, the discrete densities $p_i, i = 1, \ldots, N$ are calculated by normalizing it.

4. The information characteristics $C_{SQ}(p)$, $C_{JSD}(p)$, $C_{TV}(p)$ are calculated for the obtained set p_i.

5. The obtained sequence of information characteristic values is shown along with the signal on the time axis.

It should be noted that the parameters f_s and N were chosen to exclude the effect of spectrum spreading, i.e., to obtain clear spectral components corresponding to K harmonic functions from the formula (52). The signal-to-noise ration is chosen to be close to one.

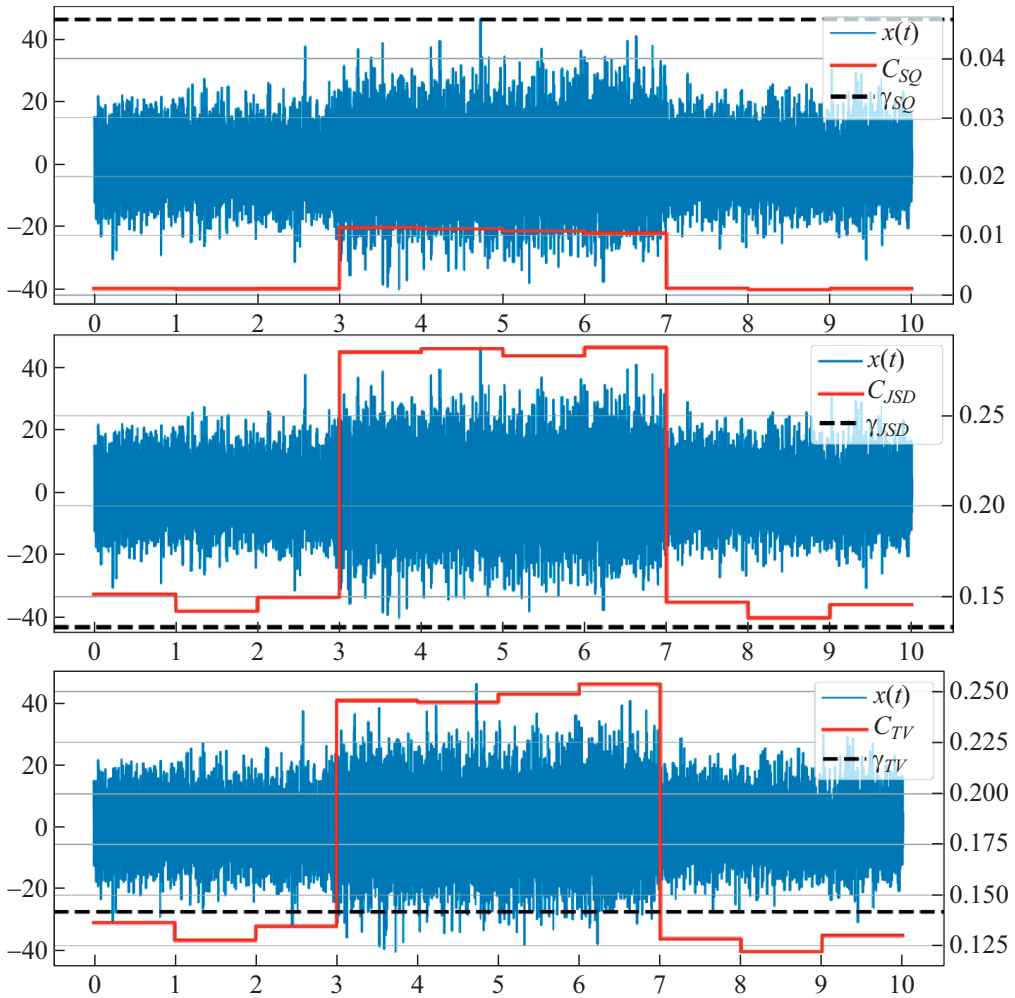


Fig. 11. Thirty components, $K = 30$.

The threshold γ for the decision rule is proposed to be chosen as 25% of the maximum criterion value for selected N from the Tables 1, 3, 4:

$$\begin{aligned} \gamma_{CQ} &= 0.25 \times 0.1861 = 0.0465; \\ \gamma_{JSD} &= 0.25 \times 0.5312 = 0.1328; \\ \gamma_{TV} &= 0.25 \times 0.5667 = 0.1417. \end{aligned} \tag{54}$$

The convenience of choosing such a threshold is that it does not depend on a particular noise realization and is based on analytically derived maximum values of statistical complexity functions.

In all plots, the blue color indicates the amplitude of the original signal, and the red color indicates the statistical complexity, which is calculated using the algorithm described above. The horizontal axis represents time in seconds, and the vertical axes represent the magnitude of the signal amplitude (left) and the criterion (right). The black dashed line shows the value of the selected threshold γ.

In the first experiment, the number of sinusoidal signals and respectively spectral components is equal to $K = 3$ at $N = 2048$. Figure 10 shows the dependencies of statistical complexities on time for the synthesized signal.

As can be seen, the values of C_{SQ} and C_{TV} exceed the selected threshold for the interval of signal presence, which allows us to confidently conclude that the signal has occurred. As for C_{JSD},

the a priori threshold selection was unsuccessful because the true value of its maximum is unknown, as shown in Section 3.2. If we change the threshold value upwards by 20%, the detection based on C_{JSD} will be as successful as that based on C_{TV}.

In the second experiment, the number of spectral components $K = 30$. In this case, C_{SQ} ceases to show a satisfactory result in the sense of exceeding the chosen threshold, since the function C_{SQ} degrades strongly with increasing of K, but still allows a signal indication, as can be seen in Fig. 11. The complexity function C_{TV} still confidently exceeds the threshold, as in the first experiment, and C_{JSD} exceeds the threshold on the whole signal, as in the first experiment.

Thus, we can conclude that C_{TV} is the most convenient in a practical sense, since it works well on signals with a large number of spectral components and allows one to decide on the appearance of an useful signal, using a fairly simple rule, associated with the choice of threshold based on the theoretical maximum value for the statistical complexity.

5. CONCLUSION

The article provides a theoretical justification for the usage of statistical complexity as a criterion for solving the problem of hypothesis testing when its error probability is close to one. Three variants of statistical complexity for different disequilibrium functions are considered. New notions of disequilibrium and statistical complexity based on the total variation measure are introduced. Information criteria are compared and the classes of discrete distributions which provide maximum for different types of statistical complexity are discovered. The values of maxima for fixed numbers of distribution samples are found. It is shown that the statistical complexity C_{TV} based on total variation is directly related to the problem of hypothesis testing, while the statistical complexity C_{JSD} based on Jensen–Shannon entropy gives a close estimate of C_{TV} on sample distributions. In turn, C_{SQ} is most promising for detecting an individual component over a uniform distribution. We propose a method for selecting the threshold for the decisive rule for the detection of a useful signal, taking into account the maximum values of the criteria obtained and show the effectiveness of this approach on the synthesized signals.

Future work will be devoted to the study of information criteria based on bivariate and multivariate distributions, as well as investigation of typical acoustic signals with realistic background noise.

FUNDING

This work was supported by the Russian Scientific Foundation, project no. 23-19-00134.

REFERENCES

1. Shannon, C.E., A Mathematical Theory of Communication, *Bell Syst. Tech. J.*, 1948, vol. 27, pp. 379–423.

2. Gray, R.M., *Entropy and Information Theory*, New York: Springer, 2011. https://doi.org/10.1007/978-1-4419-7970-4

3. Holub, A., Perona, P., and Burl, M.C., Entropy-based Active Learning for Object Recognition, *Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE*, 2008, pp. 1–8. https://doi.org/10.1109/CVPRW.2008.4563068

4. Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., et al. Supervised Machine Learning Algorithms: Classification and Comparison, *Int. J. Comput. Trends Technol. (IJCTT)*, 2017, vol. 48, no. 3, pp. 128–138. https://doi.org/10.14445/22312803/IJCTT-V48P126

5. Shen, J., Hung, J., and Lee, L., Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments, *Proc. 5th International Conference on Spoken Language Processing (ICSLP)*, 1998. https://doi.org/10.21437/icslp.1998-527

6. Ribeiro, M., Henriques, T., Castro, L., Souto, A., Antunes, L., Costa-Santos, C., and Teixeira, A., The Entropy Universe, *Entropy*, 2021, vol. 23, no. 2, art. 222. https://doi.org/10.3390/e23020222

7. Ramirez, J., Segura, J.C., Benitez, C., et al., A New Kullback-Leibler VAD for Speech Recognition in Noise, *IEEE Signal Proc. Lett.*, 2004, vol. 11, no. 2, pp. 266–269. https://doi.org/10.1109/LSP.2003.821762

8. Horie, T., Burioka, N., Amisaki, T., and Shimizu, E., Sample Entropy in Electrocardiogram During Atrial Fibrillation, *Yonago Acta Medica*, 2018, vol. 61, no. 1, pp. 49–57. https://doi.org/10.33160/yam.2018.03.007

9. Lamberti, P.W., Martin, M.T., Plastino, A., and Rosso, O.A., Intensive Entropic Non-Triviality Measure, *Phys. A: Stat. Mech. Appl.*, 2004, vol. 334, no. 1, pp. 119–131. https://doi.org/10.1016/j.physa.2003.11.005

10. Lopez-Ruiz, R., Shannon Information, LMC Complexity and Renyi Entropies: A Straightforward Approach, *Biophys. Chem.*, 2005, vol. 115, no. 3, pp. 215–218. https://doi.org/10.1016/j.bpc.2004.12.035

11. Zunino, L., Soriano, M.C., and Rosso, O.A., Distinguishing Chaotic and Stochastic Dynamics from Time Series by Using a Multiscale Symbolic Approach, *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys.*, 2012, vol. 86, no. 4, pp. 1–5. https://doi.org/10.1103/PhysRevE.86.046210

12. Ronald, L.A. and Duncan, W.M., *Signal Analysis: Time, Frequency, Scale, and Structure*, New Jersey: IEEE Press, 2004.

13. Shiryaev, A.N., *Veroyatnostno-statisticheskie metody v teorii prinyatiya reshenii* (Probabilistic-Statistical Methods in Decision-Making Theory), Moscow: MTSNMO: NMU, 2020.

14. Kishan, G.M., Chilukuri, K.M., and HuaMing Huang, *Anomaly Detection Principles and Algorithms*, Cham: Springer, 2017. https://doi.org/10.1007/978-3-319-67526-8

15. Berlin, L.M., Galyaev, A.A., and Lysenko, P.V., Comparison of Information Criteria for Detection of Useful Signals in Noisy Environments, *Sensors*, 2023, vol. 23, no. 4, art. 2133. https://doi.org/10.3390/s23042133

16. Johnson, P., Moriarty, J., and Peskir, G., Detecting Changes in Real-Time Data: A User's Guide to Optimal Detection, *Philos. Trans. Royal Soc. A*, 2017, vol. 375, p. 16, art. 2100. https://doi.org/10.1098/rsta.2016.0298

17. Li, Z., Li, Y., and Zhang, K.A., Feature Extraction Method of Ship-Radiated Noise Based on Fluctuation-Based Dispersion Entropy and Intrinsic Time-Scale Decomposition, *Entropy*, 2019, vol. 21, no. 7, art. 693. https://doi.org/10.3390/e21070693

18. Sason, I., On f-Divergences: Integral Representations, Local Behavior, and Inequalities, *Entropy*, 2018, vol. 20, no. 5, art. 383. https://doi.org/10.3390/e20050383

This paper was recommended for publication by A.V. Nazin, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 872–888.

On Asymptotically Optimal Approach for Finding of the Minimum Total Weight of Edge-Disjoint Spanning Trees with a Given Diameter

E. Kh. Gimadi[*,a] **and A. A. Shtepa**[**,b]

[*] *Sobolev Institute of Mathematics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*
[**] *Novosibirsk State University, Novosibirsk, Russia*
e-mail: [a] *gimadi@math.nsc.ru,* [b] *shoomath@gmail.com*

Received January 23, 2023
Revised March 27, 2023
Accepted April 28, 2023

Abstract—We consider the intractable problem of finding several edge-disjoint spanning trees of the minimum total weight with a given diameter in complete undirected graph in current paper. The weights of edges of a graph are random variables from several continuous distributions: uniform, biased truncated exponential, biased truncated normal. The approximation algorithm with time complexity $\mathcal{O}(n^2)$, where n is number of vertices in graph, is proposed for solving this problem. The asymptotic optimality conditions for constructed algorithm is presented for each considered probabilistic distribution.

Keywords: minimum spanning tree with given diameter, approximation algorithm, probabilistic analysis, asymptotic optimality

DOI: 10.25728/arcRAS.2023.42.85.001

1. INTRODUCTION

The Minimum Spanning Tree (MST) Problem is one of well-known problems of discrete optimization. Is consists of finding spanning tree (connected acyclic subgraph on all vertices) of minimum weight in given edge-weighted graph $G = (V, E)$. Polynomial solvability of this problem was proved by construction of polynomial algorithms Boruvka (1926), Kruskal (1956), and Prim (1957). These algorithms have time complexities $\mathcal{O}(u \log n)$, $\mathcal{O}(u \log u)$, and $\mathcal{O}(n^2)$ respectively, where $u = |E|$ and $n = |V|$. It is interesting to note, that expected value of MST's weight in graph with random edge weights can be surprisingly small. For example, MST's weight with high probability is close to constant 2.02 for complete graph with edge weights from uniform distribution on interval $(0; 1)$ [1]. Similar results were obtained in [2, 3].

One possible generalization of the above problem is the bounded diameter version of the MST problem. The diameter of a tree is the number of edges in the longest simple path within the tree connecting a pair of vertices. This problem is as follows: given edge-weighted graph and parameter $d = d_n$, it is necessary to find MST in this graph with diameter bounded from above or below by the parameter d. Both problems are NP-hard in general formulation.

The bounded from above MST problem is polynomially solvable for diameters two or three, and NP-hard for any diameter between 4 and $(n-1)$, even for the edge weights equal to 1 or 2 [4, pp. 206]. The MST problem bounded from below is NP-hard, because its particular case for $d = n-1$ is the problem "Hamiltonian Path" [4].

Recently, the authors of this article have began to study another modification of the MST problem with a bounded diameter, when the diameter of this tree is equal to a given number. It is noteworthy that the algorithm for solving such a problem can be transformed into an algorithm for solving a problem with a diameter bounded from above or from below. Thus, the scope of such a problem covers the scope of problems with a bounded diameter both from above and from below.

There are several applications for MST problem with bounded diameter from above in wireless ad-hoc networks [5], network design [6], in development of data compression algorithm [7] and distributed mutual exclusion algorithm [8] (for a detailed description see, for example, [9]).

The problem of finding several edge-disjoint spanning trees of minimum total weight with bounded from below diameter in complete graph arises in the theory of reliability of communication networks, when it is necessary to construct m-connected graph of minimum total weight for a set of objects excluding such configuration of graph, for which after failure of few nodes, the total structure of the graph becomes unreliable. Thus it is necessary to bound the diameter of constructed trees forming m-connected graph. It must be noted that in [10, 11] a probabilistic analysis of an approximation algorithm for this problem was carried out and conditions for its asymptotic optimality were obtained.

In [12, 13] the probabilistic analysis of polynomial algorithm is carried out and asymptotically optimal conditions for this algorithm were proposed for the problem of finding one and several MST with given in the case of complete directed graph. Unfortunately, the algorithm analysis is not accepted for the case of complete undirected graph. The appearance of the difficulty for probabilistic analysis in the case of undirected graph arises from the need to take into account the possible dependence between different objects (random variables) in the course of the algorithm.

We consider the problem of finding m edge-disjoint spanning trees of minimum total weight with a given diameter $d = d_n$ in complete undirected graph (this problem is denoted as m-d-UMST). The approximation algorithm for solving this problem and its conditions of asymptotic optimality are presented. Probabilistic analysis is accomplished for the case of complete edge-weighted undirected graph G without loops under assumption that weights of edges of a graph positive independent identically distributed random variables. The probability distribution functions (p.d.f.) of the weights of graph G are considered from three probabilistic distributions: uniform distribution $\mathrm{UNI}(a_n; b_n)$ on the finite segment $[a_n; b_n]$, as well as biased truncated distributions: exponential $\mathrm{EXP}(a_n, \lambda_n)$ and normal $\mathrm{NORM}(a_n, \sigma_n)$ on the unbounded semiopen interval $[a_n; \infty)$. The probabilistic density functions for these distributions are as follows

$$p(x) = \begin{cases} \dfrac{1}{b_n - a_n} & \text{if } a_n \leqslant x \leqslant b_n \text{ for } \mathrm{UNI}(a_n; b_n); \\ \dfrac{1}{\lambda_n} \exp\left(-\dfrac{x - a_n}{\lambda_n}\right) & \text{if } a_n \leqslant x < \infty \text{ for } \mathrm{EXP}(a_n, \lambda_n); \\ \dfrac{2}{\sigma_n \sqrt{2\pi}} \exp\left(-\dfrac{(x - a_n)^2}{2\sigma_n^2}\right) & \text{if } a_n \leqslant x < \infty \text{ for } \mathrm{NORM}(a_n, \sigma_n); \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

2. THE FINDING OF SEVERAL EDGE-DISJOINT SPANNING TREES OF MINIMUM TOTAL WEIGHT WITH GIVEN DIAMETER IN UNDIRECTED GRAPH

First of all, we formulate considered problem and then propose approximation algorithm for its solution.

Given complete n-vertex edge-weighted undirected graph $G = (V, E)$ and positive integer numbers $m \geqslant 2$, $d \geqslant 4$ such that $m(d+1) \leqslant n$. The m-d-UMST is to find m edge-disjoint spanning trees $T_1, \ldots, T_m$ such that the diameter of each of them is equal to $d = d_n$ and their total weight is minimum. For solving this problem, the next deterministic algorithm is proposed.

Description of algorithm $\mathcal{A}$

Preliminary Step 0. In graph G, choose arbitrary $(n - m(d+1))$-vertex subset V' and arbitrarily split remaining $m(d+1)$ vertex into m subsets $V_1, V_2, \ldots, V_m$ with $(d+1)$ vertices in each set.

Step 1. In each subgraph $G(V_s)$ $s = 1, \ldots, m$, beginning with arbitrary vertex construct $(d+1)$-vertex Hamiltonian path P_s using greedy heuristic "go to the nearest unvisited vertex".

Put $T_s = P_s$, $s = 1, \ldots, m$.

Step 2. Hereafter we assume without loss of generality that d is odd (see remark 1 below). For each pair of paths P_i and P_j, $1 \leqslant i < j \leqslant m$, add vertices from P_j to T_i and from P_i to T_j in such a way that constructed subgraph consists of two edge-disjoint $2(d+1)$-vertex subtree with diameter equals d. Each path P_s, $1 \leqslant s \leqslant m$, is considered as two halves (subpaths) P_s^1 and P_s^2, each of which contains one *end* vertex and $\frac{d-1}{2}$ *inner* vertices of path P_s totally $\frac{d+1}{2}$ vertices in each half.

Construction of edge-disjoint spanning trees T_i and T_j with help of vertices from halves P_i^1, P_i^2 and halves P_j^1, P_j^2 is described in following items 2.1–2.6.

2.1. Connect each inner vertex of P_i^1 by the shortest edge to the inner vertex of P_j^1. So we add this edge to T_j.

2.2. Connect each inner vertex of P_i^2 by the shortest edge to the inner vertex of P_j^2. We add this edge to T_j.

2.3. Connect each inner vertex of P_j^1 by the shortest edge to the inner vertex of P_i^2. Thus, we add this edge to T_i.

2.4. Connect each inner vertex of P_j^2 by the shortest edge to the inner vertex of P_i^1. We add this edge to T_i.

2.5. Connect each end vertex of the path P_i by the shortest edge to the inner vertex of the path P_j. We add this edge to T_j.

2.6. Connect each end vertex of the path P_j by the shortest edge to the inner vertex of the path P_i. So we add this edge to T_i.

Step 3. For $s = 1, \ldots, m$ each vertex of subgraph $G(V')$ is connected by the shortest edge to the inner vertex of the path P_s. Thus, we add this edge to corresponding tree T_s.

The construction of m edge-disjoint spanning trees $T_1, \ldots, T_m$ is completed (see example in Figs. 1–3).

Remark 1. In the case of even d algorithm must be slightly modified. On the Step 1 for the first chosen vertex it is necessary to find the closest vertex v_s, $s = 1, \ldots, m$ in d actions. The first chosen vertex is marked and used after all steps of algorithm. Hereafter all steps of the algorithm must be carried out for $d' = d - 1$, where the first vertex of each path is v_s, and after Step 3 marked vertices are connected with v_s, $s = 1, \ldots, m$. Thus, desired spanning trees are constructed with property that diameter of each tree equals exactly d, and time complexity of presented algorithm remains the same.

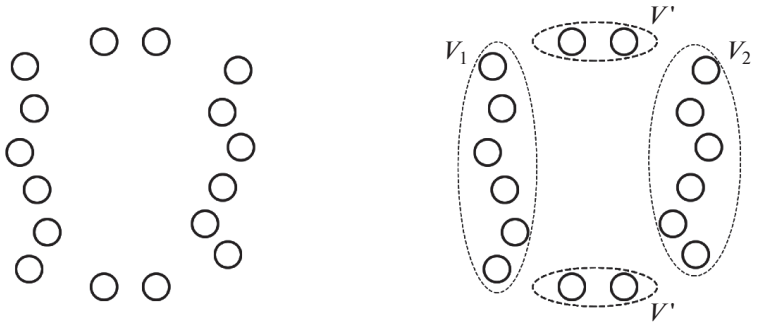


Fig. 1. Initial vertices of the graph and Step 0 of the work of the Algorithm $\mathcal{A}$ in 16-vertex complete graph, $m = 2$, $d = 5$.

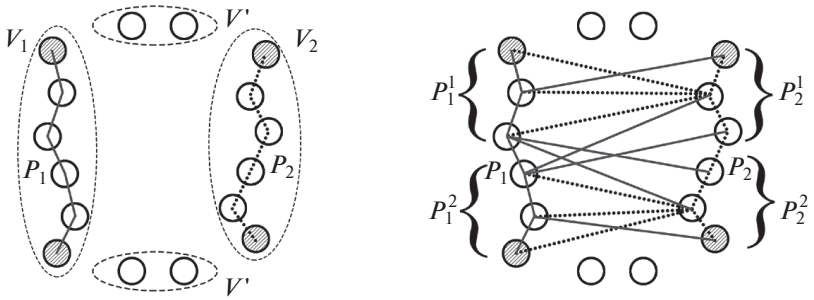


Fig. 2. Steps 1 and 2 of the work of the algorithm $\mathcal{A}$ in 16-vertex complete graph, $m = 2$, $d = 5$. The hatched vertices are end vertices. The solid edges belong to T_1. The dotted edges belong to T_2.

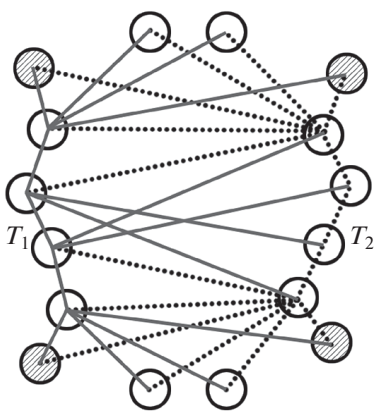


Fig. 3. Step 3 of the work of the Algorithm $\mathcal{A}$ in 16-vertex complete graph, $m = 2$, $d = 5$. The hatched vertices are end vertices. The solid edges belong to T_1. The dotted edges belong to T_2.

Let us introduce the notations: $W_{\mathcal{A}}$ is total weight of all spanning trees $T_1, \ldots, T_m$, which are constructed by algorithm $\mathcal{A}$, W_1, W_2, and W_3 are total weights of edges, which are added to the trees on Steps 1, 2, and 3 respectively. Then $W_{\mathcal{A}} = W_1 + W_2 + W_3$.

Let's formulate two statements concerning algorithm $\mathcal{A}$.

Statement 1. *Algorithm $\mathcal{A}$ constructs feasible solution for the m-d-*UMST*.*

Statement 2. *Time complexity of algorithm $\mathcal{A}$ is estimated by $\mathcal{O}(n^2)$.*

3. PROBABILISTIC ANALYSIS OF ALGORITHM $\mathcal{A}$

Let $F_A(I)$ and $OPT(I)$ be approximation (obtained using some algorithm A) and optimal value of objective function of problem on input I, respectively.

Definition 1. Algorithm A has estimates (performance guarantees) $(\varepsilon_n, \delta_n)$ on a set I of random inputs of a n-sized problem (where n is amount of input data required to describe the problem, see [4]), if

$$\mathbb{P}\{|F_A(I) - OPT(I)| > \varepsilon_n OPT(I)\} \leqslant \delta_n, \tag{2}$$

where $\varepsilon_n = \varepsilon_A(n)$ is an estimate of relative error of a solution obtained by algorithm A, $\delta_n = \delta_A(n)$ is an estimate of the failure probability of the algorithm, which is equal to the proportion of cases when the algorithm does not hold the relative error ε_n or does not produce any answer at all.

Definition 2 [14]. Approximation algorithm A is called asymptotically optimal on a class of input data of a problem, if there exist such performance guarantees that for all input I of size n

$$\varepsilon_n \to 0 \text{ and } \delta_n \to 0 \text{ as } n \to \infty.$$

Hereafter random variable, which is equal to minimum over k independent identically distributed random variables $\eta^1, \ldots, \eta^k$, is denoted as η_k.

According to the description of algorithm $\mathcal{A}$ for Steps 1–3 the following relations are true:

$W_1 = \sum\limits_{s=1}^{m} \sum\limits_{k=1}^{d} \eta_k$, since m paths $P_1, \ldots, P_m$ with d edges in each path are constructed on Step 1.

$W_2 = C_m^2 \big(4\frac{d-1}{2}\eta_{(d-1)/2} + 4\eta_{(d-1)}\big)$, because connection of new edges to constructed set of spanning trees is carried out for each pair of paths from $C_m^2 = \frac{m(m-1)}{2}$ such pairs on corresponding items 2.1–2.6 of Step 2 as follows:

— firstly, each of $\frac{d-1}{2}$ inner vertices of one half of a path connected by shortest edge to each of $\frac{d-1}{2}$ inner vertices of half of another path;

— secondly, each end vertex of each path is connected by shortest edge to one of $d-1$ inner vertices of another path.

The multiplier 4 arises since inner vertices from two halves of one path are connected by shortest edges with inner vertices from two halves of another path for each pair of paths in items 2.1–2.4. In items 2.5–2.6, corresponding shortest edges connects 4 end vertices with inner vertices of considered paths.

$W_3 = m(n - m(d+1))\eta_{(d-1)}$, since each vertex over $n - m(d+1)$ vertices of set V' is connected by shortest edge to inner vertex of the m paths P_s, $1 \leqslant s \leqslant m$ on Step 3 considering $(d-1)$ inner vertices of each path.

Remark 2. It must be noted that for even d:

$$W_1 = m\sum_{k=1}^{d'} \eta_k, \quad W_2 = C_m^2 \left(4\frac{d'-1}{2}\eta_{(d'-1)/2} + 4\eta_{(d'-1)}\right), \quad W_3 = m\left(n - m(d'+1)\right)\eta_{(d'-1)},$$

where $d' = d - 1 \geqslant 3$. Thus, replacing d by d' in all cases it is possible to accomplish probability analysis and prove all the proposed statements.

Hereafter we pass from random variables η, η_k to *normalized* random variables $\xi = \frac{\eta - a_n}{\beta_n}$, $\xi_k = \frac{\eta_k - a_n}{\beta_n}$, respectively, where $\beta_n = \begin{cases} b_n - a_n & \text{for UNI}(a_n; b_n); \\ \lambda_n & \text{for EXP}(a_n, \lambda_n); \\ \sigma_n & \text{for NORM}(a_n, \sigma_n). \end{cases}$

Let us consider random variables W_1, W_2, W_3:

$$W_1 = m\sum_{k=1}^{d}\eta_k = m\sum_{k=1}^{d}(\beta_n\xi_k + a_n) = mda_n + \beta_n m\sum_{k=1}^{d}\xi_k = mda_n + \beta_n W_1',$$

$$W_2 = C_m^2\left(4\frac{d-1}{2}\eta_{(d-1)/2} + 4\eta_{(d-1)}\right) = C_m^2\left(4\frac{d-1}{2}\left(\beta_n\xi_{(d-1)/2} + a_n\right) + 4\left(\beta_n\xi_{(d-1)} + a_n\right)\right)$$
$$= m(m-1)(d+1)a_n + \beta_n m(m-1)\Big((d-1)\xi_{(d-1)/2} + 2\xi_{(d-1)}\Big)$$
$$= \left(m^2(d+1) - md - m\right)a_n + \beta_n W_2',$$

$$W_3 = m(n - m(d+1))\eta_{(d-1)} = m(n-m(d+1))\left(\beta_n\xi_{(d-1)} + a_n\right)$$
$$= m(n-m(d+1))a_n + \beta_n m(n-m(d+1))\xi_{(d-1)} = \left(mn - m^2(d+1)\right)a_n + \beta_n W_3',$$

where W_1', W_2', W_3' are normalized random variables for W_1, W_2, W_3 respectively, and β_n is parameter of corresponding distribution.

So the following relation is obtained for the sum of the weights of the constructed spanning trees: $W_{\mathcal{A}} = m(n-1)a_n + \beta_n W_{\mathcal{A}}'$, where $W_{\mathcal{A}}' = W_1' + W_2' + W_3'$.

Lemma 1. *Algorithm $\mathcal{A}$ for the m-d-*UMST *in n-vertex complete graph with weights of edges from probabilistic distributions* (UNI$(a_n; b_n)$, EXP(a_n, λ_n) *or* NORM(a_n, σ_n)) *is the algorithm with the next estimate of relative error ε_n and failure probability δ_n:*

$$\varepsilon_n = \frac{2\beta_n}{m(n-1)a_n}\widehat{\mathbb{E}W_{\mathcal{A}}'}, \quad \delta_n = \mathbb{P}\{\widetilde{W_{\mathcal{A}}'} > \widehat{\mathbb{E}W_{\mathcal{A}}'}\}, \tag{3}$$

where β_n is parameter of corresponding distribution, $\widehat{\mathbb{E}W_{\mathcal{A}}'}$ is some upper bound for expected value $\mathbb{E}W_{\mathcal{A}}'$, $\widetilde{W_{\mathcal{A}}'} = W_{\mathcal{A}}' - \mathbb{E}W_{\mathcal{A}}'$.

Henceforth, the following statement from theory of probability is useful for probabilistic analysis of algorithm $\mathcal{A}$.

Theorem 1 [15]. *Let us consider random variables $X_1, \ldots, X_n$. We define positive constants T and $h_1, \ldots, h_n$ such that for all $k = 1, \ldots, n$ and $0 \leqslant t \leqslant T$ the inequality is true*

$$\mathbb{E}e^{tX_k} \leqslant e^{\frac{h_k t^2}{2}}. \tag{4}$$

Let $S = \sum\limits_{k=1}^{n} X_k$ and $H = \sum\limits_{k=1}^{n} h_k$. Then

$$\mathbb{P}\{S > x\} \leqslant \begin{cases} \exp\left\{-\dfrac{x^2}{2H}\right\}, & \text{if } \ 0 \leqslant x \leqslant HT, \\ \exp\left\{-\dfrac{Tx}{2}\right\}, & \text{if } \ x \geqslant HT. \end{cases}$$

Also the following statement will be useful for further analysis.

Statement 3. *For all integers $d \geqslant 3$ the following inequality is correct*

$$\sum_{k=1}^{d}\frac{1}{k} \leqslant \ln d + \frac{3}{4}.$$

It is assumed that d is odd and is defined on two semiopen intervals: case 1 ($\ln n \leqslant d < \frac{n}{\ln n}$) and case 2 ($\frac{n}{\ln n} \leqslant d < \frac{n}{m}$).

3.1. Probability Distribution $\mathrm{UNI}(a_n; b_n)$

We pass from random variables η and η_k to normalized random variables $\xi = \frac{\eta - a_n}{b_n - a_n}$ and $\xi_k = \frac{\eta_k - a_n}{b_n - a_n}$ for uniform distribution $\mathrm{UNI}(a_n; b_n)$.

Lemma 2. *For* $\mathbb{E}W'_{\mathcal{A}}$ *the following inequality holds*

$$\mathbb{E}W'_{\mathcal{A}} \leqslant m \ln d + \frac{2mn}{d}.$$

Lemma 3. *Let constants* $T = 1$ *and* $h_k = \frac{1}{(k+1)^2}$ *are defined. Then for* $\mathrm{UNI}(a_n; b_n)$ *and biased random variables* $\widetilde{\xi}_k = \xi_k - \mathbb{E}\xi_k$ *the next inequalities are true* $\mathbb{E}e^{t\widetilde{\xi}_k} \leqslant e^{\frac{h_k t^2}{2}}$ *in Petrov's theorem* [15, pp. 54–55] *for each* $0 \leqslant t \leqslant T$ *and* $1 \leqslant k \leqslant d$.

Lemma 4. *In the case of* $\ln n \leqslant d < \frac{n}{m}$ *the following upper bound is correct*

$$H \leqslant \frac{mn}{d}$$

for sums of constants $h_k = \frac{1}{(k+1)^2}$, *which correspond to the added edges in the constructed trees.*

Lemma 5. *For the case of* $\ln n \leqslant d < \frac{n}{\ln n}$ *the next inequality holds*

$$\mathbb{E}W'_{\mathcal{A}} \leqslant \frac{3mn}{d} = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Lemma 6. *For* $\frac{n}{\ln n} \leqslant d < \frac{n}{m}$ *the following inequality is true:*

$$\mathbb{E}W'_{\mathcal{A}} \leqslant 3m \ln n = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

With help of previous lemmas, it is possible to prove the main result of this section.

Theorem 2. *Let parameter* $d = d_n$ *is defined as* $\ln n \leqslant d < \frac{n}{m}$. *Then algorithm* $\mathcal{A}$ *for the* m*-*d*-*UMST *with weights of edges from* $\mathrm{UNI}(a_n; b_n)$ *is asymptotically optimal with failure probability* $\delta_n = n^{-m} \to 0$ *as* $n \to \infty$ *and the next conditions on scatter of weights of edges of graph* G

$$\frac{b_n}{a_n} = \begin{cases} o(d), & \text{if } \ln n \leqslant d < \dfrac{n}{\ln n}, \\ o\left(\dfrac{n}{\ln n}\right), & \text{if } \dfrac{n}{\ln n} \leqslant d < \dfrac{n}{m} \text{ and } m < \ln n. \end{cases} \tag{5}$$

3.2. Probability Distribution $\mathrm{EXP}(a_n, \lambda_n)$

We pass from random variables η, η_k to normalized random variables $\xi = \frac{\eta - a_n}{\lambda_n}$, $\xi_k = \frac{\eta_k - a_n}{\lambda_n}$ respectively. In the terms of this variables, p.d.f. is $\mathfrak{P}_\xi(x) = 1 - e^{-x}$ and probability density function is as follows (1)

$$p(\xi) = \begin{cases} e^{-\xi}, & \text{if } 0 \leqslant \xi < \infty, \\ 0 & \text{otherwise} \end{cases}$$

for random variable ξ. For random variable ξ_k, p.d.f. has the following form

$$\mathfrak{P}_{\xi_k}(x) = 1 - (1 - \mathfrak{P}_\xi(x))^k. \tag{6}$$

Lemma 7. *Mathematical expectation of random variable* ξ_k *equals* $\mathbb{E}\xi_k = 1/k$.

Lemma 8. *In the case of* $\mathrm{EXP}(a_n, \lambda_n)$*, the next upper bound is valid for expected value of a solution of algorithm* $\mathcal{A}$*:*

$$\mathbb{E}W'_{\mathcal{A}} \leqslant m \ln d + \frac{2mn}{d-1} = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Lemma 9. *Let* $T = \frac{1}{2}$*,* $h_k = \frac{3}{k^2}$*. Then for all* $1 \leqslant k \leqslant d$ *and* $0 \leqslant t \leqslant T$*, the conditions of Petrov's theorem* [15, pp. 54–55] *are true* $\mathbb{E}e^{t\widetilde{\xi}_k} \leqslant e^{\frac{h_k t^2}{2}}$ *for biased random variables* $\widetilde{\xi}_k = \xi_k - \mathbb{E}\xi_k$*.*

Lemma 10. *Let* $\ln n \leqslant d < \frac{n}{m}$*. Then for sufficiently large* n *the following upper bound is correct for the sums of constants* $h_k = \frac{3}{k^2}$*, which correspond to added edges in constructed trees*

$$H \leqslant \frac{3mn}{d-1}.$$

Lemma 11. *In the case of* $\ln n \leqslant d < \frac{n}{\ln n}$ *the next upper bound is valid*

$$\mathbb{E}W'_{\mathcal{A}} \leqslant \frac{3mn}{d-1} = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Lemma 12. *For* $\frac{n}{\ln n} \leqslant d < \frac{n}{m}$ *the inequality holds:*

$$\mathbb{E}W'_{\mathcal{A}} \leqslant 5m \ln n = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Using previous lemmas we can postulate the following theorem.

Theorem 3. *Let parameter* $d = d_n$ *be such that* $\ln n \leqslant d < \frac{n}{m}$*. Then algorithm* $\mathcal{A}$ *for solving the* m*-*d*-*UMST *with edge weights from* $\mathrm{EXP}(a_n, \lambda_n)$ *is asymptotically optimal with failure probability* $\delta_n = n^{-m} \to 0$ *as* $n \to \infty$ *under the conditions on scatter of weights of edges of graph* G*:*

$$\frac{\lambda_n}{a_n} = \begin{cases} o(d), & \text{if } \ln n \leqslant d < \dfrac{n}{\ln n}, \\ o\left(\dfrac{n}{\ln n}\right), & \text{if } \dfrac{n}{\ln n} \leqslant d < \dfrac{n}{m} \text{ and } m < \ln n. \end{cases} \tag{7}$$

3.3. Probability Distribution $\mathrm{NORM}(a_n, \sigma_n)$

For distribution $\mathrm{NORM}(a_n, \sigma_n)$, we introduce normalized random variables $\xi = \frac{\eta - a_n}{\sigma_n}$ and $\xi_k = \frac{\eta_k - a_n}{\sigma_n}$ for corresponding weights of edges of graph instead of η and η_k.
For random variable ξ, the following probability density function according to (1) and p.d.f. are true

$$p(\xi) = \begin{cases} \sqrt{\dfrac{2}{\pi}} \exp\left(-\dfrac{\xi^2}{2}\right), & \text{if } 0 \leqslant \xi < \infty, \\ 0 & \text{otherwise.} \end{cases} \qquad \mathfrak{F}(x) = \sqrt{\frac{2}{\pi}} \int\limits_0^x \exp\left(-\frac{u^2}{2}\right) du.$$

Definition 3. We say that p.d.f. $\mathfrak{F}_1(x)$ dominates p.d.f. $\mathfrak{F}_2(x)$, if $\mathfrak{F}_1(x) \geqslant \mathfrak{F}_2(x)$ for all x.

Statement 4. *The p.d.f.* $\mathfrak{F}(x)$ *of normal random variable with parameter* σ_n *dominates exponential p.d.f. with parameter* $\lambda_n = 2\sigma_n$*:*

$$\mathfrak{F}(x) \geqslant \mathfrak{P}(x/2) \ \ \forall x \geqslant 0. \tag{8}$$

Lemma 13 [16]. *Let $\chi_1, \ldots, \chi_k$ are independent identically distributed random variables with p.d.f. $F(x)$, $\widehat{F}(x)$ is p.d.f. of random variable $\chi = \min\limits_{i=1,\ldots,k} \chi_i$. Also $\zeta_1, \ldots, \zeta_k$ are independent identically distributed random variables with p.d.f. $G(x)$, analogically $\widehat{G}(x)$ is p.d.f. of random variable $\zeta = \min\limits_{i=1,\ldots,k} \zeta_i$. Then for all x*

$$F(x) \leqslant G(x) \Rightarrow \widehat{F}(x) \leqslant \widehat{G}(x).$$

Lemma 14 [16]. *Let $P_\vartheta, P_\omega, P_\zeta, P_\chi$ are p.d.f.'s of random variables $\vartheta, \omega, \zeta, \chi$ respectively, where ϑ and ζ are independent, also ω and χ are independent too. Then*

$$(\forall x\ P_\vartheta(x) \leqslant P_\omega(x)) \wedge (\forall y\ P_\zeta(y) \leqslant P_\chi(y)) \Rightarrow (\forall z\ P_{\vartheta+\zeta}(z) \leqslant P_{\omega+\chi}(z)).$$

Lemma 15 [16]. *Let p.d.f.'s $F(x)$ and $P(x)$ such that $F(x) \geqslant P(x)$ for all x. Then performance guarantees $(\varepsilon_{\mathcal{A}}, \delta_{\mathcal{A}})$ for algorithm $\mathcal{A}$ on inputs with p.d.f. $F(x)$ are the same as for inputs with p.d.f. $P(x)$.*

Let us put $F(x) = \mathfrak{F}(x)$ and $P(x) = \mathfrak{P}(x/2)$. From Statement 4 and Lemmas 13–15 the following theorem implies for biased truncated normal distribution.

Theorem 4. *Let parameter $d = d_n$ be defined as $\ln n \leqslant d < \frac{n}{m}$. Then algorithm $\mathcal{A}$ for the m-d-*UMST *in n-vertex complete undirected graph with edge weights from unbounded semiopen interval $[a_n; \infty)$ according to* NORM(a_n, σ_n) *asymptotically optimal with failure probability $\delta_n = n^{-m} \to 0$ as $n \to \infty$ and the next asymptotically optimal conditions:*

$$\frac{\sigma_n}{a_n} = \begin{cases} o(d), & \text{if } \ \ln n \leqslant d < \dfrac{n}{\ln n}, \\ o\left(\dfrac{n}{\ln n}\right), & \text{if } \ \dfrac{n}{\ln n} \leqslant d < \dfrac{n}{m} \ \text{ and } m < \ln n. \end{cases}$$

4. CONCLUSION

In this work, deterministic approximation algorithm, which solves the problem of finding several edge-disjoint spanning trees with given diameter in edge-weighted complete undirected graph, has been presented. This algorithm finds feasible solution in time $\mathcal{O}(n^2)$, where n is number of vertices in graph. The probabilistic analysis has been carried out for several probabilistic distributions of weights of edges of graph: uniform UNI$(a_n; b_n)$, biased truncated exponential EXP(a_n, λ_n), and biased truncated normal NORM(a_n, σ_n). Sufficient conditions of asymptotic optimality for this algorithm have been obtained in the case of each considered distribution. It would be interesting to investigate this problem on inputs with discrete probabilistic distributions. Also it would be useful to consider the problem of finding several edge-disjoint spanning trees of maximum total weight with given or bounded diameter.

FUNDING

The study was carried out within the framework of the state contract of the Sobolev Institute of Mathematics SB RAS (project FWNF-2022-0019).

APPENDIX

Proof of Statement 1. Each edge-disjoint construction consists of n vertices and $(n-1)$ edges, since, first of all, the $(d+1)$-vertex path is constructed on Step 1, and then all remaining vertices are connected to it without increasing diameter of spanning tree on Step 2 and 3. At the end, m such constructions are made presenting feasible solution for the m-d-UMST.

Proof of Statement 2. Preliminary Step 0 requires $\mathcal{O}(n)$ elementary operations.

On Step 1, each path is constructed in $\mathcal{O}(d^2)$ time, so this step is carried out completely with time complexity $\mathcal{O}(md^2)$ or $\mathcal{O}(nd)$ (because $m(d+1) \leqslant n$).

Each pair of paths (P_i, P_j), $1 \leqslant i < j \leqslant m$ is interconnected with $\mathcal{O}(d^2)$ actions on items 2.1–2.4. For all $\frac{m(m-1)}{2}$ pairs of paths, it is required $\mathcal{O}(m^2d^2)$ or $\mathcal{O}(n^2)$ elementary operations.

Items 2.5–2.6 are carried out with time complexity $\mathcal{O}(md)$.

Step 3 requires $\mathcal{O}(mdn)$ or $\mathcal{O}(n^2)$ time for connection of $|G(V')| < n$ vertices by shortest edges with inner vertices of path P_s in each spanning tree T_s, $1 \leqslant s \leqslant m$.

Thus, total time complexity of algorithm $\mathcal{A}$ is equal to $\mathcal{O}(n^2)$.

Proof of Lemma 1. Let us consider inequality (2) for performance guarantees of the quality of the algorithm in relation to the considered case of the minimum problem.

$$\begin{aligned}
\mathbb{P}\Big\{W_{\mathcal{A}} - OPT(I) > \varepsilon_n OPT(I)\Big\} &= \mathbb{P}\Big\{W_{\mathcal{A}} > (1+\varepsilon_n)OPT(I)\Big\} \\
&\leqslant \mathbb{P}\Big\{W_{\mathcal{A}} > (1+\varepsilon_n)m(n-1)a_n\Big\} \\
&= \mathbb{P}\Big\{m(n-1)a_n + \beta_n W'_{\mathcal{A}} > (1+\varepsilon_n)m(n-1)a_n\Big\} \\
&= \mathbb{P}\left\{W'_{\mathcal{A}} - \mathbb{E}W'_{\mathcal{A}} > \frac{\varepsilon_n m(n-1)a_n}{\beta_n} - \mathbb{E}W'_{\mathcal{A}}\right\} \\
&= \mathbb{P}\left\{\widetilde{W'_{\mathcal{A}}} > \frac{\varepsilon_n m(n-1)a_n}{\beta_n} - \mathbb{E}W'_{\mathcal{A}}\right\} \\
\leqslant \mathbb{P}\left\{\widetilde{W'_{\mathcal{A}}} > \frac{\varepsilon_n m(n-1)a_n}{\beta_n} - \widehat{\mathbb{E}W'_{\mathcal{A}}}\right\} &= \mathbb{P}\left\{\widetilde{W'_{\mathcal{A}}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\right\} = \delta_n,
\end{aligned}$$

the penultimate equality is true for $\varepsilon_n = \frac{2\beta_n \widehat{\mathbb{E}W'_{\mathcal{A}}}}{m(n-1)a_n}$.

Proof of Statement 3. It is easy to understand that

$$\sum_{k=1}^{d} \frac{1}{k} \leqslant 1 + \frac{1}{2} + \frac{1}{3} + \int_3^d \frac{dx}{x} = \frac{11}{6} + \ln d - \ln 3 \leqslant \ln d + \frac{3}{4}.$$

Proof of Lemma 2. It is easy to establish that $\mathbb{E}\xi_k = \frac{1}{k+1}$ for inputs $\mathrm{UNI}(a_n; b_n)$. Let us estimate from above each of the mathematical expectations of random variables W'_1, W'_2, and W'_3.

$$\mathbb{E}W'_1 = \sum_{s=1}^{m}\sum_{k=1}^{d} \mathbb{E}\xi_k = m\sum_{k=1}^{d} \frac{1}{k+1} \leqslant m \ln d,$$

penultimate inequality is correct for $d \geqslant 3$ due to the Statement 3 and relation

$$\sum_{k=1}^{d} \frac{1}{k+1} = \sum_{k=1}^{d} \frac{1}{k} - 1 + \frac{1}{d+1} \leqslant \sum_{k=1}^{d} \frac{1}{k} - \frac{3}{4} \leqslant \ln d.$$

$$\mathbb{E}W'_2 = C_m^2 \left(4\frac{d-1}{2}\mathbb{E}\xi_{(d-1)/2} + 4\mathbb{E}\xi_{(d-1)}\right) = \frac{m(m-1)}{2}\left(\frac{4(d-1)/2}{(d-1)/2+1} + \frac{4}{d}\right) \leqslant 2m^2;$$

$$\mathbb{E}W'_3 = m(n - m(d+1))\mathbb{E}\xi_{(d-1)} = m\frac{n - m(d+1)}{d} \leqslant \frac{mn}{d} - m^2.$$

Summing three inequalities and taking into account that $m(d+1) \leqslant n$, we obtain

$$\mathbb{E}W'_{\mathcal{A}} = \mathbb{E}\Big(W'_1 + W'_2 + W'_3\Big) \leqslant m \ln d + 2m^2 + \frac{mn}{d} - m^2 \leqslant m \ln d + \frac{2mn}{d}.$$

Proof of Lemma 3. Let's estimate $\mathbb{E}e^{t\xi_k}$ from above using formula

$$\mathbb{E}e^{t\xi_k} = \sum_{i=0}^{\infty} \frac{t^i}{(k+1)\cdots(k+i)}$$

from monograph [17, pp. 129]. Introducing also the notations $\alpha = \frac{t}{k+1}$ and

$$Q_{k,t} = \frac{(k+1)}{(k+2)(1-\frac{t}{k+3})} \leqslant Q_{k,T} = \frac{(k+1)(k+3)}{(k+2)^2} < 1$$

for all $t \leqslant T$ and for all natural k, we obtain

$$\mathbb{E}e^{t\xi_k} = \sum_{i=0}^{\infty} \frac{t^i}{(k+1)\cdots(k+i)} \leqslant 1 + \alpha + \alpha^2 Q_{k,t} \leqslant 1 + \alpha + \alpha^2 \leqslant e^{\alpha+\frac{\alpha^2}{2}} = e^{t\mathbb{E}\xi_k} e^{\frac{h_k t^2}{2}},$$

since $\mathbb{E}\xi_k = \frac{1}{k+1}$ for inputs UNI$(a_n; b_n)$.
Consequently,

$$\mathbb{E}e^{t(\xi_k - \mathbb{E}\xi_k)} = \mathbb{E}e^{t\widetilde{\xi}_k} \leqslant e^{\frac{h_k t^2}{2}},$$

where $\widetilde{\xi}_k = \xi_k - \mathbb{E}\xi_k$.

Proof of Lemma 4. In the case of $\ln n \leqslant d < \frac{n}{m}$, parameter H is equal to sum of H_1, H_2, and H_3 according to steps of algorithm $\mathcal{A}$. Taking into account the notations and estimates obtained earlier, we arrive at the following:

$$H_1 = m\sum_{k=1}^{d} h_k = m\sum_{k=1}^{d} \frac{1}{(k+1)^2} < \psi m,$$

where $\psi \approx 0.645$. Here we use Euler estimation for the sum of inverse squares $1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\ldots = \frac{\pi^2}{6} < 1.645$.

$$\begin{aligned} H_2 = 4C_m^2\left(\frac{d-1}{2}h_{(d-1)/2} + h_{(d-1)}\right) &\leqslant 2m^2\left(\frac{(d-1)/2}{((d-1)/2+1)^2} + \frac{1}{d^2}\right) \\ &= 2m^2\left(\frac{2(d-1)}{(d+1)^2} + \frac{1}{d^2}\right) \leqslant 4m^2\frac{d}{(d+1)^2}. \end{aligned}$$

The last inequality holds for $d \geqslant 3$.

$$H_3 = m(n - m(d+1))h_{(d-1)} \leqslant \frac{mn}{d^2} - m^2\frac{d}{(d+1)^2}.$$

Since $n \geqslant m(d+1)$ and $m \geqslant 2$, we get

$$\begin{aligned} H = H_1 + H_2 + H_3 &< \psi m + 4m^2\frac{d}{(d+1)^2} + \left(\frac{mn}{d^2} - m^2\frac{d}{(d+1)^2}\right) \\ \leqslant \frac{mn}{d}\left(\frac{d\psi}{n} + \frac{1}{d}\right) + 3m^2\frac{d}{(d+1)^2} &\leqslant \left(\frac{\psi d}{2(d+1)} + \frac{1}{d} + \frac{3d^2}{(d+1)^3}\right)\frac{mn}{d}. \end{aligned}$$

It is easy to verify that the expression in parentheses is less than 1 for all $d \geqslant 3$. Then we can obtain the next estimation $H \leqslant \frac{mn}{d}$.

Proof of Lemma 5. Taking into account that $\ln d \leqslant \ln n$ and $d < \frac{n}{\ln n}$, it is true that

$$\mathbb{E}W'_{\mathcal{A}} \leqslant m\ln d + \frac{2mn}{d} \leqslant m\ln n + \frac{2mn}{d} < m\frac{n}{d} + \frac{2mn}{d} = \frac{3mn}{d} = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Proof of Lemma 6. Because of Lemma 2, $\ln d \leqslant \ln n$, and $\frac{n}{d} \leqslant \ln n$, we get

$$\mathbb{E}W'_{\mathcal{A}} \leqslant m \ln d + \frac{2mn}{d} \leqslant 3m \ln n = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Proof of Theorem 2. First of all, it must be noted that in the course of the Algorithm $\mathcal{A}$ we have deal with random variables of the type ξ_k, $1 \leqslant k \leqslant d$. In the case of graphs with weights of edges from UNI$(a_n; b_n)$ these biased variables satisfy the conditions $\mathbb{E}e^{t\widetilde{\xi_k}} \leqslant e^{\frac{h_k t^2}{2}}$ of Petrov's theorem [15, pp. 54–55] for constants $T = 1$ and $h_k = \frac{1}{(k+1)^2}$ (see Lemma 3).

We will carry out the proof of the theorem for two cases of possible semiopen intervals of parameter d.

$$\textit{Case 1:} \quad \ln n \leqslant d < \frac{n}{\ln n}.$$

According to Lemma 5 and formula (3) for relative error, we obtain

$$\varepsilon_n = \frac{2(b_n - a_n)}{m(n-1)a_n} \widehat{\mathbb{E}W'_{\mathcal{A}}} = \frac{2(b_n - a_n)}{m(n-1)a_n} \frac{3mn}{d} \leqslant \frac{6n}{(n-1)} \frac{b_n/a_n}{d}.$$

We can see that $\varepsilon_n \to 0$ as $n \to \infty$, if the following conditions are satisfied on scatter of weights of edges of graph G: $\frac{b_n}{a_n} = o(d_n)$.

Using Lemmas 1 and 5, we can estimate failure of probability:

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\} = \mathbb{P}\Big\{\widetilde{W'_{\mathcal{A}}} > \frac{3mn}{d}\Big\}.$$

From Lemma 4 and inequality $d < \frac{n}{\ln n}$ it follows that $TH \leqslant \frac{mn}{d} < \frac{3mn}{d} = x$. According to Petrov's theorem [15, pp. 54–55], we get the next estimate of failure probability of algorithm $\mathcal{A}$: $\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > x\} \leqslant \exp\Big\{-\frac{Tx}{2}\Big\}$.

Since $\ln n < \frac{n}{d}$ and $\frac{Tx}{2} = \frac{3mn}{2d} > m \ln n$, then

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > x\} \leqslant \exp\left\{-\frac{Tx}{2}\right\} < \exp(-m \ln n) = \frac{1}{n^m} \to 0 \text{ as } n \to \infty.$$

Therefore, in Case 1 Algorithm $\mathcal{A}$ gives asymptotically optimal solution for the m-d-UMST in graph with weights of edges from UNI$(a_n; b_n)$.

$$\textit{Case 2:} \quad \frac{n}{\ln n} \leqslant d < \frac{n}{m}.$$

According to Lemma 6 and formula (3) we get the following equation for relative error:

$$\varepsilon_n = \frac{2(b_n - a_n)}{(n-1)a_n} \widehat{\mathbb{E}W'_{\mathcal{A}}} = \frac{2(b_n - a_n)}{m(n-1)a_n} 3m \ln n \leqslant \frac{6(b_n/a_n) \ln n}{(n-1)}.$$

It is clear that $\varepsilon_n \to 0$ as $n \to \infty$, if $\frac{b_n}{a_n} = o(\frac{n}{\ln n})$.

Now using Lemmas 1 and 6, we can estimate failure probability $\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\} = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > 3m \ln n\}$. If $T = 1$, $d \geqslant \frac{n}{\ln n}$ and bearing in mind Lemma 4, then the next inequality is valid: $TH \leqslant \frac{mn}{d} < 3m \ln n = x$. Since $\frac{Tx}{2} > m \ln n$ and Petrov's theorem [15, pp. 54–55] we obtain the following estimate for failure probability of algorithm $\mathcal{A}$: $\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > x\} \leqslant \exp\{-\frac{Tx}{2}\} \leqslant \exp(-m \ln n) = \frac{1}{n^m} \to 0$.

From this it follows that in the Case 2 Algorithm $\mathcal{A}$ gives asymptotically optimal solution for the problem m-d-UMST on n-vertex complete undirected graph with weights of edges from UNI$(a_n; b_n)$.

We conclude, that within the values of the parameter d for both cases, under conditions (5) we have that estimates of the relative error $\varepsilon_n \to 0$ and failure probability $\delta_n \to 0$ as $n \to \infty$.

Proof of Lemma 7. With reference to (6) we obtain

$$\mathbb{E}\xi_k = \int\limits_0^\infty x d\,\mathfrak{P}_{\xi_k}(x) = \int\limits_0^\infty xk(1-\mathfrak{P}_\xi(x))^{k-1} d\,\mathfrak{P}_\xi(x) = \int\limits_0^\infty xke^{-kx}dx$$

$$= -x\,e^{-kx}\Big|_0^\infty + \int\limits_0^\infty e^{-kx}dx = -\frac{1}{k}\,e^{-kx}\Big|_0^\infty = \frac{1}{k}.$$

Proof of Lemma 8. Let us estimate each expected value for random variables W_1', W_2', and W_3':

$$\mathbb{E}W_1' = \sum_{s=1}^{m}\sum_{k=1}^{d}\mathbb{E}\xi_k = m\sum_{k=1}^{d}\frac{1}{k} \leqslant m\left(\ln d + \frac{3}{4}\right)$$

taking into account Statement 3 and Lemma 7;

$$\mathbb{E}W_2' = C_m^2\left(4\frac{d-1}{2}\mathbb{E}\xi_{(d-1)/2} + 4\mathbb{E}\xi_{d-1}\right) = 2m(m-1)\left(1+\frac{1}{d-1}\right) \leqslant \frac{2d}{d-1}m^2 - 2m;$$

$$\mathbb{E}W_3' = m(n-m(d+1))\mathbb{E}\xi_{(d-1)} = m\frac{n-m(d+1)}{d-1} = \frac{mn}{d-1} - \frac{d+1}{d-1}m^2.$$

Adding the left and right parts of three ratios for $\mathbb{E}W_1'$, $\mathbb{E}W_2'$, $\mathbb{E}W_3'$ and bearing in mind that $m \leqslant \frac{n}{d+1}$, we get

$$\mathbb{E}W_{\mathcal{A}}' \leqslant m\ln d - \frac{5}{4}m + \frac{mn}{d-1} + m^2 \leqslant m\ln d + \frac{mn}{d-1} + \frac{mn}{d+1} \leqslant m\ln d + \frac{2mn}{d-1}.$$

Proof of Lemma 9. The following is true for quantities $\mathbb{E}e^{t\xi_k}$ according to formula (6).

$$\mathbb{E}e^{t\xi_k} = \int\limits_0^\infty e^{tx} d\,\mathfrak{P}_{\xi_k}(x) = \int\limits_0^\infty e^{tx}ke^{-kx}dx = \int\limits_0^\infty ke^{-(k-t)x}dx$$

$$= -\frac{k}{k-t}e^{-(k-t)x}\Big|_0^\infty = \frac{1}{1-t/k} = \sum_{s=0}^{\infty}\left(\frac{t}{k}\right)^s \leqslant 1 + \frac{t}{k} + \left(\frac{t}{k}\right)^2\frac{1}{1-t/k}.$$

Taking into account the inequality $\frac{t}{k} \leqslant \frac{1}{2}$, which is true under the conditions of the lemma, we estimate the value $\mathbb{E}e^{t\xi_k}$ from above:

$$\mathbb{E}e^{t\xi_k} \leqslant 1 + \frac{t}{k} + 2\left(\frac{t}{k}\right)^2 = 1 + \frac{t}{k} + \frac{1}{2}\left(\frac{t}{k}\right)^2 + \frac{3}{2}\left(\frac{t}{k}\right)^2$$

$$\leqslant \left(1 + \frac{t}{k} + \frac{1}{2}\left(\frac{t}{k}\right)^2\right)\left(1 + \frac{3}{2}\left(\frac{t}{k}\right)^2\right) \leqslant e^{t/k}\exp\left(\frac{3}{2}\left(\frac{t}{k}\right)^2\right) = e^{t\mathbb{E}\xi_k}\exp\left(\frac{h_k t^2}{2}\right),$$

because of Lemma 7 $\mathbb{E}\xi_k = \frac{1}{k}$ for EXP(a_n, λ_n). Consequently the conditions $\mathbb{E}e^{t\widetilde{\xi_k}} \leqslant e^{\frac{h_k t^2}{2}}$ of Petrov's theorem are true for constants $T = 1/2$, $h_k = 3/k^2$.

Proof of Lemma 10. For $\ln n \leqslant d < \frac{n}{m}$ parameter H is equal to sum of quantities H_1, H_2, and H_3 according to steps of algorithm $\mathcal{A}$. Taking into account previous notation and obtained estimations, we have

$$H_1 = m\sum_{k=1}^{d} h_k = m\sum_{k=1}^{d}\frac{3}{k^2} < 3(1+\psi)m < 5m,$$

where ψ equals Euler estimation for the sum of inverse squares minus 1 ($\psi \approx 0.645$).

$$H_2 = 4C_m^2\left(\frac{d-1}{2}h_{(d-1)/2} + h_{(d-1)}\right) = 6m(m-1)\left(\frac{2}{d-1} + \frac{1}{(d-1)^2}\right) \leqslant 6m^2\frac{2d-1}{(d-1)^2}.$$

$$H_3 = m\big(n - m(d+1)\big)h_{(d-1)} = \frac{3}{(d-1)^2}\Big(mn - m^2(d+1)\Big) \leqslant \frac{3mn}{(d-1)^2} - 3m^2\frac{d+1}{(d-1)^2}.$$

With $n \geqslant m(d+1)$ and $m \geqslant 2$ we get

$$\begin{aligned} H = H_1 + H_2 + H_3 &< 5m + \frac{3m^2}{(d-1)^2}\Big((4d-2) - (d+1)\Big) + \frac{3mn}{(d-1)^2} \\ &= 3m\left(\frac{5}{3} + \frac{3m}{d-1}\right) + \frac{3mn}{(d-1)^2} \\ \leqslant \frac{3n}{d+1}\left(\frac{5}{3} + \frac{3m}{d-1}\right) + \frac{3mn}{(d-1)^2} &\leqslant \frac{3n}{d+1}\left(\frac{5m}{6} + \frac{3m}{d-1}\right) + \frac{3mn}{(d-1)^2} \\ &= \frac{3mn}{d-1}\left(\frac{5(d-1)}{6(d+1)} + \frac{3}{d+1} + \frac{1}{d-1}\right) \leqslant \frac{3mn}{d-1}. \end{aligned}$$

The last sign of inequality is due to the fact that, when n is sufficiently large, the value in parentheses is less than 1 since $d \geqslant \ln n$.

Proof of Lemma 11. Taking into account $\ln d \leqslant \ln n$, $d < \frac{n}{\ln n}$ and Lemma 8 we obtain

$$\mathbb{E}W'_{\mathcal{A}} \leqslant m\ln d + \frac{2mn}{d-1} \leqslant m\frac{n}{d} + \frac{2mn}{d-1} \leqslant \frac{3mn}{d-1} = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Proof of Lemma 12. According to Lemma 8 and inequality $\ln d \leqslant \ln n$ and $n \leqslant d\ln n$ we get

$$\mathbb{E}W'_{\mathcal{A}} \leqslant m\ln d + \frac{2mn}{d-1} \leqslant m\ln n + \frac{2md}{d-1}\ln n \leqslant 5m\ln n = \widehat{\mathbb{E}W'_{\mathcal{A}}}.$$

Proof of Theorem 3. First of all, let us note that random variables $\widetilde{\xi}_k = \xi_k - \mathbb{E}\xi_k$ satisfy conditions $\mathbb{E}e^{t\widetilde{\xi}_k} \leqslant e^{\frac{h_k t^2}{2}}$ of Petrov's theorem for constants $T = 1/2$ and $h_k = \frac{3}{k^2}$ (see Lemma 9).

Let's carry out the proof of the theorem for two cases of possible semiopen intervals of the value of the parameter d.

$$\textit{Case 1:}\quad \ln n \leqslant d < \frac{n}{\ln n}.$$

Bearing in mind Lemma 11 and formula (3) for relative error we obtain

$$\varepsilon_n = \frac{2\lambda_n}{m(n-1)a_n}\frac{3mn}{(d-1)} \leqslant \frac{6n}{(n-1)}\frac{\lambda_n/a_n}{(d-1)}.$$

So we can see that $n \to \infty$ as $\varepsilon_n \to 0$, if $\frac{\lambda_n}{a_n} = o(d_n)$.

Now using Lemmas 1 and 11, we can estimate failure probability:

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\} = \mathbb{P}\Big\{\widetilde{W'_{\mathcal{A}}} > \frac{3mn}{d-1}\Big\} = \mathbb{P}\Big\{\widetilde{W'_{\mathcal{A}}} > \frac{3mn}{d-1}\Big\}.$$

For each edge with weight, which corresponds to random variable ξ_k, we define constants $T = 1/2$ and $h_k = \frac{3}{k^2}$.

From Lemma 10, it implies that $TH \leqslant \frac{3mn}{2(d-1)} < \frac{3mn}{d-1} = x$.

According to Petrov's theorem we get the next estimate for failure probability of algorithm $\mathcal{A}$:

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > x\} \leqslant \exp\left\{-\frac{Tx}{2}\right\}.$$

Since $\frac{n}{d} > \ln n$, then $\frac{Tx}{2} = \frac{3mn}{2(d-1)} > m \ln n$. So we get that

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > x\} \leqslant \exp\left\{-\frac{Tx}{2}\right\} < \exp(-m \ln n) = \frac{1}{n^m} \to 0 \text{ as } n \to \infty.$$

Thus, in Case 1, Algorithm $\mathcal{A}$ gives asymptotically optimal solution for the m-d-UMST in n-vertex complete undirected graph with weights of edges from $\mathrm{EXP}(a_n, \lambda_n)$.

$$\textit{Case 2:} \quad \frac{n}{\ln n} \leqslant d < \frac{n}{m}.$$

Knowing Lemma 12 and formula (3) for relative error ε_n we get

$$\varepsilon_n = \frac{2\lambda_n}{(n-1)a_n}\widehat{\mathbb{E}W'_{\mathcal{A}}} = \frac{2\lambda_n}{m(n-1)a_n} 5m \ln n \leqslant \frac{10(\lambda_n/a_n)\ln n}{n-1}.$$

It is clear that $\varepsilon_n \to 0$ as $n \to \infty$, if the conditions $\frac{\lambda_n}{a_n} = o\Big(\frac{n}{\ln n}\Big)$ are satisfied.

So using Lemmas 1 and 12, we can estimate failure probability

$$\delta_n = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > \widehat{\mathbb{E}W'_{\mathcal{A}}}\} = \mathbb{P}\{\widetilde{W'_{\mathcal{A}}} > 5m \ln n\}.$$

Putting constants h_k as in the Case 1, we set $T = 1/2$ and $x = 5m \ln n$.

Taking into account Lemma 10, quantities x, T, H, and $d \geqslant \frac{n}{\ln n}$, we arrive at the following inequality $TH \leqslant \frac{3mn}{2d} < 5m \ln n = x$.

Since $\frac{Tx}{2} > m \ln n$, according to Petrov's theorem we obtain the next estimate for failure probability of Algorithm $\mathcal{A}$:

$$\delta_n = \mathbb{P}\left\{\widetilde{W'_{\mathcal{A}}} > x\right\} \leqslant \exp\left\{-\frac{Tx}{2}\right\} \leqslant \exp(-m \ln n) = \frac{1}{n^m} \to 0 \text{ as } n \to \infty.$$

Consequently, in Case 2, Algorithm $\mathcal{A}$ gives asymptotically optimal solution for the m-d-UMST in n-vertex complete undirected graph with weights of edges from $\mathrm{EXP}(a_n, \lambda_n)$.

Therefore, for values of parameter d we have estimate of relative error $\varepsilon_n \to 0$ and failure probability $\delta_n \to 0$ as $n \to \infty$ in both cases under conditions (7).

Proof of Statement 4. We present the proof of this statement, as in [18]. The difference of left and right sides of inequality (8) is denoted as

$$h(x) = \sqrt{\frac{2}{\pi}} \int_0^x e^{-\frac{u^2}{2}} du - (1 - e^{-\frac{x}{2}}).$$

It is easy to check that for function $h(x)$ and its derivative

$$h'(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} - \frac{1}{2} e^{-\frac{x}{2}}$$

the following is correct $h(0) = 0$, $\lim\limits_{x\to\infty} h(x) = 0$, $h'(x) > 0$.

Since on positive positive semiaxis, the inequality $h'(x) = 0$ holds only in one unique point $x_0 = \frac{1}{2}(1 + \sqrt{1 + 12\ln(2) - 4\ln(\pi)} \geqslant 0$, we can conclude that $h(x) \geqslant 0$ as $x \geqslant 0$, which implies the validity of the statement.

REFERENCES

1. Frieze, A., On the Value of a Random MST Problem, *Discrete Applied Mathematics*, 1985, vol. 10, pp. 47–56. https://doi.org/10.1016/0166-218X(85)90058-7
2. Angel, O., Flaxman, A.D., and Wilson, D.B., A Sharp Threshold for Minimum Bounded-Depth and Bounded-Diameter Spanning Trees and Steiner Trees in Random Networks, *Combinatorica*, 2012, vol. 32, pp. 1–33. https://doi.org/10.1007/s00493-012-2552-z
3. Cooper, C., Frieze, A., Ince, N., Janson, S., and Spencer, J., On the Length of a Random Minimum Spanning Tree, *Combinatorics, Probability and Computing*, 2016, vol. 25, no. 1, pp. 89–107. https://doi.org/10.1017/S0963548315000024
4. Garey, M.R. and Johnson, D.S., *Computers and Intractability*, 1979, San Francisco: Freeman, 340 p.
5. Clementi, A.E.F., Ianni, M.D., Monti, A., Rossi, G., and Silvestri, R., Experimental Analysis of Practically Efficient Algorithms for Bounded-Hop Accumulation in Ad-Hoc Wireless Networks *Proc. 19th IEEE Int. Parallel Distributed Processing Symposium (IPDPS'05)*, 2005, pp. 8–16. https://doi.org/ 10.1109/IPDPS.2005.210
6. Bala, K., Petropoulos, K., and Stern, T.E., Multicasting in a Linear Lightwave Network, *Proc. of IEEE INFOCOM'93*, 1993, pp. 1350–1358. https://doi.org/10.1109/INFCOM.1993.253399
7. Bookstein, A. and Klein, S.T., Compression of Correlated Bit-Vectors, *Inform. Syst.*, 1996, vol. 16, no. 4, pp. 110–118.
8. Raymond, K., A Tree-Based Algorithm for Distributed Mutual Exclusion, *ACM Trans. on Comput. Syst.*, 1989, vol. 7, no. 1, pp. 61–77. https://doi.org/10.1145/58564.59295
9. Gruber, M., Exact and Heuristic Approaches for Solving the Bounded Diameter Minimum Spanning Tree Problem, *PhD Thesis*, Vienna University of Technology, 2009.
10. Gimadi, E.Kh. and Serdyukov, A.I., A Probabilistic Analysis of Approximation Algorithm for the Minimum Weight Spanning Tree Problem with a Bounded Below Diameter, in *Oper. Res. Proceed.*, vol. 99, Inderfurth, K. et. al, Eds., Berlin: Springer, 2000, pp. 63–68. https://doi.org/10.1007/978-3-642-58300-1_12
11. Gimadi, E.Kh., Istomin, A.M., and Shin, E.Yu., On Algorithm for the Minimum Spanning Tree Problem Bounded Below, *Proc. Conference DOOR 2016*, Vladivostok, Russia, CEUR-WS, vol. 1623, 2016, pp. 11–17.
12. Gimadi, E.Kh. and Shin, E.Yu., On Given Diameter MST Problem on Random Input Data, in *MOTOR 2019. Communications in Computer and Information Science, vol. 1090*, Bykadorov, I., Strusevich, V., and Tchemisova, T., Eds., Cham: Springer, 2019, pp. 30–38. https://doi.org/10.1007/978-3-030-33394-2_3
13. Gimadi, E.Kh., Shevyakov, A.S., and Shtepa, A.A., A Given Diameter MST on a Random Graph, in *Optimization and Applications — 11th International Conference OPTIMA*, Olenev, N., Evtushenko, Y., Khachay, M., and Malkova, V., Eds., 2020, LNCS, vol. 12422, pp. 110–121. https://doi.org/10.1007/978-3-030-62867-3_9

14. Gimadi, E.Kh., Glebov, N.I., and Perepelitsa, V.A., Algorithms with Bounds for Discrete Optimization Problems, *Problemy Kibernetiki*, 1975, no. 31, pp. 35–42 (in Russian).

15. Petrov, V.V., *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*, 1995, Oxford: Clarendon Press, 304 p.

16. Gimadi, E.Kh. and Glazkov, Yu.V., An Asymptotically Exact Algorithm for One Modification of Planar Three-Index Assignment, *J. Appl. Industr. Math.*, 2007, vol. 1, no. 4, pp. 442–452.

17. Gimadi, E.Kh. and Khachay, M.Yu., *Extremal Problems on Sets of Permutations*, Ekaterinburg: UMC UPI, 2016, 219 p. (in Russian).

18. Gimadi, E.Kh. and Shin, E.Yu., Probabilistic Analysis of an Algorithm for the Minimum Spanning Tree Problem with Diameter Bounded Below, *J. Appl. Industr. Math.*, 2015, vol. 9, pp. 480–488. https://doi.org/10.1134/S1990478915040043

This paper was recommended for publication by A.A. Lazarev, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 889–890.

OBITUARY

August 31, 1931—April 25, 2023

Lev Rozonoer, in full Lev Il'ich Rozonoer, was an outstanding Soviet and Russian scientist. Unfortunately, he passed away in Newton (Boston, the USA) at the end of April 2003.

Rozonoer joined the Institute of Automation and Remote Control (IARC), the USSR Academy of Sciences, in 1955 after graduating from Moscow Power Engineering Institute. (Nowadays, IARC is the Trapeznikov Institute of Control Sciences, the Russian Academy of Sciences, simply called the Institute below.) He had worked at IARC for over 40 years until he left for family reasons for the USA (1996).

Rozonoer undoubtedly belongs to those outstanding scientists who have laid the foundations of control theory through their research. Rozonoer became globally recognized for his investigations on optimal control. First and foremost, he formulated and proved the correctness of a fundamentally new concept, subsequently developed by mathematicians from the Steklov Institute and called Pontryagin's maximum principle. This concept was presented by Rozonoer in his candidate's dissertation in 1966, and some members of the dissertation council even suggested conferring the doctoral degree to him. The main results of that study were published in three issues of *Automation and Remote Control.* In 1970, Rozonoer defended his doctoral dissertation. Other significant Rozonoer's results at IARC were connected with thermodynamics (optimal control of thermodynamic processes) and systems theory (information aggregation in large-scale systems and models of biological evolution). They were published in Russia as well as in leading foreign journals.

For many years, Rozonoer was a member of the Editorial Board of *Automation and Remote Control.* Dozens of his fundamental works first appeared in the journal.

Rozonoer was a brilliant teacher. Having delved into a new science, he immediately shared accumulated knowledge with his students at the Moscow Institute of Physics and Technology or Institute's employees. Such lectures attracted a large audience.

Rozonoer's uniqueness as a scientist was the breadth of his research interests rather than his separate results (no matter how significant they were). At the Institute, he was actively involved

in research on pattern recognition, mathematical logic, and the theory of algorithms and finite automata. But the real breadth of thought showed up during his life in the USA. In 2018, Fizmatlit, a famous Russian publishing house, released his book *Poslednie teksty. Teoriya sistem. Fizika. Chelovek, nauka, sotsium* (Recent Writings. Systems Theory. Physics. Man, Science and Socium). The book included only Rozonoer's results obtained after 2000. These results have turned out unexpected even for his colleagues at the Institute. For example, note his work entitled in the manner of Kant: *How the Science of the Spirit is Possible*. It attempts to lay the foundations of a scientific approach to studying the human spirit, a mysterious phenomenon that was previously considered only in philosophy and religion.

However, Rozonoer considered his main result to be the hypothesis of the random flow of time at the micro level as the cause of decoherence in quantum mechanics. Decoherence is a phase shift that spontaneously arises in quantum mechanical systems; as a result of this shift, the laws of the microworld turn into the laws of the macroworld. The random process-based conceptualization of time at the micro level seems so unusual that it has caused confusion among physicists: they will have to puzzle it out. It will be uneasy since experts in quantum mechanics are not deeply versed in the theory of random processes.

The Institute is proud that such a huge scientist grew up and worked in it for many years!

Employees of the Trapeznikov Institute of Control Sciences, the Russian Academy of Sciences

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 7, pp. 891–899.

AUTOMATION IN INDUSTRY

Soft Sensors Based on Digital Models

A. A. Chereshko

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
e-mail: chereshko@phystech.edu

Received May 16, 2022
Revised May 20, 2022
Accepted June 17, 2022

Abstract—The article proposes a method for creating soft sensors using identification models obtained by associative search algorithm. The method consists in constructing an approximating hypersurface of the space of input vectors and their corresponding one-dimensional outputs at each time instant. Case studies are presented and the advantages of the author's method over traditional approaches are evaluated are revealed.

Keywords: soft sensor, model predictive control (MPC), identification, machine learning, clustering methods, associative search

DOI: 10.25728/arcRAS.2023.93.50.001

1. INTRODUCTION

Product quality is the most important parameter in the process optimization. However, quality metrics are difficult to measure in real time because product properties can change over time due to many factors. At modern manufacturing enterprises, various methods and means of physical and chemical physical and chemical analysis of product samples are widely used. Nevertheless, the use of embedded analysis systems is not always justified due to their inertia and high cost. Thats why soft sensors have gained popularity as a low-cost alternative to complex analytical systems. Soft sensors make predictions of product quality in MPC [1, 2]. They allow to control product properties in real time with an acceptable accuracy at relatively modest deployment and maintenance costs. Soft sensors can control physical and chemical properties, that cannot be controlled by conventional analyzers. They can also be used to monitor product quality, where the use of inline analyzers is economically feasible or technically impossible [1]. Soft sensors based on linear regression with automatic fit of free term according to laboratory control data have performed well in practice, but for some nonlinear objects they give inadequate prediction. The article offers a new approach to predicting quality metrics for a wide class of processes described by non-linear models. Models based on associative search algorithms use the knowledge base of processes to build, at each time step, the best model by the least squares method (LSM). A case studies was conducted based on ore grinding process data, the advantages of the proposed method in comparison with the traditional ones are shown.

2. CLASSICAL METHODS FOR SOFT SENSORS CONSTRUCTING

Model Predictive Control (MPC) allows direct control of the products quality, that is evaluated by soft sensors. Usually, soft sensors in MPC systems are realized in the form of simple regression models:

$$Y = \sum_{1}^{N} b_i x_i + b_0, \tag{1}$$

where Y is product quality, x_i are the input variables, b_i are the coefficients for each input variable, b_0 is the free term of linear regression.

Model (1) is developed on the basis of historical process data. Factory data often contain many outliers, which can lead to poor modeling. It is easy to show that a least squares method (LSM) model is highly distorted by even a few noticeable outliers. There is a so-called weighted least squares (WLS) method, an improved method of constructing a linear regression, when individual outliers in the data do not distort the constructed model so much.

For soft sensors of the form (1) we often use the algorithm for constructing a free linear regression term on the basis of the laboratory control data. For this purpose, the coefficient b_0 is recalculated as follows:

$$b_{0new} = b_0 + k(Y_{lab} - Y_{model}), \tag{2}$$

where Y_{lab} is the laboratory quality score, Y_{model} is the quality score calculated by a model, b_{0new} is the new value of free term b_0, k is the weight coefficient of accounting for incoming laboratory control data. Soft sensors based on linear regression with automatic fit of free term b_0 proved to be a good solution to various practical problems. For nonlinear processes, such models may not be satisfactory. In this case, it seems reasonable to use an approach to predicting quality metrics based on data mining of the object functioning and formation of an inductive knowledge base for it.

3. SOFT SENSORS BASED ON INTELLIGENT ANALYSIS OF TECHNOLOGICAL DATA

Models based on associative search algorithms use the inductive knowledge base of process to build, at each time step, the best model by the LSM. The concept of inductive knowledge — the regularities extracted from the data of the object functioning was introduced by V.N. Vapnik [3]. Such models are formed at each time step by the system identifier on the basis of the analysis of the information on the process dynamics accumulated by current time instant. This information makes possible to replenish inductive knowledge base and additionally train the system.

The model, generated by the soft sensors at a certain time instant, replenishes at each time step the appropriate "library of models" in the knowledge base. This digital model is fully characterized by a set of values of the following attributes: inputs, output, coefficients. Further, these models can be used in the traditional MPC scheme.

The models can be derived using well-known identification methods, such as LSM. Prediction with the associative search method [4] for solving control problems has high accuracy of identification model for a wide class of nonlinear and nonstationary objects [5, 6]. In addition, pre-training (clustering) in real time provides the algorithm with a speed gain, which may be important for a certain class of process control problems.

The identification model is fully described by the sets of inputs and corresponding outputs of the system, that are stored in the archive, and in this aspect, it can be considered as a digital model. The combination of statistical data sets (feature values) gives a "digital portrait" of the process dynamics.

4. DESCRIPTION OF THE ASSOCIATIVE SEARCH ALGORITHM

The process of inductive knowledge is reduced to the restoration (associative search) of knowledge by its fragment [7, 8]. In this case knowledge can be interpreted as an associative connection between images. As an image, we will use the vectors of inputs, that is, input variables.

The criterion of closeness of images can be formulated in different ways. In the most general case, it can be represented as a logical function, that is, a predicate. In the particular case, when

the sets of features are vectors in n-dimensional space, the proximity criterion may be a distance in this space.

Associative search can be performed as a process of either restoring an image according to partially specified features (or restoring a fragment of knowledge under conditions of incomplete information. This process is simulated in various models of associative memory), or searching for other images related associatively to the given one, but representing other time instants.

Various schemes of associative search are known [9]. Thus, in frame-based systems, the search task is implemented in the form of matching frames. In semantic networks the search is performed by matching fragments of the network and the graph-query.

The approach based on the method of verbal analysis of decisions proved effective for solving discrete multi-criteria selection problems [10]. This approach involves decomposition of the description of objects by many criteria into their partial descriptions of smaller dimensions, which are offered to the decision maker for comparison (under the assumption of pairwise equal evaluations by the criteria that are not included in such descriptions).

There is a well-known model that describes the process of associative thinking as a sequential recall based on the application of associations — pairs of images characterized by their own set of features. This model appears to be an intermediate stage between neural network models and logical models used in classical artificial intelligence systems.

The statement of the Y signal modeling problem is a result of operator's influence (in general case, nonlinear) on vector signal $x_1, x_2, \ldots, x_n$ in discrete time [11]. At any chosen time instant a new linear model in the neighborhood of the operating point is created (instead of approximation of real signal in time).

The linear dynamic model has the following form:

$$Y = \sum_{i=1}^{m} a_i y_{N-i} + \sum_{j=1}^{n} \sum_{s=1}^{S} b_{js} x_{N-j,s},\ \forall j \in 1, \ldots, N, \tag{3}$$

where y_N is the output prediction at the moment N, x_N is the input vector, m is the memory depth for output, n is the memory depth for input, S is the length of the vector of inputs.

This model is not a classical regression model: not the entire dynamic chronological "tail" is selected, but only certain inputs, in accordance with a given criterion. The coefficients at other inputs are assumed to be zero.

To construct a virtual model corresponding to a certain point in time, the input vectors, close to the current one according to the criterion, are selected from the archive. Then, based on the classical (nonrecurrent) LSM, the value of the output at the next point in time is calculated.

Described algorithm does not create a single approximating model of the real process for all moments of time, but builds a new model for each fixed moment t (Fig. 1). In this case each point of the global nonlinear regression surface is formed as a result of using "local" linear models.

In contrast to classical regression models, for each fixed point in time, input vectors are selected from the archive that are close to the current input vector according to a certain criterion rather than in chronological order by steps backwards.

The selection criteria are called associative search criteria or associative pulses. Thus, in (3) n represents the number of vectors from the archive (from time moment 1 to moment N) selected by the associative search criterion. At each time interval $\{N-1, N\}$ a certain set of n vectors is chosen, $1 \leqslant n \leqslant N$.

The criterion for selecting input vectors from the archive to build a virtual model at a given time by the current state of the object can be as follows. Let the distance in R^p between the points

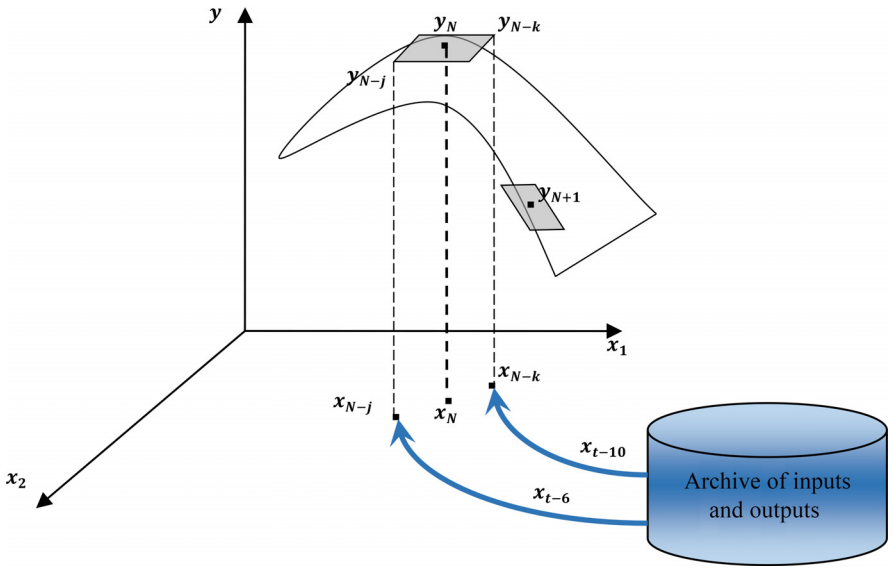


Fig. 1. The approximating hypersurface of the space of input vectors and one-dimensional output.

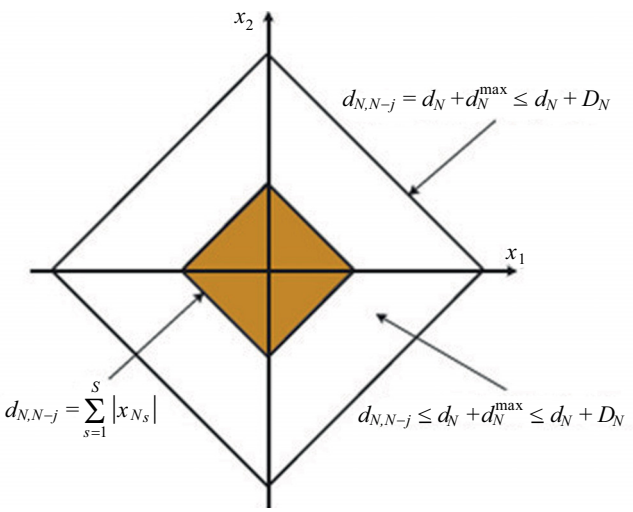


Fig. 2. The area of input vectors acceptably close to the input vector.

of the input space of dimension P be like:

$$d_{t,t-j} = \sum_{p=1}^{P} |x_{tp} - x_{t-j,p}|, \quad j = 1, \ldots, s, \tag{4}$$

where $s < t$, x_{tp} are the components of input vector on current time moment t. Suppose that for the current input vector x_t:

$$\sum_{p=1}^{P} |x_{tp}| = d_t. \tag{5}$$

To construct the approximating hypersurface x_t, let us choose such vectors x_{t-j}, $j = 1, \ldots, s$, from the archive of historical data, that for a given D_t the following condition is satisfied (Fig. 2):

$$d_{t,t-j} \leqslant d_t + \sum_{p=1}^{P} |x_{t-j,p}| \leqslant d_t + D_t, \quad j = 1, \ldots, s. \tag{6}$$

The preliminary value of D_t is determined on the basis of the knowledge of the process. If the chosen domain does not contain enough inputs for the application of LSM, i.e., the corresponding system of linear algebraic equations has no solution, then the given point criterion can be relaxed by increasing the threshold value D_t.

In order to increase the speed of the identification algorithm (both at the stage of training and at the subsequent operation of the object), one of the methods of intelligent data analysis, that is, clustering (dynamic classification, automatic grouping of data, "learning without a teacher"), is used. There are a lot of clustering methods, namely hierarchical algorithms, k-means algorithm, minimum covering tree algorithm, nearest neighbor method and others. All of them determine to which region, into which this space is divided, point in a multidimensional space belongs to.

As a result, at each time instant each investigated point in the multidimensional space can be assigned to some group and gain special cluster label. In the associative search problem for selection of input vectors close to the current one, the cluster label is defined according to the criterion of associative selection of input vectors from the archive. To build soft sensors, vectors are selected within the corresponding cluster.

For a dynamic linear model, the following algorithm is used to determine the unknown coefficients. The model is formed as:

$$y_N = \sum_{i=1}^{Q} a_i \hat{x}_i, \tag{7}$$

where $\hat{x}_i = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_r)$, $r = m + nS$, $\hat{x}$ is extended input vector for which:

$$\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m\} = \{y_{N-1}, y_{N-2}, \ldots, y_{N-m}\}; \tag{8}$$

$$\{\hat{x}_{m+1}, \hat{x}_{m+2}, \ldots, \hat{x}_{m+nS}\} = \{y_{N-1,1}, y_{N-1,2}, \ldots, y_{N-1,S}, \ldots, y_{N-n,S}\}; \tag{9}$$

α is an extended vector of input coefficients:

$$\{\alpha_1, \alpha_2, \ldots, \alpha_m\} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}; \tag{10}$$

$$\{\alpha_{m+1}, \alpha_{m+2}, \ldots, \alpha_{m+nS}\} = \left\{b_{1,1}, b_{1,2}, \ldots, b_{1,S}, \ldots, b_{n,S}\right\} \tag{11}$$

To build the model (7), the input vectors, which are close to the current one according to chosen criterion, are selected from the process data archive. After vector selection, a matrix of extended input vectors is compiled:

$$\hat{X} = \begin{pmatrix} \hat{x}_1^1 & \cdots & \hat{x}_r^1 \\ \vdots & \ddots & \vdots \\ \hat{x}_1^P & \cdots & \hat{x}_r^P \end{pmatrix}; \quad P \gg r. \tag{12}$$

To find the coefficients α_i you need to solve a system of linear equations:

$$\hat{X}\alpha = \hat{y}, \tag{13}$$

where $\hat{y}$ is system output at next time step for the selected extended vectors of process inputs.

When solving the system of linear Eqs. (13), assuming that $\operatorname{rank} \hat{X} = r$, LSM can be applied to find the estimate $\hat{\alpha}$:

$$\left(\hat{y} - \hat{X}\hat{\alpha}\right)^T \left(\hat{y} - \hat{X}\hat{\alpha}\right) = \min_{\alpha} \left(\hat{y} - \hat{X}\alpha\right)^T \left(\hat{y} - \hat{X}\alpha\right). \tag{14}$$

Assuming that $\hat{X}$ it is a matrix of full rank:

$$\hat{\alpha} = (\hat{X}^T \hat{X})^{-1} \hat{X}^T \hat{y}. \tag{15}$$

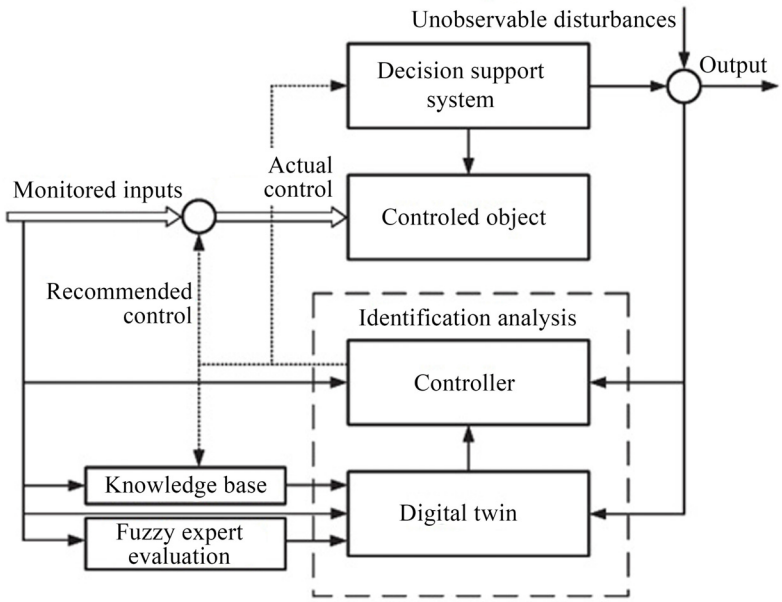


Fig. 3. Regulation scheme with an identifier in the feedback circuit.

$\hat{\alpha}$ is an LSM-estimate and, according to the Gauss–Markov theorem, has minimal dispersion in the class of unbiased linear estimates of the parameter α.

For dynamic models, we have the case of poor matrix conditioning due to statistical dependence of the components of the extended vector of inputs, so the estimation (15) may be inadequate. In this case we propose to use the following Moore–Penrose procedure. It is proposed to search estimation of $\hat{\alpha}_0$ such that:

$$\hat{\alpha}_0^T \hat{\alpha}_0 = \min \hat{\alpha}^T \hat{\alpha}. \tag{16}$$

According to the Moore–Penrose theorem, the estimation $\hat{\alpha}$ minimizes the left part of (14) if and only if it is represented in the form:

$$\hat{\alpha} = \hat{X}^+ \hat{y} + (1 - \hat{X}^+ \hat{X})p, \tag{17}$$

where p is some r-dimensional vector.

Asymptotically the normal estimation $\hat{\alpha}_0$ (16) is in the form:

$$\hat{\alpha}_0 = \hat{X}^+ \hat{y}. \tag{18}$$

The same procedure can be applied to form digital identification models in a closed loop control of the process.

The associative search algorithm makes it possible to obtain models, which for each time step are described by synchronized sets of values of inputs, outputs, control actions, coefficients, i.e., digital models, which are formed by the identifier in the feedback circuit of the MPC. In this case, the identifier in the feedback circuit of the control system is a digital twin (Fig. 3), because it forms a digital predictive model based on the current and statistical data of the process.

For non-stationary processes, as we know from [11], the associative search method also offers a constructive solution to the identification problem, namely, wavelet transform. This approach has demonstrated efficiency both for nonstationary input signal and in the case of impossibility of modeling internal dynamics of the control object. To apply the associative search algorithm for the

purpose of predicting the dynamics of nonstationary processes, it is necessary to select from the technological archive the vectors that are close to the current one according to the criterion formed for the coefficients of the multiple-scale wavelet decomposition.

5. CONSTRUCTION OF INTELLIGENT SOFT SENSOR BASED ON ASSOCIATIVE MODELS

We propose the following method of developing a software and algorithmic complex for the formation of real-time models of associative search — associative soft sensor.

It is necessary to take into account the specifics of laboratory work in production for process, when building the model. Often the laboratory analysis of products is performed not on one-time, but on averaged samples. Modeling should provide not only a sufficiently accurate description of the process class, but also adequately reflect the specifics of the production situation and the characteristics of a particular process. It is important to use all the available a priori information as much as possible. In particular, it is necessary to take into account all the constraints, determined by both technological regulations and expert opinions. The decision maker (operator or technologist) acts as an expert analyzing the situation.

Thus, the following main elements are highlighted in the development of soft sensors for process.

Description of the process: its features, allowing to formalize it with the help of certain mathematical models; accounting for certain data of the technical documentation, in particular, the technological regulations (or similar documents), which allows you to determine all the necessary restrictions:

— Base of technological regulations contains (in formalized form) the description of the following items: equipment, technological standards, rules of operations in various situations, detailed order of process, mode rules, mode parameters;

— Database of libraries for formalized representation of processes and their mathematical models.

The scheme of information flows of the investigated process, formed by the user of the system using the interactive interface, which allows to formalize the description of the simulated process in the form of differential or finite-difference equations.

Process inductive knowledge base must contain an archive of "production experience" of a particular process: from time-synchronized inputs and their corresponding outputs to the archives of configured models and archives of formalized situations ("coded" features and characteristics of the current state). The main elements of the system for formation and storage of knowledge, interpreted as patterns characterizing the process, are the following:

— Database of the functioning process;

— Data of technological equipment — actual values at specific time instants of technological parameters: consumption, pressures, temperatures, etc.; possible deviations from standard situations (set of patterns), additional limitations;

— Base of constructed point process models (archive of constructed models): sets of values of inputs and controls, as well as their corresponding outputs (finished products, by-products, waste) according to monitoring data;

— Assessments of results and recommendations for management (including formalized ones), that is, evaluations obtained by associative identification algorithms, as well as formalized values of experts' assessments (e.g., by means of fuzzy models).

The identifier in the feedback circuit of the automatic / automated process control system (digital twin) generates a digital model at each moment of time. Various elements of the formalized description of both the process itself and its current state come to the input of the identifier.

6. BUILDING AN INTELLIGENT SOFT SENSOR FOR A CONCENTRATOR OF MINING PRODUCTION

An enrichment plant is a mining facility for the primary processing of solid minerals in order to obtain technically valuable products suitable for industrial use. By means of various technologies (flotation, magnetic separation and others), a concentrate is obtained from the mined ore, in which the content of the useful component is much higher than in the original raw material. Concentrators process enrich ores of ferrous and nonferrous metals, nonmetallic minerals, and coal.

Below we compare results of laboratory control of product quality (iron concentration) with the quality metrics generated by soft sensors, developed on the basis of multiple linear regression and regression with associative search. The performance quality criteria of soft sensors are the mean absolute error (MAE) of the model and Pearson's correlation coefficient.

Values of laboratory analysis and soft sensors predictions of different types on iron concentration are presented on Figs. 4 and 5. Table shows the corresponding performance quality criteria of two types of soft sensors.

The formula for the linear regression model of the iron concentration:

$$F = 76.34 + 0.00545D, \tag{19}$$

where F — iron content in the sectional concentrate, D — drain density in hydrocyclones.

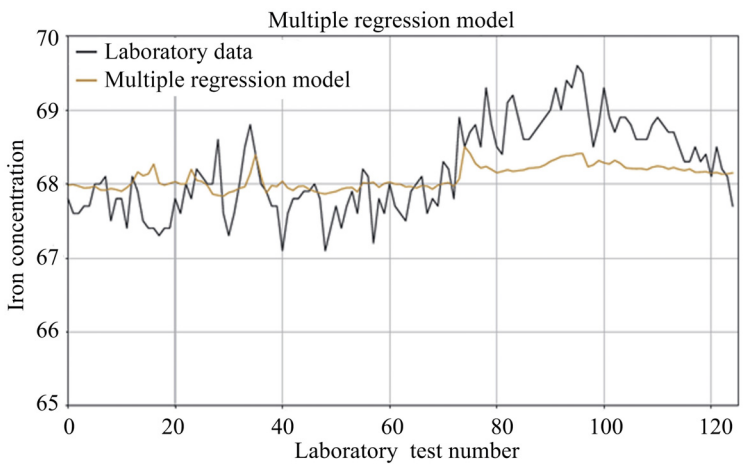


Fig. 4. Comparison of soft sensors values (multiple linear regression) with laboratory control data.

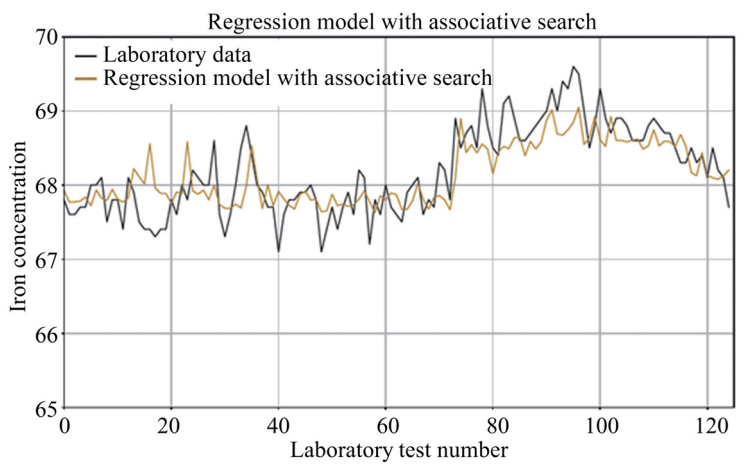


Fig. 5. Comparison of soft sensors values (regression with associative search) with laboratory control data.

Summary table of evaluation parameters

Model quality parameter	Linear regression	Regression with associative search
Mean absolute error (MAE)	0.416	0.282
Mean square error (MSE)	0.285	0.129
Pearson's correlation coefficient	0.479	0.803

Table shows that MAE and MSE for soft sensor based on associative model are smaller, and Pearson correlation coefficient is larger in comparison with soft sensor based on regression model. Thus, the regression model with associative search gives a more accurate prediction than the linear regression.

7. CONCLUSION

In this work the method of building soft sensors using intelligent identification algorithm, which forms real-time models of process by means of intelligent data analysis and machine learning is proposed. The new algorithm of soft sensors based on regression with associative search in inductive knowledge base is presented. Numerical simulation of the proposed methods for the process of a concentrator of mining production was conducted. The results demonstrate the higher accuracy and efficiency of regression model with associative search over classical regression models.

FUNDING

The reported study was funded: by the Russian Science Foundation, project no. 19-19-00673, by RFBR, according to the research project by RFBR and NSFC, project no. 21-57-53005.

REFERENCES

1. Bakhtadze, N.N., Virtual Analyzers: Identification Approach, *Autom. Remote Control*, 2004, vol. 65, no. 11, pp. 1691–1709.
2. Lototsky, V., Chadeev, V., Maksimov, E., and Bakhtadze, N., Prospects of Application of Virtual Analyzers in Production Control Systems, *Automation in Industry*, 2004, no. 5, pp. 23–29.
3. Vapnik, V., *Vosstanovlenie zavisimostei po empiricheskim dannym* (Reconstructing Dependencies from Empirical Data), Moskow: Nauka, 1979.
4. Bakhtadze, N., Kulba, V., Lototsky, V., and Maximov, E., Identification-based Approach to Soft Sensors Design, *Proceedings of IFAC Workshop of Intelligent Manufacturing Systems*, 2007, vol. 40, no. 3, pp. 86–92.
5. Bakhtadze, N., Sakrutina, E., and Pyatetsky, V., Predicting Oil Product Properties with Intelligent Soft Sensors, *IFACPapersOnLine*, 2017, vol. 50, no. 1, pp. 14632–14637.
6. Bakhtadze, N., Sakrutina, E., Pavlov, B., Lototsky, V., and Zaikin, O., Knowledge-based Prediction in Process Control Systems under Limited Measurement Data, *Procedia Computer Science J.*, 2017, vol. 112, pp. 1225–1237.
7. Bakhtadze, N., Chereshko, A., Elpashev, D., Suleykin, A., and Purtov, A., Predictive Associative Models of Processes and Situations, *IFACPapersOnLine*, 2022, vol. 55, no. 2, pp. 19–24.
8. Chereshko, A. and Titkina, M., Application of Associative Search Algorithms in Control Systems with a Predictive Model, *Avtomatizatsiya v Promyshlennosti*, 2022, no. 6, pp. 58–62.
9. Patel, V. and Ramoni, M., Cognitive Models of Directional Inference in Expert Medical Reasoning, in *Expertise in Context. Human and Machine*, MIT Press, 1997, pp. 67–99.
10. Razumkov, M., Verbal Analysis Methods: Research and Comparison, *Fundamental'nye Issledovaniya*, 2016, no. 10-3, pp. 642–646.
11. Bakhtadze, N., Lototsky, V., Vlasov, S., abd Sakrutina, E., Associative Search and Wavelet Analysis Techniques in System Identification, *Proceedings of the 16th IFAC Symposium on System Identification*, 2012, vol. 45, no. 16, pp. 1227–1232.