Volume 84, Number 12
December 2023

ISSN 0005-1179
CODEN: AURCAT

A JOURNAL OF * RUSSIAN ACADEMY OF SCIENCES *

AUTOMATION AND REMOTE CONTROL

Automation and Remote Control

Editor-in-Chief
Andrey A. Galyaev

Vol. 84, No. 12, December 2023

http://ait.mtas.ru

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. 12, 2023

Topical Issue

Nonlinear Systems

Robust, Adaptive, and Network Control

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1399–1400.

15th International Conference on Management on Large-Scale System Development. Opening Remarks of the Program Committee[1]

This special issue presents selected papers from the 15th International Conference on Management of Large-Scale System Development (MLSD'2022), held on September 26–28, 2022.

MLSD is an annual event organized by the Trapeznikov Institute of Control Problems, the Russian Academy of Sciences, starting from 2007. The conference program is intended for original research on the theory and practice of computer control to manage the development of production, transport, energy, financial, and social processes. Every year MLSD gathers over three hundred participants from research institutes, universities, government, and commercial organizations.

Traditionally, several high-impact conference papers are selected and placed as full-text articles in special issues of *Automation and Remote Control*. This issue of the journal includes eight best research papers presented at MLSD'2022.

Controlled thermonuclear fusion for industrial purposes is of paramount importance for the national economy. This problem is considered by Yu.V. Mitrishkin, S.L. Ivanova, and K.S. Mukhtarov in the paper "Adaptive Control Algorithm for Unstable Vertical Plasma Position in Tokamak." The authors develop and model an adaptive control algorithm for unstable vertical plasma positioning in a vertically elongated tokamak. The topic is interesting and important: at each step of automatic-mode operation, the system determines the parameters of the plant (identification) and designs a new feedback controller based on them. This system belongs to the class of robust-adaptive control systems. The parameters of the feedback controller are calculated using a given placement of the poles of the closed loop control system in the left half-plane of the complex plane. A robust system synthesized using Quantitative Feedback Theory (QFT) is used as an initial model of the control system. Note that the system was simulated on a real-time digital test bed; see https://www.ipu.ru/plasma/about.

An important area of research in various industries (energy, mechanical engineering, aviation, aerospace, and robotics) is the state monitoring of controlled objects and the controlled damping of dangerous oscillations. I.B. Yadykin and I.A. Galyaev significantly contribute to the solution of this problem in the paper "Structural Spectral Methods for Solving Continuous Lyapunov Equations." The authors develop spectral and singular decompositions for the inverse gramians of controllability and observability. As a result, invariant decompositions of energy functionals are obtained, and new stability criteria are formulated for linear systems with nonlinear mode interaction effects.

An urgent problem in developing new effective drugs and creating artificial proteins is predicting the properties of protein molecules based on their amino acid composition data. At present, molecular dynamic modeling is used to predict the properties of proteins and, in particular, their stability in the process of conformational changes. This method requires high computational and time costs. An effective approach to reduce the costs is to assess how the arrangement of amino acid residues in a protein affects its stability. In the paper "Probabilistic Assessment of a Pentapeptide Composition Influence on Its Stability" (A.I. Mikhalskii, J.A. Novoseltseva, A.A. Anashkina, and A.N. Nekrasov), this problem is solved using a cooperative game theory method. The authors

[1] The papers on pp. 1399–1467 are from the thematic issue.

calculate the Shapley value to estimate the probability of a positive or negative influence on the stability of a protein or the absence of a particular amino acid in its primary structure. The paper presents the practical implementation of the method to analyze the stability of short proteins consisting of five amino acids (pentapeptides).

The contemporary theory of managing the development of large-scale systems requires new models and methods for analyzing the attainability of goals. At present, this check is insufficiently formalized and performed mainly using the intuition and experience of decision-makers. A.D. Tsvirkun, A.F. Rezchikov, V.A. Kushnikov, O.I. Dranko, A.S. Bogomolov, and A.D. Selutin propose one approach to solving this problem; see their paper "Models and Methods for Checking the Attainability of Goals and Feasibility of Plans in Large-Scale Systems Using the Example of Goals and Plans for Elimination of the Consequences of Flood." The authors describe an algorithm for analyzing the attainability of goals and the feasibility of action plans implemented in the management of large-scale systems and consider flood management goals and plans as one example. The check is carried out in four stages; the first and second stages involve relational algebra and production models; the third stage, Markov process models and Kolmogorov–Chapman equations; the fourth stage, the system-dynamic approach and regression equations. Also, they form an algorithm for analyzing the attainability of goals and plans implemented during the development of such systems. An example is provided to illustrate the main stages of checking the attainability of goals and the feasibility of action plans.

The paper "Optimization of Group Incentive Schemes" (V.N. Burkov, I.V. Burkova, and A.R. Kashenkov) is devoted to the problem of designing a group incentive scheme to compensate for the costs of reducing the duration of project works. The theory of incentives generally considers two types of such schemes, namely, individual (a particular incentive scheme for each work) and unified (the same incentive scheme for all works). The group incentive scheme occupies an intermediate position: all works are partitioned into groups, and a particular incentive scheme is selected for each group. The problem is to partition the set of works into groups by minimizing the total incentive fund. This scheme largely offsets the disadvantages of individual and unified incentive schemes. The authors propose algorithms for solving the problem; they are based on determining shortest paths in a graph.

In the paper "Comparison of Distribution Procedures in Blended Finance," A.V. Shchepkin analyzes the following situation: the contractors of a megaproject apply to the Principal (an organization interested in this project) for funds. The Principal distributes the megaproject budget among the contractors only if they allocate their internal funds to project implementation. When distributing available funds, the Principal considers the requests for funding and the internal funds allocated by the contractors to their projects. This paper provides opportunities to improve business results.

Chair of the Program Committee
of MLSD'2022 Conference
Academician of RAS S.N. Vassilyev

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1401–1410.

THEMATIC ISSUE

Adaptive Control Algorithm for Unstable Vertical Plasma Position in Tokamak

Yu. V. Mitrishkin[*,**,a], **S. L. Ivanova**[**,b], **and K. S. Mukhtarov**[**,c]

[*] *Lomonosov Moscow State University, Moscow, Russia*
[**] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a] *yvm@mail.ru,* [b] *ivanovasvetlanamsu@gmail.com,* [c] *kirill.muhtarov@mail.ru*

Received May 31, 2023
Revised September 19, 2023
Accepted September 30, 2023

Abstract—The problem considered includes the development and modeling of an adaptive control algorithm for unstable vertical plasma positioning in a vertically elongated tokamak. At each iteration, a new PID controller is automatically synthesized for the evolving plasma model identified using the least squares method. The parameters of the feedback controller were computed based on the desired placement of the poles of the closed-loop control system in the left half-plane of the complex plane. The initial control system model utilized was a robust system synthesized using Quantitative Feedback Theory (QFT). The system was simulated on a real-time digital test bed (https://www.ipu.ru/plasma/about).

Keywords: tokamak, plasma, vertical plasma instability, QFT method, observation-based identification, adaptation, automatic synthesis, real-time digital test bed

DOI: 10.25728/arcRAS.2024.51.14.001

1. INTRODUCTION

In a vertically elongated tokamak, plasma exhibits vertical instability, necessitating the synthesis and application of feedback control systems for managing the vertical position of the plasma. This is a crucial problem in the field of plasma control in tokamaks.

The physics of vertically elongating plasma in a tokamak entails a process that significantly increases plasma pressure under the same toroidal magnetic field. However, this vertical elongation of the plasma induces its vertical instability.

This is explained by the creation of a radial magnetic field B_R, directed towards the central axis in the upper half-plane of the vertical cross-section of the tokamak and outward in the lower half-plane, resulting in the elongation of the plasma in the vertical direction (see Fig. 1).

As a result, the magnetic field lines of the total magnetic field B are convex towards the central axis Z of the tokamak. The Ampere force

$$F = [I \times B] \tag{1}$$

is directed upward in the upper half-plane and downward in the lower half-plane. While the current distribution and magnetic field are fully symmetric with respect to the central axis, the total Ampére force is zero. However, if a disturbance occurs, such as plasma displacement upward from the central axis, there will be a redistribution of currents and fields, resulting in a net force directed upward. This imbalance causes the plasma to move upward, as the resultant force is directed upward [1].

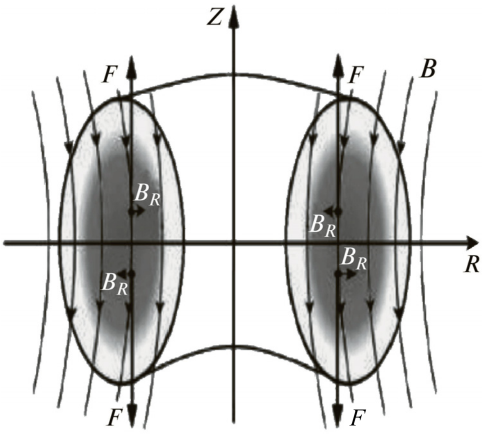


Fig. 1. Illustration of the instability of vertically elongated plasma in the tokamak.

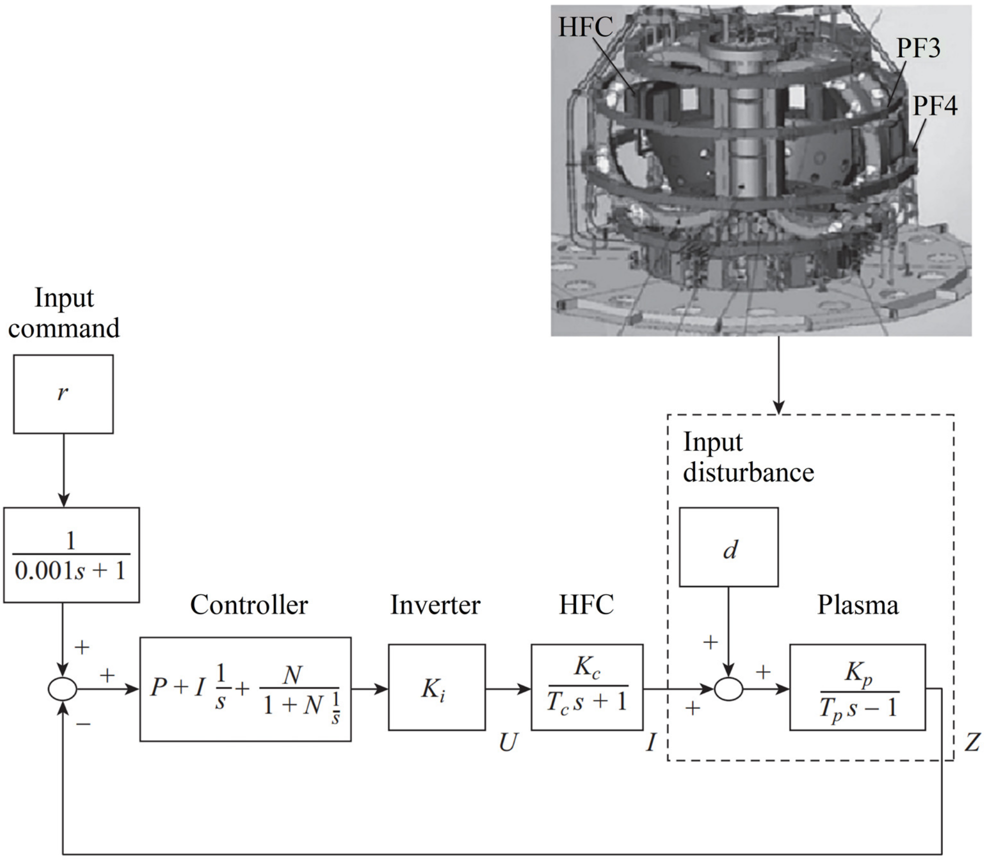


Fig. 2. Block diagram of the control system for the quantity Z without automatic tuning (PID controller with constant parameters).

The problem of controlling the vertical position of the plasma is addressed using the example of the T-15MD tokamak [2]. To suppress vertical plasma instability, the T-15MD tokamak design incorporates the Horizontal Field Coil (HFC) (see Fig. 2) [1]. The HFC is situated between the vacuum chamber and the toroidal field coil. In the design of the T-15MD tokamak, the HFC has been relocated from its position between the PF coils to the location depicted in Fig. 2. This relocation was prompted by the internal instability caused by the initial positioning of the HFC in the control system for vertical plasma position feedback [3, 4]. In the feedback control system, the HFC, in the event of a disturbance in the plasma column, generates magnetic field distributions such that the net Ampere force acting on the plasma becomes zero (compensated), thus stabilizing the vertical position of the plasma.

2. CONTROL PLANT MODEL

The plant model for the T-15MD tokamak includes a major plasma radius $R_0 = 1.48$ m, a minor radius $a = 0.67$ m, elongation $k = 1.7$–1.9, triangularity $\delta = 0.3$–0.4, plasma current $I_p = 2$ MA, pulse duration of 1 s, and toroidal magnetic field on the plasma axis up to $B = 2$ T [2]. When designing the plasma vertical position control system in the T-15MD tokamak, the plasma model (2) (the model justification history is provided in [5]) and the linear model of HFC (3) in state space were utilized:

$$T_p \frac{dZ}{dt} - Z = K_p(I + d), \tag{2}$$

$$L\frac{dI}{dt} + RI = U. \tag{3}$$

To simplify the plant model for subsequent adaptive control problem solving, the current inverter model from [6] was adopted as the actuator, which in the first approximation is modeled by a constant gain coefficient

Then the transfer function of the plant model consists of the sequential connection of transfer functions of the current inverter model K_i, the HFC model $\frac{K_c}{T_c s+1}$, the plasma model $\frac{K_p}{T_p s-1}$ with a disturbance input $d < 1$ kA (see Fig. 2) [1]. When designing a robust controller, all coefficients in this model have uncertainties. Here in (2), (3) U, I are the voltage and current of the HFC, K_p, T_p, K_a, T_a are the gain coefficients and time constants of the plasma model and the multi-phase thyristor rectifier model respectively, Z represents the vertical displacement of the plasma center.

The inductance L and the active resistance R of the HFC were calculated to be $L = 0.0042$ H, $R = 0.09$ ohm based on the data from JSC D.V. Efremov Institute of Electrophysical Apparatus (NIIEFA) [1]. Hence, the gain coefficient and time constant for the HFC model are respectively $K_c = \frac{1}{R} = 11.11\ \text{ohm}^{-1}$ and $T_c = \frac{L}{R} = 46.7$ ms. The nonlinear plasma physics code DINA, presented by employees of the Joint Stock Company “State Scientific Center of the Russian Federation Trinity Institute of Innovative and Thermonuclear Research” (JSC “SSC RF TRINITY”) (Troitsk), as referenced in [7], was identified in [8] with estimates of the time constant $T_p = 20.8$ ms and the gain coefficient $K_p = 1.78$ cm/kA for the linearized DINA-L model at the selected point in the parameter space of the T-15MD tokamak.

For the initial control system with the adaptation algorithm, a robust control system was utilized, synthesized using the Quantitative Feedback Theory (QFT) [9].

3. THE SYNTHESIS OF THE ROBUST CONTROL SYSTEM OF PLASMA VERTICAL POSITION Z USING THE QFT METHOD AND TESTING IT ON A REAL-TIME DIGITAL TEST BED

The constant magnitude and constant phase lines of the closed-loop control system in the amplitude—phase coordinates are plotted on the Nichols chart using the QFT theory (see Fig. 3a). These characteristics are referred to as QFT—boundaries and are calculated for different system parameters, thus encapsulating all the information about the uncertain model (see Fig. 3a).

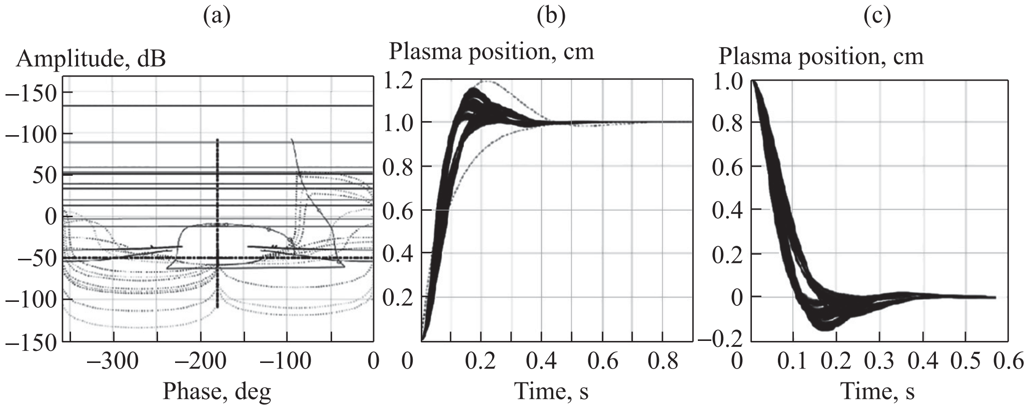
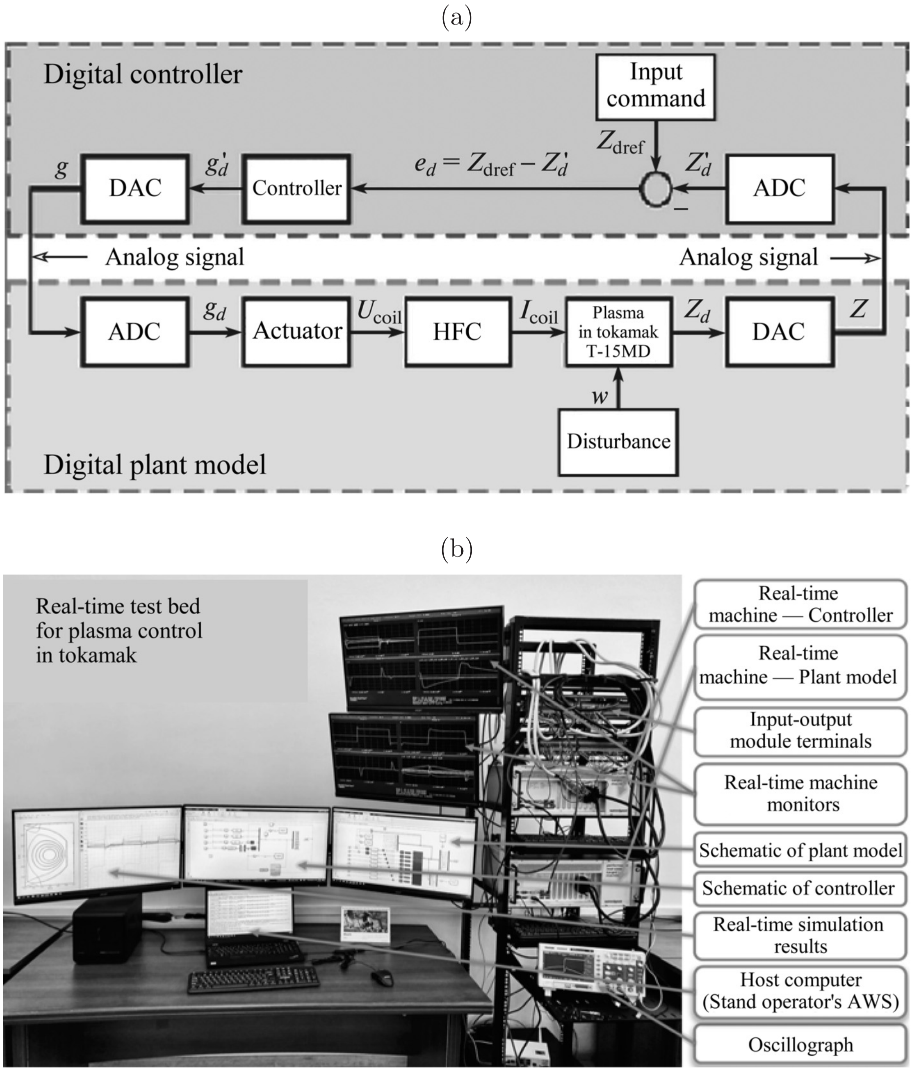
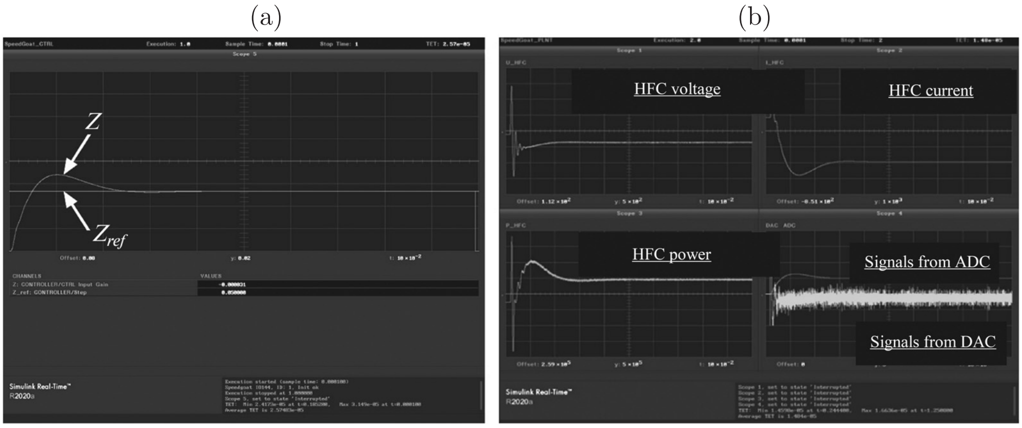


Fig. 3. (a) — Open—loop frequency response and Nichols chart boundaries, (b) — transfer functions of the feedback system for different parameters of the plant model when a setpoint input is applied, (c) — system transient responses when an external disturbance is applied.


(a)

(b)

Fig. 4. (a) — structural diagram of the control system on the real-time digital platform in discrete form with ADC and DAC; (b) — real—time digital platform for simulating control systems in tokamak plasma.



SpeedGoat_CTRL Execution: 1.0 Sample Time: 0.0001 Stop Time: 1 TET: 2.57e-05

CHANNELS	VALUES
Z: CONTROLLER/CTRL Input Gain	-0.000031
Z_ref: CONTROLLER/Step	0.050000

```
Execution started (sample time: 0.000100)
Speedgoat IO144, ID: 1, Init ok
Execution stopped at 1.000000
Scope 5, set to state 'Interrupted'
TET:  Min 2.4173e-05 at t=0.185200,   Max 3.149e-05 at t=0.000100
Average TET is 2.57483e-05
```

SpeedGoat_PLNT Execution: 2.0 Sample Time: 0.0001 Stop Time: 2 TET: 1.48e-05

```
Scope 1, set to state 'Interrupted'
Scope 2, set to state 'Interrupted'
Scope 3, set to state 'Interrupted'
Scope 4, set to state 'Interrupted'
TET:  Min 1.4598e-05 at t=0.244400,   Max 1.6636e-05 at t=1.250000
Average TET is 1.484e-05
```

Fig. 5. *a* — Step response of the control system to a 5 cm step input in real—time; *b* — signals of voltage, current, and power in the HFC, as well as signals from the ADC and DAC, to a 5 cm step input in real—time.

Using the specified boundaries and the Nichols chart (see Fig. 3a), a robust PID controller was synthesized:

$$C(s) = P + \frac{I}{s} + D\frac{N}{1 + \frac{N}{s}}$$

with parameters $P = 39$, $I = 563$, $D = 1.38$, $N = 12\,291$. The control system with this controller has no steady-state error, a settling time of about 300 ms (see Fig. 3b), and suppresses external disturbances within 300 ms (see Fig. 3c).

The obtained control system was discretized using the ZOH (zero-order hold) method with a sample time of 100 μs and tested on the Speedgoat Performance real-time target machine on the SimulinkRT operating system [10–12]. The real-time target machines, connected in a feedback loop “plant model—controller”, facilitate the fastest transition from control system modeling in the MATLAB/Simulink environment to real-time testing on the digital test bed (see Fig. 4a). The digital controller and digital plant model on the test bed exchange analog signals with each other using DAC and ADC (see Fig. 4b).

The real-time system’s performance is determined by the task execution time (TET). It consists of the time required for calculating the models of tokamak components and control algorithms, as well as the time for polling input—output modules. For the developed control system with the robust controller, the TET was approximately 14.6 μs. For nominal real-time system operation, the TET should not exceed the sample time in the numerical algorithm solving difference equations (in this case, 100 μs). The graphs depicting the plasma position and the voltage, current, and power changes in the HFC are shown in Figs. 5a and 5b.

4. ADAPTIVE PLASMA CONTROL DURING A SINGLE DISCHARGE

The problem involves identifying the changing plasma model and subsequently tuning the controller within one plasma discharge, which lasts approximately 1 s.

The plasma model with two time-varying parameters $K(t)$ and $T(t)$ was adopted as the control plant model:

$$T(t)\frac{dZ(t)}{dt} - Z(t) = K(t)I(t), \tag{4}$$

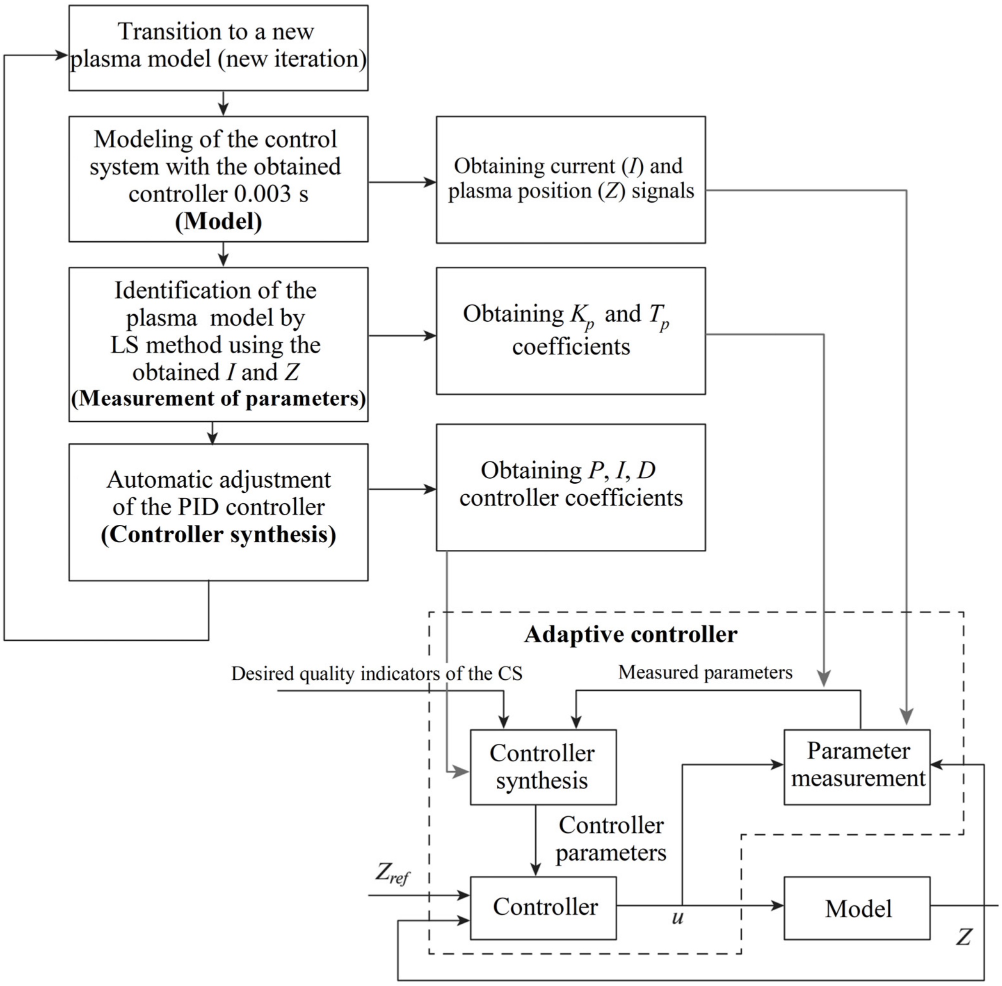

Transition to a new plasma model (new iteration)
Modeling of the control system with the obtained controller 0.003 s (Model)
Obtaining current (I) and plasma position (Z) signals
Identification of the plasma model by LS method using the obtained I and Z (Measurement of parameters)
Obtaining K_p and T_p coefficients
Automatic adjustment of the PID controller (Controller synthesis)
Obtaining P, I, D controller coefficients
Adaptive controller
Desired quality indicators of the CS
Measured parameters
Controller synthesis
Parameter measurement
Controller parameters
Z_{ref}
Controller
u
Model
Z

Fig. 6. The system with an adaptive control algorithm for the vertical position of the plasma during a discharge.

connected in series with the linear model of the HFC with known constant parameters

$$L\frac{dI(t)}{dt} + RI(t) = U(t).$$

While simulating the evolution of the plasma model (4), the coefficients of the plasma model change linearly from the lower bound to the upper bound during the algorithm's operation — coefficient $K \in [1.78; 7.61]$ cm/kA, coefficient $T \in [0.0208; 0.093]$ s. Simultaneously, plasma model identification and synthesis of a new PID controller are performed. Figure 6 depicts the system with the adaptive control algorithm for vertical plasma position throughout the discharge.

The parameter identification problem for the plasma model was solved using linear regression and the method of least squares [13]. For thirty consecutive measurements at discrete points with a quantization step of the input and output signals $Z(k)$, $I(k)$, an estimate $\widehat{T}$ of the parameter T and an estimate $\widehat{K}$ of the parameter K are computed. These estimates are obtained by minimizing the following functional:

$$J_k = \sum_{k=1}^{30}\Bigl(T\frac{Z(k+1) - Z(k)}{\triangle\, t} - Z(k) - KI(k)\Bigr)^2. \qquad (5)$$

By taking partial derivatives with respect to the estimated parameters in the functional (5), we can derive formulas for their estimation:

$$\begin{aligned} J_k = {} & \widehat{K}^2 I(k)^2 + 2\widehat{K} I(k) Z(k) - 2\widehat{K} I(k) \widehat{T}\left(\frac{Z(k+1) - Z(k)}{\Delta t}\right) + Z(k)^2 \\ & - 2Z(k)\widehat{T}\left(\frac{Z(k+1) - Z(k)}{\Delta t}\right) + \widehat{T}^2\left(\frac{Z(k+1) - Z(k)}{\Delta t}\right)^2 \to \min, \end{aligned}$$

$$\frac{dJ_k}{d\widehat{K}} = 2\widehat{K} I(k)^2 + 2I(k)Z(k) - 2I(k)T\frac{Z(k+\widehat{1) - }Z(k)}{\Delta t} = 0, \tag{6}$$

$$\frac{dJ_k}{d\widehat{T}} = 2\widehat{T}\left(\frac{Z(k+1) - Z(k)}{\Delta t}\right)^2 - 2Z(k)\frac{Z(k+1) - Z(k)}{\Delta t} - 2I(k)T\frac{Z(k+\widehat{1) - }Z(k)}{\Delta t} = 0. \tag{7}$$

Transform equations (6), (7):

$$\widehat{T}\frac{Z(k+1) - Z(k)}{\Delta t} - Z(k) - \widehat{K} I(k)\frac{Z(k+1) - Z(k)}{\Delta t} = 0, \tag{8}$$

$$KI(k) + I(k)Z(k) - I(k)T\frac{Z(k+1) - Z(k)}{\Delta t} = 0. \tag{9}$$

Express the estimates for the coefficients K and T from (8) and (9):

$$\widehat{T} = \frac{Z(k)}{\frac{Z(k+1) - Z(k)}{\Delta t}},$$

$$\widehat{K} = \frac{\widehat{T}\frac{Z(k+1) - Z(k)}{\Delta t} - Z(k)}{I(k)}.$$

After measuring the signals I and Z and estimating the parameters T and K of the changing plasma model, it is necessary to synthesize the controller. To solve this problem, a PID controller was chosen, as described in [14], which automatically adjusts itself using the method of placing the characteristic polynomial roots in the left half—plane of the complex plane at each iteration of controller tuning (every 0.023 s). During the first iteration of the control system modeling, a PID controller synthesized using the QFT method was employed.

Transform the transfer function of the PID controller with a filter (10)

$$C(s) = K_c\left(1 + \frac{1}{\tau_I s} + \frac{\tau_D s}{\tau_f s + 1}\right) \tag{10}$$

to a common denominator and introduce the following notations:

$$C(s) = \frac{c_2 s^2 + c_1 s + c_0}{s(s + l_0)},$$

where

$$c_2 = \frac{K_c(\tau_I \tau_D + \tau_I \tau_f)}{\tau_I \tau_f}, \quad c_1 = \frac{K_c(\tau_I + \tau_f)}{\tau_I \tau_f}, \quad c_0 = \frac{K_c}{\tau_I \tau_f}, \quad l_0 = \frac{1}{\tau_f}.$$

For the PID controller, the unstable control plant model will take the form:

$$G(s) = \frac{K_p K_c K_a}{(T_p s - 1)(T_c s + 1)} = \frac{K}{T_p T_c s^2 + (T_p - T_c)\, s - 1}.$$

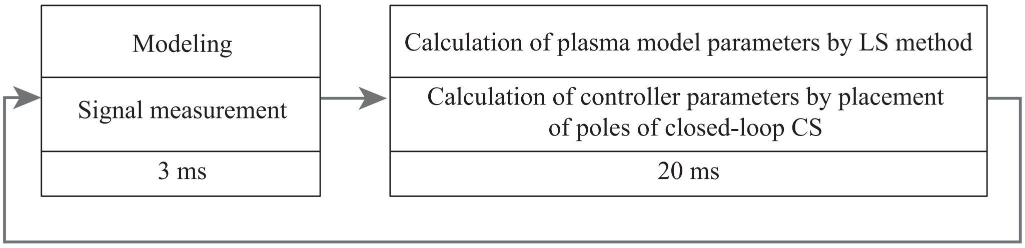

Modeling
Signal measurement
3 ms
Calculation of plasma model parameters by LS method
Calculation of controller parameters by placement of poles of closed-loop CS
20 ms

Fig. 7. The adaptive control algorithm for the unstable vertical position of the plasma.

The transfer function of the closed-loop control system is given by:

$$\frac{K(c_2s^2+c_1s+c_0)}{T_pT_cs^4+(T_p-T_c)\,s^3+(Kc_2+l_0T_pT_c-1)\,s^2+(l_0T_p-l_0T_c+KC_1)\,s+c_0K-l_0}.$$

Write down the characteristic equation and equate it to the polynomial with the given coefficients:

$$s^4+\frac{T_p-T_c+l_0T_pT_c}{T_pT_c}s^3+\left(\frac{l_0T_p-l_0T_c+Kc_1-1}{T_pT_c}\right)s^2+\frac{c_0K-l_0}{T_pT_c}s+\frac{c_0K}{T_pT_c}$$
$$=s^4+a_3s^3+a_2s^2+a_1s+a_0.$$

By comparing the coefficients of both sides of the polynomial, we obtain four linear equations:

$$\begin{cases}\dfrac{1}{T_c}-\dfrac{1}{T_p}+l_0=a_3,\\ Kc_2-1+(T_p-T_c)\,l_0=a_2,\\ \dfrac{K}{T_pT_c}c_1-\dfrac{l_0}{T_pT_c}=a_1,\\ \dfrac{K}{T_pT_c}c_0=a_0.\end{cases}\tag{11}$$

The parameters of the PID controller are found by solving the system of linear equations (11), which can be expressed as

$$\begin{bmatrix}l_0\\c_2\\c_1\\c_0\end{bmatrix}=\begin{bmatrix}1&0&0&0\\T_p-T_c&K&0&0\\\dfrac{-1}{T_pT_c}&0&\dfrac{K}{T_pT_c}&0\\0&0&0&\dfrac{K}{T_pT_c}\end{bmatrix}^{-1}\begin{bmatrix}a_3-\dfrac{T_p-T_c}{T_pT_c}\\a_2+1\\a_1\\a_0\end{bmatrix}.$$

Figure 7 illustrates the adaptive control algorithm for the unstable vertical position of the plasma, consisting of two stages: measuring and storage the input and output signals of the plasma model, i.e., I and Z over 3 ms with a 100 μs sample time, and computing the parameters of the plasma model and, based on them, the parameters of the PID controller over 0.02 s. Thus, in the discrete system, there are two steps: the overall system operation step of 100 μs and the step of identifying the parameters of the plant model and tuning the controller parameters, which is equal to 0.023 s. Therefore, within one discharge, which lasts approximately 1 second, it is possible to perform 43 iterations of controller tuning (see Fig. 7). The results of the adaptive control algorithm in the closed-loop system are presented in Fig. 8.

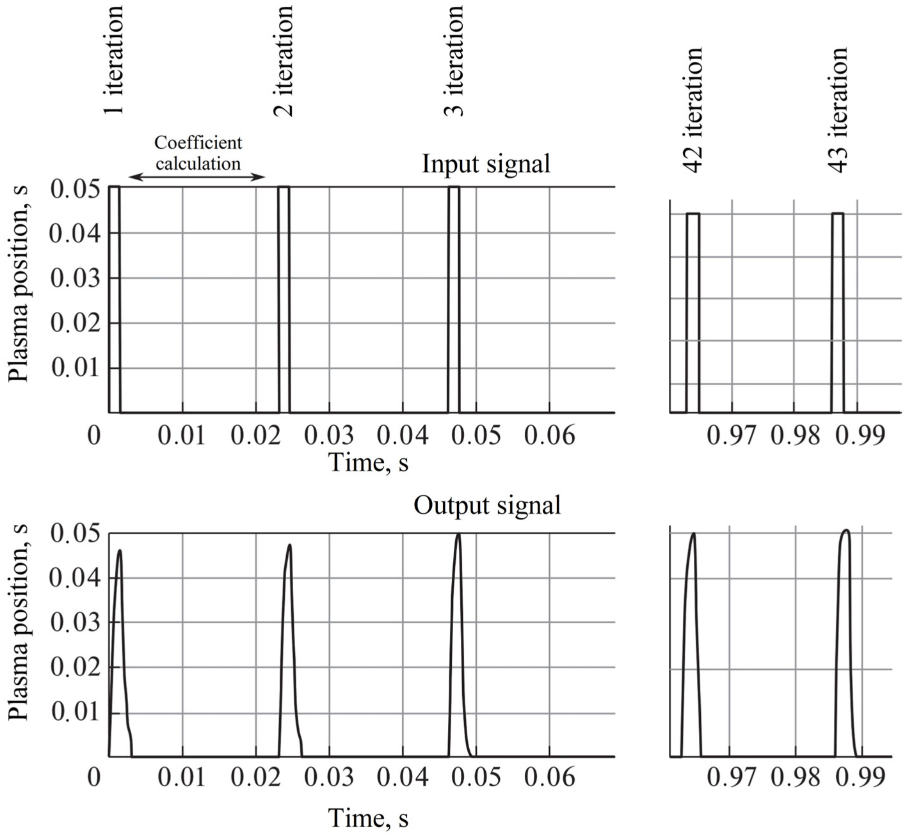


Fig. 8. Results of simulating the control system for the unstable vertical position of the plasma, performing 43 iterations of controller tuning under the changing plasma model.

5. CONCLUSION

During each iteration, the coefficients of the plasma model $T_p \in [0.0208;\ 0.093]$ s and $K_p \in [1.78;\ 7.61]$ cm/kA were linearly changed. The least squares method was used to estimate these coefficients, and the PID controller was adjusted using the root locus method to ensure stability of the closed-loop system in the left half-plane of the complex plane. The specified coefficients of the characteristic equation $a_0 = -0.0004$, $a_1 = 6e - 08$, $a_2 = -4e - 12$, $a_3 = 1e - 16$ were chosen for tuning the controller. The adaptation algorithm performs 43 controller adjustments within one second, which is sufficient for a real control plant like the T15-MD tokamak.

Currently, robust [15], adaptive [16], and robust-adaptive [17] control systems continue to evolve [18]. Robust-adaptive control systems with the application of neural networks [19] deserve the most attention and can also be applied to plasma control in tokamaks in the near future.

FUNDING

The work was supported by the Russian Science Foundation, project no. 21-79-20180.

REFERENCES

1. Mitrishkin, Y.V., Pavlova, E.A., Kuznetsov, E.A., and Gaydamaka, K.I., Continuous, Saturation, and Discontinuous Tokamak Plasma Vertical Position Control Systems, *Fusion Eng. Des.*, 2016, vol. 108, pp. 35–47.

2. Khvostenko, P.P., Anashkin, I.O., Bondarchuk, E.N., Inyutin, N.V., Krylov, V.A., Levin, I.V., Mineev, A.B., and Sokolov, M.M., Experimental Thermonuclear Facility Tokamak T-15MD, *VANT. Termoyad. Sint.*, 2019, vol. 42, no. 1, pp. 15–38.

3. Mitrishkin, Yu.V., Kartsev, N.M., and Zenkov, S.M., Stabilization of Unstable Vertical Position of Plasma in T-15 Tokamak. I, *Autom. Remote Control*, 2014, vol. 75, no. 2, pp. 281–293.

4. Mitrishkin, Yu.V., Kartsev, N.M., and Zenkov, S.M., Stabilization of Unstable Vertical Position of Plasma in T-15 tokamak. II, *Autom. Remote Control*, 2014, vol. 75, no. 9, pp. 1565–1576.

5. Mitrishkin, Y.V., Konkov, A.E., and Korenev, P.S., Comparative Study of Real-Time Control Systems for Vertical Plasma Position in a Tokamak with Different Power Sources for the Horizontal Field Coil, *VANT. Termoyad. Sint.*, 2022, vol. 45, no. 3, pp. 34–49.

6. Kuznetsov, E.A., Mitrishkin, Y.V., and Kartsev, N.M., Current Inverter as Auto-Oscillation Actuator in Applications for Plasma Position Control Systems in the Globus-M/M2 and T-11M Tokamaks, *Fusion Eng. Des.*, 2019, vol. 143, no. 3, pp. 247–258.

7. Khayrutdinov, R.R. and Lukash, V.E., Studies of Plasma Equilibrium and Transport in a Tokamak Fusion Device with the Inverse-Variable Technique, *J. Comput. Phys.*, 1993, vol. 109, no. 2, pp. 193–201.

8. Mitrishkin, Y.V., Kartsev, N.M., and Zenko, S.M., Plasma Vertical Position, Shape, and Current Control in T-15 Tokamak, in *Proceedings of the IFAC Conference on Manufacturing Modelling, Management and Control*, Saint Petersburg, Russia, 19–21 June 2013, pp. 1820–1825.

9. Garcia-Sanz, M., *Robust Control Engineering. Practical QFT solutions*, USA: CRC Press, 2017.

10. Mitrishkin, Y.V., Plasma Magnetic Control Systems in D-Shaped Tokamaks and Imitation Digital Computer Platform in Real Time for Controlling Plasma Current and Shape, *Adv. Syst. Sci. Appl.*, 2022, vol. 22, no. 1, pp. 1–14.

11. Mitrishkin, Y.V., Konkov, A.E., and Korenev, P.S., Digital Real-Time Modeling Stand for Plasma Control in tokamaks, *Proceedings of the XVI International Conference on Stability and Oscillations of Nonlinear Control Systems (Pyatnitsky Conference)*, Moscow, 2022, pp. 286–289.

12. Mitrishkin, Y.V., Method of Magnetic Plasma Control in Real Time in a Tokamak and Device for its Implementation, *Patent RF no. 2773508*, 2022.

13. Ljung, L., *System Identification: Theory for the User*, Englewood Cliffs: Prentice Hall, 1987. Translated under the title *Identifikatsiya sistem. Teoriya dlya pol'zovatelya*, Moscow: Nauka, 1991.

14. Wang, L., *PID Control System Design and Automatic Tuning using MATLAB/Simulink*, UK: Wiley, 2020.

15. Skogestad, S. and Postlethwaite, I., *Multivariable Feedback Control. Analysis and Design*, UK: Wiley, 2005.

16. Tyukin, I.Yu. and Terekhov, V.A., *Adaptatsiya v nelineinykh dinamicheskikh sistemakh* (Adaptation in Nonlinear Dynamic Systems), Moscow: LKI Publishing, 2008.

17. *Adaptive Robust Control Systems*, Anh Tuan Le, Ed., IntechOpen, March 2018. 362 p. https://doi.org/10.5772/intechopen.68813

18. Abdalla, T., Adaptive Data-Driven Control for Linear Time Varying Systems, *Machines*, 2021, vol. 9, no. 8, p. 167.

19. Yechiel, O. and Guterman, H., A Survey of Adaptive Control, *International Robotics & Automation Journal*, 2017, 3(2), pp. 290–292. https://doi.org/10.15406/iratj.2017.03.00053

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1411–1427.

THEMATIC ISSUE

Structural Spectral Methods for Solving Continuous Lyapunov Equations

I. B. Yadykin[*,a] **and I. A. Galyaev**[*,b]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a] *Jad@ipu.ru,* [b] *ivan.galyaev@yandex.ru*

Received May 31, 2023
Revised September 10, 2023
Accepted September 30, 2023

Abstract—For linear multivariable continuous stationary stable control systems with a simple spectrum, presented in the form of a canonical diagonal form, controllability and observability forms, a method was developed and analytical formulas for spectral decompositions of gramians in the form of various Xiao matrices were obtained. A method and algorithm for calculating generalized Xiao matrices in the form of the Hadamard product for multiply connected continuous linear systems with many inputs and many outputs have been developed. This allows us to calculate the elements of the corresponding controllability and observability gramians in the form of products of the corresponding elements of the multiplier matrices and a matrix that is the sum of all possible products of the numerator matrices of the matrix transfer function of the system. New results are obtained in the form of spectral and singular decompositions of the inverse gramians of controllability and observability. This makes it possible to obtain invariant decompositions of energy functionals and formulate new criteria for the stability of linear systems taking into account the nonlinear effects of mode interaction.

Keywords: : spectral decompositions of gramians, singular numbers, inverse gramian matrix, stability that takes into account the interaction of modes, Xiao matrices, Lyapunov equation

DOI: 10.25728/arcRAS.2024.59.12.001

1. INTRODUCTION

Monitoring the state of control objects and controlling the damping of dangerous vibrations are important areas of research in various fields of industry (energy, mechanical engineering, aviation and astronautics, robotics). New modeling technologies require the development of tools for approximating mathematical models of complex systems of various natures [1–3]. An important role is played by the methods of calculating the Lyapunov and Sylvester matrix equations and the study of the structural properties of solutions to these equations [4–11]. The fundamental properties of linear dynamic systems associated with solutions to these equations are controllability, observability and stability. Important results were obtained in the field of computing gramians for systems which models are presented in the canonical forms of controllability and observability. In [12], methods for calculating gramians based on the use of matrices of periodic structure were first proposed for linear systems specified by equations in the forms of controllability and observability. A new approach was developed in terms of use the properties of the impulse transition function and gramian matrices in the form of the zero-plaid structure of the controllability gramian in [13, 14]. In [15], the approach was developed to compute spectral decompositions of a more general class of linear time-invariant (LTI) multiple-input multiple-output (MIMO) systems. Using this approach, a method for optimal selection of locations for sensors and actuators on the graph of a distributed control system was developed in [16]. The paper shows that for a diagonalized system the controllability gramian can be represented as the Hadamard product of two positive semidefinite matrices. In [17], the problem of optimizing the capacity of an urban transport network was solved based on minimizing the trace of the gramian controllability matrix taking into account restrictions. Various problems related to the usage of controllability, observability and cross-gramians for calculating system invariants and energy stability indices can be found in [18, 19].

The goal of this work is to develop structural methods for solving matrix Lyapunov equations and obtain spectral and singular decompositions of controllability and observability gramians, based on reducing the equations of state of a linear stationary system to the following canonical forms: diagonal, controllability and observability.

2. FORMULATION OF THE PROBLEM

We consider a stable continuous MIMO LTI dynamic system of the form

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = 0, \quad y(t) = Cx(t), \tag{2.1}$$

where $x(t) \in R^n$, $u(t) \in R^m$, $y(t) \in R^m$. We will consider real matrices of corresponding sizes A, B, C. Let us assume that the system (2.1) is completely controllable and observable and all eigenvalues of matrix A are different. In this case, the implementation of (2.1) is minimal and there is the only one transfer function $W(s)$ in the form

$$W\left(s\right) = \sum_{i=1}^{n} M_i s^i N^{-1}\left(s\right),$$

where $N(s)$ is characteristic polynomial of matrix A, M_i is a matrix of the form

$$M_i = \sum_{i=0}^{n-1} A_i B.$$

Above, A_i denotes the "i"th Faddeev matrix in the decomposition of the resolvent of matrix A in the Faddeev–Le Verrier series [6, 7]. In accordance with [20], we write a general formula for calculating the controllability gramian from the pair spectrum of the system (2.1)

$$P^c = -\sum_{j=1}^{n}\sum_{\rho=1}^{n} \frac{1}{s_j + s_\rho} Res\left[\left(Is - A\right)^{-1}, s_j\right] BB^* Res\left[\left(Is - A^*\right)^{-1}, s_\rho\right]. \tag{2.2}$$

We consider a continuous dynamic MISO LTI system of the form

$$\begin{gathered}\dot{x}\left(t\right) = Ax\left(t\right) + b_\gamma u_\gamma\left(t\right), \quad x\left(0\right) = 0, \\ y\left(t\right) = cx\left(t\right),\end{gathered} \tag{2.3}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^1$, $u_\gamma\left(t\right) \in \mathbb{R}^m$, $\gamma = 1, \dots m$, b_γ is column of matrix B.

We consider the transformation of the equation (2.1) of a general system to equations of state in canonical forms: diagonal, controllability and observability.

If all eigenvalues s_r of matrix A are different, then the linear system can be reduced to diagonal form using a non-degenerate coordinate transformation

$$\begin{gathered}x_d = Tx, \quad \dot{x}_d = A_d x_d + B_d u, \quad y_d = C_d x_d, \\ A_d = TAT^{-1}, \quad B_d = TB, \quad C_d = CT^{-1}, \quad Q_d = TBB^{\mathrm{T}}T^{\mathrm{T}},\end{gathered}$$

or

$$\mathrm{A} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & s_n \end{bmatrix} \begin{bmatrix} \nu_1^* \\ \nu_2^* \\ \vdots \\ \nu_n^* \end{bmatrix} = T\Lambda T^{-1},$$

where matrix T is composed of right eigenvectors u_i, and matrix T^{-1} is composed of left eigenvectors ν_i^* corresponding to the eigenvalue s_i. The gramian of the diagonalized linear part is a solution to the Lyapunov equation, which is determined from the formula [15]

$$P_d^c = -\sum_{j=1}^{n}\sum_{\rho=1}^{n} \frac{1}{s_j + s_\rho} Res\left[(Is - A_d)^{-1}, s_j\right] B_d B_d^* Res\left[(Is - A_d)^{-1}, s_\rho\right].$$

The controllability gramian P_d^c is related to the gramian P^c by a relation of the form

$$P^c = T P_d^c T^{\mathrm{T}}.$$

From (2.2) results the following separable spectral decomposition of the gramian controllability of a system transformed into a diagonal canonical form [21]

$$P_d^c = \sum_{j=1}^{n}\sum_{\rho=1}^{n} \frac{-b_{j\rho}}{s_j + s_\rho} \mathbb{1}_{j\rho}, \quad b_{j\rho} = [B_d B_d^*]_{j\rho},$$

where the designation is introduced

$$\mathbb{1}_{j\rho} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1_{j\rho} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Further we consider the channel "γ" of the MISO LTI system in the canonical form of controllability [1, 21]

$$\begin{gathered} x(t) = \sum_{\gamma=1}^{m} R_{c\gamma}^F x_{c\gamma}(t). \\ \dot{x}_c(t) = A_c^F x_{c\gamma}(t) + b_\gamma^F u_\gamma(t), \quad x_c(0) = 0, \\ y_c^F(t) = c_\gamma^F x_c(t), \quad \gamma = 0, 1 \dots, m. \end{gathered} \tag{2.4}$$

$$A_c^F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-1} \end{bmatrix}, \quad b_\gamma^F = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix}^{\mathrm{T}},$$

$$a = \begin{bmatrix} -a_0 & -a_1 & \dots & -a_{n-2} & -a_{n-1} \end{bmatrix}, \quad c_\gamma^F = \begin{bmatrix} \xi_0 & \xi_1 & \dots & \xi_{n-2} & \xi_{n-1} \end{bmatrix}.$$

If we use a non-degenerate transformation of variables with the matrix R_c^F, we can consider the MISO LTI system in the canonical form of controllability. The vector B_γ of the MISO system has the form

$$B_\gamma = \begin{bmatrix} 0 & \dots & b_\gamma & \dots & 0 \end{bmatrix}^T.$$

The following relations are valid [14]:

$$\left(R^F_{c\gamma}\right)^{-1} A R^F_{c\gamma} = A^F_c,\ \left(R^F_{c\gamma}\right)^{-1} B_\gamma = b^F_\gamma,\ C R^F_{c\gamma} = c^F_\gamma,$$

$$P^c = \sum_{\gamma=1}^{m} R^{cF}_\gamma P^{cF}_\gamma (R^{cF}_\gamma)^T.$$

In relation to the systems (2.1) and (2.3) we will assume that various structural conditions of stability, controllability, observability and properties of the spectrum of the dynamics matrix are fulfilled. The following spectral decomposition of the controllability gramian was obtained in [15]:

$$P^{cF}_\gamma = \sum_{k=1}^{n}\sum_{\eta=0}^{n-1}\sum_{j=0}^{n-1} \frac{s_k^j(-s_k)^\eta}{\dot N\left(s_k\right) N\left(-s_k\right)} \mathbb{1}_{j+1\eta+1}.$$

Next, we consider the "γ" SIMO LTI channel of a linear system in the canonical form of observability [15]. In this case, the formulas are valid

$$x_o(t) = \sum_{\gamma=1}^{m} R^F_{o\gamma} x_{o\gamma}(t),$$

$$\dot x_{o\gamma}\left(t\right) = A^F_c x_{o\gamma}\left(t\right) + b^F_{o\gamma} u_\gamma\left(t\right), \quad x_o\left(0\right) = 0,$$

$$y^F_{o\gamma}\left(t\right) = c^F_{o\gamma} x_{o\gamma}\left(t\right), \quad \gamma = 0, 1 \ldots m.$$

$$A^F_o = \begin{bmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_{-1} \\ 0 & 1 & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & -a_{n-2} \\ 0 & 0 & \ldots & 1 & -a_{n-1} \end{bmatrix}, \quad b^F_{o\gamma} = \begin{bmatrix} \xi_0 & \xi_1 & \ldots & \xi_{n-2} & \xi_{n-1} \end{bmatrix}^{\mathrm{T}},$$

$$c^F_{o\gamma} = \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 \end{bmatrix}.$$

We obtain following expressions in accordance with the principle of duality [14]

$$P^F_{o\gamma} = \sum_{k=1}^{n}\sum_{\eta=0}^{n-1}\sum_{j=0}^{n-1} \frac{s_k^j(-s_k)^\eta}{\dot N\left(s_k\right) N\left(-s_k\right)} \mathbb{1}_{j+1\eta+1},$$

$$P^o = \sum_{\gamma=1}^{m} R^F_{o\gamma} P^{oF}_\gamma (R^F_{o\gamma})^T.$$

Definition 1. We call the Xiao matrix (Zero plaid structure) a matrix of the form [12, 13]

$$Y = \begin{bmatrix} y_1 & 0 & -y_2 & 0 & y_3 \\ 0 & y_2 & 0 & -y_3 & 0 \\ -y_2 & 0 & y_3 & 0 & \ldots \\ 0 & -y_3 & 0 & \ldots & 0 \\ y_3 & 0 & \ldots & 0 & y_n \end{bmatrix}.$$

The matrix elements are calculated using the formulas

$$y_{j\eta} = \begin{cases} 0, \text{ if } j+\eta = 2k+1, \quad k = 1, \ldots, n; \\ y_n = \dfrac{1}{2Y_{n,1}}, \\ y_{n-l} = \dfrac{-\sum\limits_{i=1}^{m-1} (-1)^i Y_{n-l,i+1} y_{n-l+i}}{Y_{n-l,1}}, \quad \text{if } j+\eta = 2k, \quad k = 1, \ldots, n, \quad l = \overline{1, n-1}, \end{cases}$$

where $Y_{i,j}$ is the element of the Routh table for the system located at the intersection of i row and j column.

3. MAIN RESULTS

3.1. Identities for One Class of Stable Polynomials Which Roots are Different over the Field of Complex Numbers

We consider the spectral decomposition of the controllability gramian in simple and pair spectrum (2.2).

$$\sum_{k=1}^{n}\sum_{j=0}^{n}\sum_{\eta=0}^{n}\frac{s_k^j(-s_k)^\eta}{\dot N(s_k)\,N(-s_k)}\equiv\sum_{k=1}^{n}\sum_{\rho=1}^{n}\sum_{j=0}^{n-1}\sum_{\eta=0}^{n-1}\frac{-1}{s_k+s_\rho}\frac{s_k^j s_\rho^\eta}{\dot N(s_k)\,\dot N(s_\rho)},\quad s_k+s_\rho\neq 0. \tag{3.1}$$

We introduce the notation

$$\omega(n,s_k,j,\eta)=\sum_{k=1}^{n}\frac{s_k^j(-s_k)^\eta}{\dot N(s_k)\,N(-s_k)},$$
$$\omega(n,s_k,s_\rho,j,\eta)=\sum_{k=1}^{n}\sum_{\rho=1}^{n}\frac{-1}{s_k+s_\rho}\,\frac{s_k^j s_\rho^\eta}{\dot N(s_k)\,\dot N(s_\rho)}.$$

Taking into account the introduced notation, the identity (3.1) takes the form

$$\omega(n,s_k,j,\eta)\equiv\omega(n,s_k,s_\rho,j,\eta)\ \text{ for }\forall s_k,s_\rho\in\mathbb{C}^-,\ s_k+s_\rho\neq 0.$$

The proof follows from the decomposition of the fractional rational function $\frac{s_k^j(-s_k)^\eta}{\dot N(s_k)\dot N(-s_k)}$ by the roots of the characteristic equation $N(-s_k)=0$.

Lemma 1. *Consider the polynomial* $\gamma(n,s_k,-s_k)$ *over the field of complex numbers of the following form:*

$$\gamma(n,s_k,-s_k)=\sum_{i=0}^{n-1}\sum_{\mu=0}^{n-1}s_k^i(-s_k)^\mu,\ \forall k=1,\ldots,n,$$

where s_k *are roots of the characteristic equation of the system* (2.1), *and* $-s_k$ *are roots of the characteristic equation of its antistable conjugate system. We assume that all eigenvalues of the systems are prime, non-zero complex numbers. Then the polynomial* $\gamma(n,s_k,-s_k)$ *contains only all even powers of the numbers* s_k *and does not contain their odd powers.*

$$\gamma(n,s_k,-s_k)=\gamma\left(n,\ s_k^0,s_k^2,\ldots,s_k^{2m}\right),\quad n=2m,$$
$$\gamma(n,s_k,-s_k)=\gamma\left(n,\ s_k^1,s_k^3,\ldots,s_k^{2m-1}\right)\equiv 0,\quad n=2m-1.$$

Proof. It is easy to verify that the result of the lemma is valid for $n=1,2,3$:

$$\gamma(1,s_k,-s_k)=1,$$
$$\gamma(2,s_k,-s_k)=1-s_k^2,$$
$$\gamma(3,s_k,-s_k)=s_k^4-s_k^2+1.$$

Further, we apply the method of mathematical induction. We assume that the result of the lemma is true for the polynomial $\gamma(n,s_k,-s_k)$:

$$\gamma(n,s_k,-s_k)=\begin{cases}\gamma\left(2m,s_k^0,\ldots,s_k^{2m}\right) & \text{for even } n=2m,\\ \gamma\left(2m-1,s_k^0,\ldots,s_k^{2m}\right) & \text{for odd } n=2m-1.\end{cases}$$

We show that it is also valid for the polynomial $\gamma\,(n+1, s_k, -s_k)$. As n increases by one, the polynomial $\gamma\,(n, s_k, -s_k)$ takes the form

$$\gamma\,(n+1, s_k, -s_k) = \begin{cases} \gamma\,(2m+1, s_k, -s_k) \ \text{for even } n = 2m, \\ \gamma\,(2m, s_k, -s_k) \ \text{for odd } n = 2m-1. \end{cases}$$

We first consider the case of even n.

$$\gamma\,(n+1, s_k, -s_k) = \gamma\left(2m, s_k^0, \ldots, s_k^{2m}\right) + \gamma\left(2m, s_k^0, \ldots, s_k^{2m}\right) s_k^{2m+1}$$
$$- \gamma\left(2m, s_k^0, \ldots, s_k^{2m}\right) s_k^{2m+1} + s_k^{2m+1}(-s_k)^{2m+1} = \gamma\left(2m, s_k^0, \ldots, s_k^{2m}\right)^2 - s_k^{2(2m+1)}.$$

For the case of odd n, we similarly obtain

$$\gamma\,(n+1, s_k, -s_k) = \gamma\left(2m-1, s_k^0, \ldots, s_k^{2m}\right) + \gamma\left(2m-1, s_k^0, \ldots, s_k^{2m+1}\right) s_k^{2(2m+1)}$$
$$- \gamma\left(2m-1, s_k^0, \ldots, s_k^{2m+1}\right) s_k^{2(m+1)} + s_k^{2(2m+1)} = \gamma\left(2m-1, s_k^0, \ldots, s_k^{2m}\right) + s_k^{2(2m+1)},$$

where the first three terms contain even powers of s_k by assumption.

Corollary 1. *We consider the multiplier $\omega\,(n, s_k, j, \eta)$ in the spectral decomposition of the controllability gramian in the simple spectrum* (2.2). *The identities are valid:*

$$\omega\,(n, s_k, j, \eta) \equiv 0, \ \textit{if} \ \ j + \eta = \ 2m - 1, \tag{3.2}$$
$$\omega\,(n, s_k, j, \eta) \equiv \sum_{k=1}^{n} \frac{s_k^j(-s_k)^\eta}{\dot{N}\,(s_k)\,N\,(-s_k)}, \ \textit{if} \ \ j + \eta = \ 2m. \tag{3.3}$$

Proof. We express the multiplier through a polynomial $\gamma(n, s_k, -s_k)$

$$\omega\,(n, s_k, j, \eta\ ) \equiv \sum_{k=1}^{n} \frac{s_k^j(-s_k)^\eta}{\dot{N}\,(s_k)\,N\,(-s_k)} = \sum_{k=1}^{n} \frac{\gamma\,(n, s_k, -s_k, j, \eta)}{\dot{N}\,(s_k)\,N\,(-s_k)}$$

and apply the lemma.

Corollary 2. *Let us consider the multiplier $\omega\,(n, s_k, s_\rho, j, \eta)$ in the spectral decomposition of the controllability gramian in the pair spectrum* (2.2). *The identities are valid:*

$$\omega\,(n, s_k, s_\rho, j, \eta) \equiv 0, \ \textit{if} \ \ j + \eta = \ 2m - 1, \tag{3.4}$$
$$\omega\,(n, s_k, s_\rho, j, \eta) \equiv \sum_{k=1}^{n} \sum_{\rho=1}^{n} \frac{-1}{s_k + s_\rho} \ \frac{s_k^j s_\rho^\eta}{\dot{N}\,(s_k)\,\dot{N}\,(s_\rho)}, \ \textit{if} \ \ j + \eta = \ 2m. \tag{3.5}$$

Proof. We express the multiplier through a polynomial $\gamma\,(n, s_k, -s_k)$

$$\omega\,(n, s_k, j, \eta) \equiv \omega\,(n, s_k, s_\rho, j, \eta) \ \text{for} \ \forall s_k, s_\rho \in \mathbb{C}^-, \ s_k + s_\rho \neq 0$$

and apply the lemma.

Corollaries 1 and 2 prove that for all continuous stable MIMO LTI systems with a simple spectrum, reduced to the canonical forms of controllability and observability, exist spectral decompositions in the form of Xiao matrices. For systems represented in the canonical forms of controllability and observability, this allows to calculate only n diagonal elements using the formulas (3.2)–(3.5), instead of calculating n^2 matrix elements.

Remark. The multiplier $\omega\left(n, s_k, s_\rho, j, \eta\right)$ should be used with caution in the spectral decomposition of the controllability gramian in the pair spectrum (2.2). For example, in the case of a MIMO LTI system reduced to a diagonal canonical form, the spectral decomposition of the controllability gramian has a simple form

$$P_d^c = \sum_{j=1}^{n}\sum_{\rho=1}^{n} \frac{-b_{j\rho}}{s_j + s_\rho}\mathbb{1}_{j\rho}\ , \qquad b_{j\rho} = [B_d B_d^*]_{j\rho}. \tag{3.6}$$

On the other hand, we have

$$P_d^c = \sum_{j=0}^{n-1}\sum_{\rho=0}^{n-1}\omega(n, s_j, j, \rho) A_j B_d B_d^* A_\rho^*, \omega\left(n, s_k, j, \rho\right) = \sum_{k=1}^{n} \frac{s_k^j(-s_k)^\rho}{\dot{N}\left(s_k\right) N\left(-s_k\right)}. \tag{3.7}$$

We note that both formulas (3.6), (3.7) give the same numerical result, which corresponds to *different* spectral decompositions. Let's give an example.

Illustrative example 1

We consider the problem of controlling a two-zone furnace. The model of the control object of the heating furnace can be described by equations of state of the form

$$\Sigma_1\colon \begin{cases} \dfrac{dx}{dt} = Ax\left(t\right) + Bu\left(t\right), \quad x\left(0\right) = 0, \\ y\left(t\right) = Cx\left(t\right). \end{cases}$$

$$A = \begin{bmatrix} -0.5 & 0 \\ 0 & -1 \end{bmatrix},\ B = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix},\ C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In this case, you can evaluate the expressions

$$N\left(s\right) = s^2 + 1.5s + 0.5, \quad \dot{N}\left(s\right) = 2s + 1.5,$$

$$\left(Is - A\right)^{-1} = \begin{bmatrix} s+1 & 0 \\ 0 & s+0.5 \end{bmatrix}\left(s^2 + 1.5s + 0.5\right)^{-1},$$

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ A_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix},\ BB^T = \begin{bmatrix} 1.25 & 1.5 \\ 1.5 & 4.25 \end{bmatrix}.$$

The controllability gramian calculated using the formula (3.6), is equal to

$$P^c = \begin{bmatrix} 1.25 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 2.125 \end{bmatrix}.$$

The expression for the decomposition of the controllability gramian has the form

$$P^c = \sum_{j=0}^{n-1}\sum_{\rho=0}^{n-1}\sum_{k=1}^{2} \frac{s_k^j(-s_k)^\rho}{\dot{N}\left(s_k\right)N\left(-s_k\right)} A_j BB^{\mathrm{T}} A_\rho^{\mathrm{T}},$$

$$P^c = \sum_{j=0}^{n-1}\sum_{\rho=0}^{n-1}\sum_{k=1}^{2} \frac{s_k^j(-s_k)^\rho}{\dot{N}\left(s_k\right)N\left(-s_k\right)} \frac{-b_{j\rho}}{s_j + s_\rho}\mathbb{1}_{j\rho},$$

where A_j is the Faddeev matrix, constructed for the matrix A using the Faddeev–Le Verrier algorithm [6, 7]. Let's calculate the matrices $A_j BB^{\mathrm{T}} A_\rho^{\mathrm{T}}$:

$$A_0 BB^T A_0^T = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.0625 \end{bmatrix},\ A_0 BB^{\mathrm{T}} A_1^{\mathrm{T}} = \begin{bmatrix} 1.25 & 0.75 \\ 1.5 & 2.125 \end{bmatrix},$$

$$A_1 BB^{\mathrm{T}} A_0^{\mathrm{T}} = \begin{bmatrix} 1.25 & 1.5 \\ 0.75 & 2,125 \end{bmatrix},\ A_1 BB^{\mathrm{T}} A_1^{\mathrm{T}} = \begin{bmatrix} 1.25 & 1.5 \\ 1.25 & 4.25 \end{bmatrix}.$$

Substituting these expressions into (3.7), we obtain the spectral decomposition:

$$P^c = \begin{bmatrix} 1.25 & 0.75 \\ 0.75 & 1.0625 \end{bmatrix} \frac{2}{3} + \begin{bmatrix} 1.25 & 1.5 \\ 1.25 & 4.25 \end{bmatrix} \frac{1}{3} = \begin{bmatrix} 1.25 & 1 \\ 1 & 2.125 \end{bmatrix}.$$

Matrices of infinite sub-gramians are symmetric and positive definite, and so is their sum. We verify that the calculated controllability gramian is a solution to the Lyapunov equation by direct substitution. The separable spectral decomposition of the controllability gramian, calculated using the formula (3.6), has the form

$$P^c = \begin{bmatrix} 1.25 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 2.125 \end{bmatrix}.$$

The matrices of infinite sub-gramians in this decomposition are not symmetric and positive definite, although their sum is. The example shows that the same gramian can have several different spectral decompositions.

3.2. Decomposition of Gramians in the Form of Hadamard Products [3]

We introduce the matrices of the gramian controllability multiplier of a continuous MIMO LTI system in the form

$$\Omega_{\mathrm{c}} = [\omega_{\mathrm{c},j\eta}]_{n\times n}$$

and its observability gramian in the form

$$\Omega_o = [\omega_{o,j\eta}]_{n\times n},$$

where j is the row index, and η is the column index of the multiplier matrices.

We introduce matrices Ψ_{c} and Ψ_o in the form

$$\Psi_{\mathrm{c}} = \sum_{i=0}^{n-1}\sum_{\mu=0}^{n-1} A_i B B^{\mathrm{T}} A_\mu^{\mathrm{T}},$$

$$\Psi_o = \sum_{i=0}^{n-1}\sum_{\mu=0}^{n-1} A_i^{\mathrm{T}} C^{\mathrm{T}} C A_\mu.$$

We introduce an element-wise representation of these matrices in the form

$$\psi_{\mathrm{c},j\eta} = e_j^{\mathrm{T}} \Psi_{\mathrm{c}}\, e_\eta,$$

$$\psi_{o,j\eta} = e_j^{\mathrm{T}} \Psi_o\, e_\eta.$$

Theorem 1 [15]. *We consider a stable continuous dynamic MIMO LTI system with a simple spectrum*

$$\dot{x}(t) = Ax(t) + Bu\,(t)\,,\quad x\,(0) = 0,$$
$$y\,(t) = Cx\,(t)\,,$$

where $x\,(t) \in R^n,\ u\,(t) \in R^m,\ y(t) \in R^m$.

Then the controllability subramian P^c *is a matrix of the form* (2.2), *and in accordance with* [7] *formulas* (2.1), (2.2) *is defined as*

$$P^c = \sum_{j=0}^{n-1}\sum_{\eta=0}^{n-1} P^c_{j,\eta},\quad P^c_{j,\eta} = \omega(n, s_k, s_\rho, j, \eta) A_j B B^{\mathrm{T}} A_\eta^{\mathrm{T}}, \tag{3.8}$$

where

$$\omega\left(n, s_k, s_\rho, j, \eta\right) = \begin{cases} 0, & \textit{if index } j+\eta \textit{ is odd,} \\ \displaystyle\sum_{k=1}^{n}\sum_{\rho=1}^{n} \frac{-1}{s_\rho + s_k} \frac{s_k^j s_\rho^\eta}{\dot N\left(s_k\right)\dot N\left(s_\rho\right)}, & \textit{if index } j+\eta \textit{ is even.} \end{cases}$$

Proof of Theorem 1. As is known, the spectral decomposition of the controllability gramian under the conditions of Theorem 1 has the form [7, 20]

$$P^c = \sum_{j=0}^{n-1}\sum_{\eta=0}^{n-1}\sum_{k=1}^{n}\sum_{\rho=1}^{n} \frac{-1}{s_\rho + s_k} \frac{s_k^j s_\rho^\eta}{\dot N\left(s_k\right)\dot N\left(s_\rho\right)} A_j B B^{\mathrm{T}} A_\eta^{\mathrm{T}}.$$

We substitute the newly introduced scalar function $\omega\left(n, s_k, s_\rho, j, \eta\right)$ into this formula and obtain the formula (3.8).

Theorem 2 [13]. *We consider a continuous MIMO LTI system of the form* (2.1)*. We assume that the system is stable and all the roots of its characteristic equation are different. Then its controllability and observability gramians have the form of the generalized Xiao matrices of the following form:*

$$P^c = \Omega_c \circ \Psi_c = \left[p_{j\eta}^c\right]_{n\times n}, \qquad j,\eta = 1,\ldots,n. \tag{3.9}$$

$$\Psi_c = \left[\psi_{c,j\eta}\right]_{n\times n}, \; j,\eta = 1,\ldots,n. \qquad \Psi_c = \sum_{i=0}^{n-1}\sum_{\mu=0}^{n-1} \Psi_{c,i\mu}, \; \Psi_{c,i\mu} = M_i M_\mu^*,$$

$$M_i = A_i B, \qquad \Omega_c = \left[\omega_c(n,j,\eta)\right]_{n\times n}, \qquad j,\eta = 1,\ldots,n.$$

$$p_{j\eta}^c = \omega_c(n,j,\eta) \times \psi_{c,j\eta}.$$

Proof of Theorem 2. We use the spectral decomposition of the controllability gramian (2.2). We introduce a representation of gramians in the form of Hadamard products

$$P^c = \Omega_{\mathrm{c}} \circ \Psi_{\mathrm{c}}, \tag{3.10}$$

$$P^o = \Omega_o \circ \Psi_o. \tag{3.11}$$

This representation allows us to write simple formulas for calculating the elements of the controllability and observability gramians of MIMO LTI systems P^c and P^o in the form [13]

$$p_{j\eta}^c = \omega_c(n,j,\eta) \times \psi_{c,j\eta}, \tag{3.12}$$

$$p_{j\eta}^o = \omega_c(n,j,\eta) \times \psi_{o,j\eta}. \tag{3.13}$$

Next, we use identities for one class of stable polynomials whose roots are different over the field of complex numbers (Lemma). Formulas (3.10)–(3.13) express algorithms for calculating elements of generalized Xiao matrices in the form of products of elements of multiplier matrices and elements of sums of all possible products of matrices $A_j B B^{\mathrm{T}} A_\eta^{\mathrm{T}}$, written in the form of products of Hadamard matrices

$$\Omega_{\mathrm{c}} \circ \Psi_{\mathrm{c}}.$$

Corollary 3. *We consider an important special case of continuous linear SISO systems represented by equations of state in the canonical forms of controllability and observability. In this case, the controllability and observability gramians are determined by the formulas* [15]

$$P^{cF} = \sum_{k=1}^{n}\sum_{\eta=0}^{n-1}\sum_{j=0}^{n-1} \frac{s_k^j(-s_k)^\eta}{\dot N\left(s_k\right) N\left(-s_k\right)} \mathbb{1}_{j+1\eta+1}, \tag{3.14}$$

$$P^{oF} = \sum_{k=1}^{n}\sum_{\eta=0}^{n-1}\sum_{j=0}^{n-1} \frac{s_k^j(-s_k)^\eta}{\dot N\left(s_k\right) N\left(-s_k\right)} \mathbb{1}_{j+1\eta+1}.$$

The representation of gramians in Hadamard form according to (3.10)–(3.11) *takes the form*

$$P^{cF} = \Omega_{cF} \circ \Psi_c, \quad \Psi_c = \sum_{\eta=0}^{n-1} \sum_{j=0}^{n-1} \mathbb{1}_{j+1\eta+1},$$
$$P^{oF} = \Omega_{oF} \circ \Psi_o, \quad \Psi_o = \sum_{\eta=0}^{n-1} \sum_{j=0}^{n-1} \mathbb{1}_{j+1\eta+1.}$$

This implies the identities

$$P^{cF} \equiv \Omega_{cF}, \tag{3.15}$$
$$P^{oF} \equiv \Omega_{oF}. \tag{3.16}$$

This means that the controllability gramian in the canonical form of controllability coincides with the multiplier matrix for this gramian, which allows us to apply the formulas (3.15), (3.16) *to calculate all elements of the gramian and establishes that the gramian belongs to the class of Xiao matrices. A similar result holds for the gramian observability in the canonical form of observability.*

Multiplier matrices in different canonical forms have the form

$$\Omega_{cF} \equiv \Omega_{oF} = \left[\omega\left(n, s_k, s_\rho, j, \eta\right)\right]_{n\times n} = \left[\omega\left(n, s_k, j, \eta\right)\right]_{n\times n}.$$

3.3. Spectral and Singular Decompositions of Inverse Matrices of Gramians

General formulas for calculating inverse gramian matrices (hereinafter inverse gramians) for continuous MIMO LTI systems have the form [1]

$$(P^c)^{-1} = \frac{-1}{\gamma_0}\left[(P^c)^{n-1} + \gamma_{n-1}(P^c)^{n-2} + \cdots + \gamma_2 P^c + \gamma_1 I\right];$$
$$(P^o)^{-1} = \frac{-1}{\gamma_0}\left[(P^o)^{n-1} + \gamma_{n-1}(P^o)^{n-2} + \cdots + \gamma_2 P^o + \gamma_1 I\right].$$

In the case of continuous SISO LTI systems, these formulas, in accordance with (3.15), (3.16), take the form

$$\left[P^{cF}(\omega(n, s_k, j, \eta))\right]^{-1} = \frac{-1}{\gamma_0}\left[(\Omega_{cF})^{n-1} + \gamma_{n-1}(\Omega_{cF})^{n-2} + \cdots + \gamma_2\Omega_{cF} + \gamma_1 I\right];$$
$$\left[P^{oF}(\omega(n, s_k, j, \eta))\right]^{-1} = \frac{-1}{\gamma_0}\left[(\Omega_{oF})^{n-1} + \gamma_{n-1}(\Omega_{oF})^{n-2} + \cdots + \gamma_2\Omega_{oF} + \gamma_1 I\right].$$

The presence of powers of the multiplier matrices on the right side of the formulas leads to the appearance of complex fractional rational functions of eigenvalues s_k, which limits the scope of application of the formulas for spectral decompositions of inverse gramians to systems of small and medium dimensions. We return to stable continuous MIMO LTI systems with a simple spectrum and note that the controllability and observability gramians are symmetric complex-valued matrices. In this case, there are their singular decompositions of the form [1]

$$P^c = P^{c*} = V_c \Lambda V_c^*,$$
$$P^o = P^{o*} = V_o \Lambda V_o^*,$$

where the matrix V_c is formed by the right singular vectors of the matrix P^c, the matrix V_c^* is formed by the left singular vectors of the matrix P^c, and the matrix Λ is a diagonal matrix of the form

$$\Lambda = diag\left\{|\lambda_1|\,|\lambda_2|\ldots|\lambda_n|\right\}.$$

We define matrices S and U in the form

$$S = \ diag\left\{sgn\lambda_1 \ sgn\lambda_2 \ldots sgn\lambda_n \ \right\}, \ U_c = V_c S,$$
$$sgn\lambda = \begin{cases} +1, & \text{if } \lambda \geqslant 0, \\ -1, & \text{if } \lambda < 0. \end{cases}$$

Then

$$P^c = U_c \Lambda V_c^*,$$
$$P^o = U_o \Lambda V_o^*,$$

where the matrix U_c is formed by the left singular vectors of the matrix P^c. Since Λ, U_c, V_c are nonsingular matrices, then

$$(P^c)^{-1} = (U_c)^{-1}\Lambda^{-1}(V_c^*)^{-1} = V_c^*\Lambda^{-1}U_c. \tag{3.17}$$

In a similar way we get

$$(P^o)^{-1} = (U_o)^{-1}\Lambda^{-1}(V_o^*)^{-1} = V_o^*\Lambda^{-1}U_o. \tag{3.18}$$

Since the matrix Λ is diagonal, its inverse matrix can be written as

$$\Lambda^{-1} = \left[|\lambda_1|^{-1}\mathbb{1}_{11} + |\lambda_2|^{-1}\mathbb{1}_{22} + \cdots + |\lambda_n|^{-1}\mathbb{1}_{nn}\right]. \tag{3.19}$$

Substituting (3.19) into (3.17), (3.18), we obtain the following singular decompositions of the inverse gramians of controllability and observability in terms of their singular spectrum:

$$(P^c)^{-1} = V_c^*\left[|\lambda_1|^{-1}\mathbb{1}_{11} + |\lambda_2|^{-1}\mathbb{1}_{22} + \cdots + |\lambda_n|^{-1}\mathbb{1}_{nn}\right]U_c;$$
$$(P^o)^{-1} = V_o^*\left[|\lambda_1|^{-1}\mathbb{1}_{11} + |\lambda_2|^{-1}\mathbb{1}_{22} + \cdots + |\lambda_n|^{-1}\mathbb{1}_{nn}\right]U_o.$$

Theorem 3. *Consider a continuous stable and fully controllable dynamic MIMO LTI system of the form* (2.1).

The singular decompositions of its inverse controllability gramian in terms of the eigenvalues of the gramian matrix have the following form.

For a simple spectrum of the gramian matrix

$$(P^c)^{-1} = \frac{\sum_{\lambda=1}^{n}\sum_{j=0}^{n-1} P_j^c \sigma_\lambda^j}{\dot{N}_c(\sigma_\lambda)} \frac{1}{\sigma_\lambda}, \tag{3.20}$$

where P^c is the gramian controllability matrix, P_j^c is the Faddeev matrix in the decomposition of the gramian resolvent, σ_λ is the eigenvalue of the gramian matrix P^c.

For the multiple spectrum of the gramian matrix

$$(P^c)^{-1} = -\sum_{\delta=1}^{q}\sum_{\rho=1}^{m_\delta} \frac{K_{\delta\rho}}{(-\sigma_\delta)^{m_\delta - j + 1}}, \tag{3.21}$$

$$K_{\delta\rho} = \frac{1}{(\rho-1)!}\left\{\frac{d^{\rho-1}}{d\sigma^{\rho-1}}\left[\frac{(\sigma-\sigma_\delta)^{m_\delta}\sum_{j=0}^{n-1}\sigma^j P_j^c}{\prod_{\delta=1}^{n}(\sigma-\sigma_\delta)^{m_\delta}}\right]\right\}\Bigg|_{s=s_\delta}, \tag{3.22}$$

where P^c is the gramian controllability matrix, P_j^c is the Faddeev matrix in the decomposition of the gramian resolvent, σ_δ is the eigenvalue of the gramian matrix P^c multiplicity m_δ, ρ is the multiplicity index of the eigenvalue σ_δ.

Proof of Theorem 3. We consider the decomposition of the resolvent of the gramian controllability matrix in the form of a segment of the Faddeev series [6]

$$(I\sigma - P^c)^{-1} = \frac{\sum_{j=0}^{n-1} P_j^c \sigma^j}{N_c(\sigma)}\ . \tag{3.23}$$

We denote: $N_c\,(\sigma) = s^n + a_{c,n-1}\sigma^{n-1} + \ldots a_{c,1}\sigma + a_{c,0}$, $j = 1, \ldots, n$; $N_c\,(\sigma)$ is characteristic polynomial of the resolvent of the gramian matrix, P_j^c is the Faddeev matrix in the decomposition of the resolvent in the Faddeev series, $j = 1, \ldots, n$.

We first consider the case when all singular values σ_λ of the gramian are different. In this case, the decomposition (3.23) is transformed to the form

$$(I\sigma - P^c)^{-1} = \frac{\sum_{\lambda=1}^{n} \sum_{j=0}^{n-1} P_j^c \sigma_\lambda^j}{\dot{N}_c(\sigma_\lambda)} \frac{1}{\sigma - \sigma_\lambda}. \tag{3.24}$$

Iterative algorithm for calculating Faddeev matrices and coefficients of the characteristic equation:

First step: $a_{c,n-1} = 1,\ R_n = I,$

Step "k": $a_{c,n-k} = -\frac{1}{k} tr\,(P^c R_{n-k+1})\,,\quad R_{n-k} = a_{c,n-k} I + P^c R_{n-k+1},\quad k = 1, \ldots, n;$

In accordance with the Faddeev–Le Verrier algorithm, the following matrix equalities are also valid:

$$\begin{gathered}
P_0^c = a_{c,1} I + a_{c,2} P^c + \cdots + a_{c,n} (P^c)^{n-1},\\
P_1^c = a_{c,2} I + a_{c,3} P^c + \cdots + a_{c,n} (P^c)^{n-2},\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
P_{n-2}^c = a_{c,n-1} I + a_{c,n} P^c,\\
P_{n-1}^c = a_{c,n} I.
\end{gathered}$$

The above system can be written in the form

$$P_j^c = \sum_{j=0}^{n-1} a_{c,j+1} (P^c)^j,\quad \forall j : j = 0, 1, \ldots, n-1.$$

We put $\sigma = 0$ in (3.24) and get the formula (3.20):

$$(P^c)^{-1} = \frac{\sum_{\lambda=1}^{n} \sum_{j=0}^{n-1} P_j^c \sigma_\lambda^j}{\dot{N}_c(\sigma_\lambda)} \frac{1}{\sigma_\lambda}. \tag{3.25}$$

Thus, (3.20)–(3.25) in the case of a simple spectrum, the gramian matrices determine the singular decomposition of the inverse gramian of controllability. A similar approach can be applied to the case of a multiple spectrum of the gramian matrix. We assume that the characteristic equation of the gramian matrix can be represented in the form

$$N_c\,(\sigma) = \prod_{i=1}^{n} (\sigma - \sigma_i)^{m_i},\qquad \sum_{i=1}^{n} m_i = q,\quad \sigma_i \in C^+.$$

For any square gramian matrix, its resolvent has the form of a matrix function (3.24). In accordance with [22, 23], its decomposition into simple fractions has the form

$$(I\sigma - P^c)^{-1} = \sum_{\delta=1}^{q} \sum_{\rho=1}^{m_\delta} \frac{K_{\delta\rho}}{(\sigma - \sigma_\delta)^{m_\delta - j + 1}}, \tag{3.26}$$

$$K_{\delta\rho} = \frac{1}{(\rho-1)!} \left\{ \frac{d^{\rho-1}}{d\sigma^{\rho-1}} \left[\frac{(\sigma - \sigma_\delta)^{m_\delta} \sum_{j=0}^{n-1} \sigma^j P_j^c}{\prod_{\delta=1}^{n} (\sigma - \sigma_\delta)^{m_\delta}} \right] \right\} \Bigg|_{\sigma = \sigma_\delta}.$$

We set $\sigma = 0$ in (3.26) and obtain formulas (3.21)–(3.22) for the singular decomposition of the inverse gramian of controllability for the case of a multiple spectrum of the gramian matrix.

Illustrative example 2

We consider the problem of controlling an asynchronous motor. The model of the control object can be described by equations of state of the form

$$\Sigma_1\colon \begin{cases} \dfrac{dx}{dt} = Ax\,(t) + Bu\,(t)\,, \quad x\,(0) = 0, \\ y\,(t) = Cx\,(t)\,. \end{cases}$$

$$A = \begin{bmatrix} -4.67 & 3 & -1.33 & 2.33 \\ -2.17 & 2.33 & -3.83 & 5.17 \\ 1.5 & -0.33 & -1.5 & 0.17 \\ 2.17 & -3.33 & 3.83 & -6.17 \end{bmatrix}, \ B = \begin{bmatrix} 3 \\ -3 \\ -7 \\ -4 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We present the eigenvalues of the system dynamics matrix

$$\lambda_i = -4;\ -3;\ -2;\ -1.$$

To construct a singular decomposition of the inverse gramian of the controllability of the system by the singular values of the gramian matrix, we calculate the gramian of controllability using the formula (3.6)

$$P^c = \begin{bmatrix} 2.5 & 3 & 2.5 & 0.56 \\ 3 & 11.2 & 13.2 & 5.1 \\ 2.5 & 13.2 & 16.6 & 6.9 \\ 0.56 & 5.1 & 6.9 & 3 \end{bmatrix}.$$

We note that the formula (3.6) is valid not only for stable linear systems, but also for unstable systems in which the condition $s_k + s_p \neq 0$ is not violated. It is violated in the case of $s_k = 0$ or $s_k = +j\omega$, $s_{k+1} = -j\omega$ [21].

Then the singular numbers of this gramian take the form

$$\sigma_i = 30.7;\ 2.5;\ 0.17;\ 0.0002.$$

The controllability gramian of the system is represented by a symmetric matrix, therefore there is its SVD decomposition [1]

$$P^c = \begin{bmatrix} -0.13 & 0.86 & -0.48 & -0.009 \\ -0.6 & 0.28 & 0.67 & 0.35 \\ -0.73 & -0.25 & -0.25 & -0.58 \\ -0.3 & -0.32 & -0.51 & 0.74 \end{bmatrix} \times \begin{bmatrix} 31 & 0 & 0 & 0 \\ 0 & 2.5 & 0 & 0 \\ 0 & 0 & 0.17 & 0 \\ 0 & 0 & 0 & 0.0002 \end{bmatrix}$$

$$\times \begin{bmatrix} -0.13 & 0.86 & -0.48 & -0.009 \\ -0.6 & 0.28 & 0.67 & 0.35 \\ -0.73 & -0.25 & -0.25 & -0.58 \\ -0.3 & -0.32 & -0.51 & 0.74 \end{bmatrix}.$$

In accordance with the Faddeev–Le Verrier algorithm, we calculate the Faddeev matrices and the coefficients of the characteristic equation for the inverse gramian

$$P_0^c = \begin{bmatrix} -0.01 & 0.05 & -0.07 & 0.08 \\ 0.05 & -1.62 & 2.69 & -3.4 \\ -0.07 & 2.69 & -4.5 & 5.6 \\ 0.08 & -3.4 & 5.6 & -7.2 \end{bmatrix}, \ P_1^c = \begin{bmatrix} 21.8 & -23.5 & 8.4 & 17 \\ -23.5 & 44.6 & -29.5 & -5.5 \\ 8.4 & -29.5 & 33 & -25 \\ 17 & -5.5 & -25 & 65.7 \end{bmatrix},$$

$$P_2^c = \begin{bmatrix} -31 & 3 & 2.5 & 0.56 \\ 3 & -22.1 & 13.2 & 5.1 \\ 2.5 & 13.2 & -16.7 & 6.9 \\ 0.56 & 5.1 & 6.9 & -30.3 \end{bmatrix}, \ P_3^c = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$a_{c,0} = 0.0031, \ a_{c,1} = -13.3, \ a_{c,2} = 82.6, \ a_{c,3} = -33.3, \ a_{c,4} = 1.$$

Then the inverse gramian can be calculated using the formula (3.20)

$$(P^c)^{-1} = \begin{bmatrix} 1.97 & -14.8 & 22.3 & -26.2 \\ -14.8 & -517 & -856 & 1083 \\ 22.3 & -856 & 1422 & -1803 \\ -26.2 & 1083 & -1803 & 2290 \end{bmatrix}.$$

3.4. Spectral Decompositions of Energy Functionals and New Stability Criteria

Within the framework of the SISO LTI assumptions made above, we consider a system of the form (2.4), the equations of state of which are reduced to the canonical form of controllability, and calculate the energy functional J, which is the value of the square H$_2$ is the norm of the transfer functions of the system and gives an assessment of the risk of loss of stability [1, 19, 22]. To do this, we use (3.12) and (3.14) and, for definiteness, choose the spectral decomposition of the controllability gramian in a simple spectrum

$$J = trC^F \Omega_c (C^F)^T$$
$$= \left(\frac{{\xi_0}^2}{\sum\limits_{k=1}^{n} \dot{N}(s_k) N(-s_k)} - \frac{{\xi_1}^2 \sum\limits_{k=1}^{n} s_k^2}{\sum\limits_{k=1}^{n} \dot{N}(s_k) N(-s_k)} + \cdots + \frac{(-1)^{n-1} {\xi_{n-1}}^2 \sum\limits_{k=1}^{n} s_k^{2n}}{\sum\limits_{k=1}^{n} \dot{N}(s_k) N(-s_k)} \right). \tag{3.27}$$

We obtained an invariant spectral decomposition of the energy functional over the simple spectrum of the dynamics matrix. This simple formula shows the advantage of using spectral decompositions in canonical form over the general decomposition (3.23). The decomposition does not depend on the choice of a non-singular matrix of linear transformations of the system coordinates. Two main factors influence the value of the buckling risk J:

1) the values of the diagonal terms of the Xiao matrix Ω_c,

2) the squares of the elements of the reduced output vector.

The expression (3.27) can be simplified by simplifying the SISO LTI system. In [14] it is shown that the Xiao matrix is the controllability gramian for a SISO LTI system with a transfer function

$$W(s) = \frac{y(s)}{u(s)} = \frac{1}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0}. \tag{3.28}$$

The asymptotic stability of SISO LTI systems of the form (2.4) is equivalent to the asymptotic stability of this system. Moreover, we will show that the asymptotic stability of MIMO LTI systems of the form (2.1) is equivalent to its asymptotic stability. The energy functional J for the system (3.28) according to (3.27) is equal to

$$J = \left(\frac{1}{\sum_{k=1}^{n} \dot{N}(s_k)N(-s_k)} - \frac{\sum_{k=1}^{n} s_k^2}{\sum_{k=1}^{n} \dot{N}(s_k)N(-s_k)} + \ldots + \frac{(-1)^{n-1}\sum_{k=1}^{n} s_k^{2n}}{\sum_{k=1}^{n} \dot{N}(s_k)N(-s_k)} \right). \tag{3.29}$$

Theorem 4. *We consider a continuous fully controllable dynamic MIMO LTI system with a simple spectrum of the form* (2.1)*, as well as a continuous dynamic SISO LTI system with the same spectrum, the equations of state of which are reduced to the canonical form of controllability of the form* (2.4)*.*

Then a sufficient condition for the asymptotic stability of the system (2.1) *according to Lyapunov is the boundedness of the energy functional* (3.29) *for a SISO LTI system with the same spectrum and transfer function*(3.28) *is*

$$J < +\infty, \tag{3.30}$$

$$\textit{for any } s_k \textit{ belonging to } C^-, \;\; k = 1, \ldots, n. \tag{3.31}$$

Proof of Theorem 4. Let us recall that the MIMO LTI system (2.1) is completely controllable and observable, all eigenvalues of matrix A are different, the implementation of the system (2.1) is minimal and there is a single transfer function of the system. When the specified conditions are met, the boundedness of the functional $\sqrt{J}$ is a necessary and sufficient condition for the asymptotic stability of the system (2.1) according to Lyapunov [1, Theorem 5.14]. Thus, the boundedness of the functional J is a sufficient condition for the asymptotic stability of the system (3.27)

$$J < \infty.$$

But the functional J is a trace of the Xiao matrix SISO LTI system (2.4), the equations of state of which are reduced to the canonical form of controllability. This leads to the conclusion that the boundedness of the energy functional of a simple SISO LTI system (3.28) in the form of inequality (3.30) guarantees the asymptotic stability of a complex MIMO LTI system of the form (2.1). Verification of the condition (3.31) requires the use of asymptotic gramian models [22].

Thus, a new criterion for the stability of a complex stationary linear dynamic MIMO LTI system is obtained in the form of a criterion for the boundedness of the trace of the Xiao matrix Ω_c for a simple SISO LTI system (3.21), the equations of which are reduced to the canonical form of controllability. The new criterion does not contradict the well-known criterion that the eigenvalues of the dynamics matrix of a linear system belong to the left half-plane of the plane of eigenvalues, but refines it taking into account the nonlinear effects of mode interaction (multiple eigenvalues, close aperiodic and vibrational modes) [22].

4. CONCLUSION

This article is dedicated to the development of spectral methods for solving the Lyapunov equation. The main results are obtained using structural methods in developing new methods and tools that are closely related to the fundamental properties of linear dynamic systems: controllability, observability and stability. Among the solution methods, two should be mentioned first of all: determining the structure of the solution matrix in the form of the Xiao matrix and spectral decompositions of the solution in the form of Hadamard products. A method and algorithm for calculating matrices in the form of the Hadamard product for multiply connected continuous linear systems with many inputs and many outputs has been developed. This allows us to calculate the elements of the corresponding controllability and observability gramians in the form of products of the corresponding elements of the multiplier matrices and a matrix that is the sum of all possible products of the numerator matrices of the matrix transfer function of the system. When using the canonical forms of controllability or observability, the Hadamard decomposition of the corresponding gramians is reduced to a multiplier matrix, the trace of which is equal to the energy functional of the SISO LTI system. New results are obtained in the form of spectral and singular decompositions of the inverse gramians of controllability and observability. This makes it possible to obtain invariant decompositions of energy functionals and formulate new criteria for the stability of linear systems taking into account the nonlinear effects of mode interaction [20].

FUNDING

This work was supported in part by the Russian Science Foundation, project no. 19-19-00673.

REFERENCES

1. Antoulas, A.C., *Approximation of Large-Scale Dynamical Systems*, Philadelphia: SIAM, 2005.

2. Polyak, B.T., Khlebnikov, M.V., and Rapoport, L.B., *Teoriya avtomaticheskogo upravleniya* (Theory of Automatic Control), Moscow: LENAND, 2019.

3. Zubov, N.E., Zybin, E.Yu., Mikrin, E.A., Misrikhanov, M.Sh., and Ryabchenko, V.N., General Analytical Forms for Solving the Sylvester and Lyapunov Equations for Continuous and Discrete Dynamic Systems, *Theory and control systems*, 2017, no. 1, pp. 3–20. https://doi.org/10.1134/S1064230717010130

4. Gantmakher, F.R., *Teoriya matrits* (Theory of Matrices), Moscow: Nauka, 1966. Translated into English under the title *Theory of Matrices*, New York: Chelsea, 1959.

5. Ikramov, Kh.D., *Chislennoe reshenie matrichnykh uravnenii* (Numerical Solution of Matrix Equations), Moscow: Nauka, 1984.

6. Faddeev, D.K. and Faddeeva, V.N., *Vychislitel'nye metody lineinoi algebry* (Computational Methods of Linear Algebra), Moscow: Lan', 2009.

7. Kwakernaak, H. and Sivan, R., *Linear Optimal Control Systems*, New York: Wiley, 1972. Translated under the title *Lineinye optimal'nye sistemy upravleniya*, Moscow: Mir, 1977.

8. Andreev, Yu.N., *Upravlenie konechnomernymi lineinymi ob"ektami* (Management of Finite-Dimensional Linear Objects), Moscow: Nauka, 1976.

9. Godunov, S.K., *Lektsii po sovremennym aspektam lineinoi algebry* (Lectures on Modern Aspects of Linear Algebra), Novosibirsk: Nauchnaya kniga, 2002.

10. Proskurnikov, A.V. and Fradkov, A.L., Problems and Methods of Network Control, *Autom. Remote Control*, 2016, vol. 77, no. 10, pp. 1711–1740. https://doi.org/10.1134/S0005117916100015

11. Zhabko, A.P. and Kharitonov, V.L., *Metody lineinoi algebry v zadachakh upravleniya: uchebnoi posobie* (Methods of Linear Algebra in Control Problems), St. Petersburg: S.-Peterburg. Gos. Univ., 1993.

12. Sreeram, V. and Agathoklis, P., Solution of Lyapunov Equation with System Matrix in Companion Form, *IEE Proc. D. Control. Theory Appl.*, 1991, vol. 138, no. 6, pp. 529–534. https://doi.org/10.1049/ip-d.1991.0074

13. Xiao, C., Feng, Z., and Shan, X., On the Solution of the Continuous-Time Lyapunov Matrix Equation in Two Canonical Forms, *IEE Proc.*, 1992, vol. 139, no. 3, pp. 286–290. https://doi.org/10.1049/ip-d.1992.0038

14. Hauksdottir, A. and Sigurdsson, S., The continuous closed form controllability Gramian and its inverse, *2009 American Control Conference Hyatt Regency Riverfront, St. Louis, MO, USA June 10–12*, 2009, pp. 5345–5351. https://doi.org/978-1-4244-4524-0/09

15. Yadykin, I.B., Spectral Decompositions of gramians of Continuous Stationary Systems Given by Equations of State in Canonical Forms, *Mathematics*, 2022, vol. 10, no. 13, pp. 2339. https://doi.org/10.3390/math10132339

16. Dilip, A.S.A., The Controllability Gramian, the Hadamard Product and the Optimal Actuator, *Leader Sensor Select. Problem Nature Phys.*, 2015, vol. 11, pp. 779–786. https://doi.org/10.1109/LCSYS.2019.2919278

17. Bianchin, G. and Pasqualetti, F., Gramian-Based Optimization for the Analysis and Control of Traffic Networks, *IEEE Transactions on Intelligent Transportation Systems*, 2022, vol. 21, no. 7, pp. 3013–3024. https://doi.org/10.1109/TITS.2019.2922900

18. Himpe, C., The Empirical gramian Framework, *Algorithms*, 2018, vol. 11, no. 91. https://doi.org/10.3390/a11070091

19. Benner, P., Goyal, P., and Duff, I.P., Gramians, Energy Functionals, and Balanced Truncation for Linear Dynamical Systems With Quadratic Outputs, *IEEE Transact. Autom. Control*, 2022, vol. 67, no. 2, pp. 886–893. https://doi.org/10.1109/TAC.2021.3086319

20. Yadykin, I.B., On Properties of Gramians of Continuous Control Systems, *Autom. Remote Control*, 2010, vol. 71, no. 6, pp. 1011–1021.

21. Yadykin, I.B. and Galyaev, A.A., On the Methods for Calculation of Grammians and Their Use in Analysis of Linear Dynamic Systems, *Autom. Remote Control*, 2013, vol. 74, no. 2, pp. 207–224.

22. Yadykin, I.B. and Iskakov, A.B., Energy Approach to Stability Analysis of the Linear Stationary Dynamic Systems, *Autom. Remote Control*, 2016, vol. 77, no. 12, pp. 2132–2149.

23. Gardner, M.F. and Barns, J.L., *Transients in Linear Systems Studied by the Laplace Transformation. V. 1. Lumped-Constant Systems*, New York, London: Wiley, Chapman and Hall, 1942.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1428–1436.

Probabilistic Assessment of a Pentapeptide Composition Influence on Its Stability

A. I. Mikhalskii^{*,a}**, J. A. Novoseltseva**^{*,b}**, A. A. Anashkina**^{**,c}**, and A. N. Nekrasov**^{***,d}

**Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
***Engelgardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*
****Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia*
e-mail: ^{a}*ipuran@yandex.ru,* ^{b}*novoselc.janna@yandex.ru,* ^{c}*a_anastasya@inbox.ru,* ^{d}*a_nnekrasov@mail.ru*

Received May 31, 2023
Revised September 12, 2023
Accepted September 30, 2023

Abstract—The influence of the arrangement of amino acid residues in a pentapeptide on its stability is being studied. A forecast of pentapeptide stability is made using the gradient boosting method, which allows one to evaluate the influence of each feature on the stability of the pentapeptide. Combinations of amino acid arrangements in the pentapeptide have been identified that make a significant contribution to its stability. It has been shown that the use of such combinations reduces the amount of data required to obtain a reliable prediction of pentapeptide stability.

Keywords: amino acid residue, pentapeptide, gradient boosting, prediction, information sufficiency

DOI: 10.25728/arcRAS.2024.26.65.001

1. INTRODUCTION

The problem of predicting the spatial structure of proteins is one of the priority tasks in the field of mathematical and biological modeling, leading to practical application—the design of new proteins with useful medical properties. Currently, there is a tool for predicting the tertiary structure of a protein from its amino acid sequence, AlphaFold [1], which has shown incredible accuracy of structure prediction, comparable to the accuracy of X-ray diffraction analysis on CASP [2]. However, this tool is based on a deep neural network and the principles behind the styling remain unexplored. Understanding which amino acids and in what combination contribute to increasing the stability of a protein fragment will allow us to create a method for designing protein structure. The goal of the work is to, based on experimental data, identify potential markers of pentapeptides stability (combinations and positions of amino acids in the molecule).

The study of the entropy characteristics of fragments of protein sequences showed that for five sequentially located residues a reduced level of information entropy is observed and, therefore, blocks of this particular size must be considered as elementary units of the sequence. This approximation made it possible to develop a method that reveals the hierarchical structure in protein sequences — the method of analyzing information structure (ANIS method) [3]. Analysis of the conformational stability of pentapeptides by the molecular dynamics method showed that all pentapeptides can be divided into three types [4]: conformationally stable (located in the predominant topology for more than 80% of the simulation time), triggered (having two predominant topologies, in each of which the peptide was present for at least 40% of the simulation time) and labile. Molecular dynamics is a technique in which the time evolution of a system of interacting atoms or particles is tracked by integrating their equations of motion. Classical mechanics is used to describe atoms or particles and their motion. The law of particle motion is found using analytical mechanics, and the forces of interatomic interaction are represented in the form of classical potential forces (as the gradient of the potential energy of the system).

2. DATA

The work used 44 860 pentapeptides, the sequences of which were created according to a certain rule, and 4885 previously studied pentapeptides from real proteins, the stability of which was determined by molecular dynamics modeling. In the resulting set of 49 745 pentapeptides, only 1705 pentapeptides turned out to be stable, which was about 3.43% of the total number.

During the study, all data were randomly divided into “training”, “control” and “validation” samples in proportions of 0.66, 0.17, 0.17, maintaining the original balance of classes. The training and control samples were used during the training phase. The training sample was also used at the stage of interpretation of the results.

2.1. Data Encodings

In the original data set, each pentapeptide is encoded by a sequence of five letters representing the amino acid residues included in the pentapeptide. The order of the letters corresponds to the sequence of amino acid residues in the pentapeptide molecule. For formal numerical analysis of the data, the five-letter representation was encoded using three different representations. Binary encoding (One Hot Encoding), continuous string representation (n-gram), and discontinuous string representation (broken n-gram) were considered. Each of the considered coding methods allows one to evaluate contributions in its own way and make judgments about the influence of certain combinations of amino acid residues on the stability of the pentapeptide.

2.2. Binary Encoding (OHE)

One hot encoding is an encoding in which the presence of each amino acid at its position is determined by the position of one in the vector, the remaining coordinates of which are equal to zero. The number of vector elements is 20 — the number of amino acid types. As a result, each pentapeptide is encoded by a matrix of 20 rows and 5 columns. The column corresponds to the amino acid position in the pentapeptide molecule, and the row corresponds to the amino acid. For example, when classifying amino acids by the first letter of the name, the pentapeptide DKLNV will be encoded by a matrix in which the first column in the third row is 1, the remaining elements are equal to zero, the second column in the ninth row is 1, the remaining elements are equal to zero, etc. In the calculations, each pentapeptide is represented as a vector in 100-dimensional space.

2.3. Continuous String Representation (n-gram)

An n-gram is a continuous string representation of the amino acid sequence in a pentapetide. Depending on the number of letters included in the string, n-grams of order 1, 2 or more are distinguished. For example, the peptide DKLNV is encoded by five n-grams of order 1 D, K, L, N, V, four n-grams of order 2 DK, KL, LN, NV, three n-grams of order 3 DKL , KLN, LNV. In the analysis carried out, n-grams from 1 to 3 were used. As with OHE encoding, the entire set of n-grams encoding pentapeptides is represented in the form of a table consisting of zeros and ones. In each column of the table, a certain row contains one, and the remaining elements are zeros.

2.4. Discontinuous String Representation (Broken n-gram)

A broken n-gram is a generalization of a continuous n-gram and is a string representation of the sequence of amino acids in a pentapeptide, in which there is a gap of one to three characters between groups of amino acids. When forming a broken n-gram, the amino acids included in the n-gram are indicated, the position of the first amino acid from the n-gram in the pentapeptide molecule is indicated and the number of positions between each of the amino acids included in the n-gram. For example, for the pentapeptide DKLNV there is a second-order broken n-gram 12DN, where 1 is the position of the first amino acid, 2 is the number of positions between amino acids, DN is the list of amino acids included in the broken n-gram. For this pentapeptide there are only six broken n-grams of order 2, namely 11DL, 21KN, 31LV, 12DN, 22KV, 13DV. The study considered broken n-grams of order 2 only.

3. CLASSIFICATION ALGORITHM

To classify pentapeptides into stable and unstable, we used the gradient boosted decision trees algorithm [5]. The algorithm is built on the principle that a relatively weak machine learning algorithm can be strengthened by the same algorithm, which will "refine" the predictions of the previous algorithm based on its errors. When applying this principle to random forest classification, the first row of trees is trained on real data, predicting the class label for each object. The second row of trees is trained on the same data, but giving more weight to objects where the first row of trees made mistakes and correcting them. The trees of the third row are trained by correcting the errors of the trees of the second row, etc. Currently, gradient boosting over decision trees is one of the most popular machine learning algorithms, because at low training costs it provides high accuracy and protection from overfitting due to the fact that a random forest of decision trees is used. In this case, the features and subsample are mixed to construct a new tree. In addition, the obtained result is easy to interpret.

Quality control of training was carried out using the metric F_1, specified by the formula

$$F_1 = 2\frac{precision * recall}{precision + recall}.$$

In this case, one class is considered as a class of "positive objects", for example, a class of stable pentapeptides, and the other is a class of "negative objects". The *precision* metric determines the proportion of correctly identified positive objects among all objects classified as positive. The *recall* metric determines the proportion of correctly identified positive objects among all positive objects. Metric F_1 used to assess the quality of classification in the case of data in which the classes are significantly unbalanced.

The parameters of the classification algorithm were adjusted for each encoding method used using the cross-validation procedure in a high-dimensional space [6] using the *hyperopt* package. Table 1 shows the classification results achieved with the found settings.

Table 1. Results of classification of pentapeptide stability for different encoding methods

Encoding	Metrics		
	precision	*recall*	F_1
OHE	0.39	0.54	0.45
n-grams	0.39	0.41	0.40
broken n-grams	0.32	0.54	0.40

The best quality in terms of F_1 metrics is achieved when using OHE encoding. For n-gram and broken n-gram encodings, the quality is lower. This is explained by the small sample size and large number of features when encoding using n-grams and broken n-grams.

4. PROBABILISTIC ASSESSMENT OF THE SIGNIFICANCE OF AMINO ACID POSITIONS IN A PENTAPEPTIDE

In addition to assessing the quality of classification, it is of great interest to assess the importance of individual features in the stability of pentapeptides. To construct such an estimate using gradient boosting, the SHAP (SHapley Additive exPlanations) algorithm was used in this study [7], which allows you to estimate the probabilistic contribution of each combination of amino acids to the probability of classifying a pentapeptide as stable, taking into account the interaction of factors (amino acids and their positions) with each other. This method calculates the importance of a particular feature by comparing the results obtained with and without that feature. When constructing a classification rule in the form of a tree, the result can be influenced by the order in which the elements of the training set are used. To eliminate such an influence on the assessment of the importance of a feature, elements of the training sample are received for training many times in a random sequence.

The SHAP method was justified in the theory of cooperative games, when game participants can unite in coalitions to achieve the best result. The payoff of each player is equal to his average contribution to the total payoff over all coalitions under a random, equally probable ordering of the participants. This value is called the Shapley index [7] and is calculated by summing over all sets of features that do not include feature i, the weighted effect of using the excluded feature. In this case, the effect of using feature i is understood as the difference in the classification accuracy of a pentapeptide taking into account feature i and without taking it into account. The Shapley index is calculated using the formula

$$\Phi_i = \sum_{S \in F \setminus i} \frac{n_S!\,(n_F - n_S - 1)!}{n_F!} \left(f_{S \bigcup i} - f_S \right),$$

here F denotes the set of all possible sets of features, $F \setminus i$ denotes the set of sets of features that do not include the feature i, S – a set of features without feature i, $S \bigcup i$ – set of features S with the addition of feature i, f_S and $f_{S \bigcup i}$ – classification accuracy when using feature sets S and $S \bigcup i$, respectively, n_F and n_S – number of feature sets in the sets F and S, respectively. The significance of a feature is determined by the absolute value of the corresponding Shapley index.

5. INTERPRETATION OF RESULTS

Below are the results of interpretation using the SHAP method of the results of classifying the stability of pentapeptides by the gradient boosting algorithm using three different encodings.

5.1. Binary Encoding (OHE)

Table 2 presents an example of assessing the influence of the position of amino acids in the DRNAA pentapeptide on its stability. It is important to note that the stability of a pentapeptide is affected not only by the presence of an amino acid at any position, but also by its absence.

In Table 2 rows are ordered in order of decreasing probabilistic contribution of amino acids and their positions to the stability of the pentapeptide. Negative values indicate a negative impact on stability. It follows from the table that the presence of amino acid D in the first position in the fifth position increases the probability that the pentapeptide is stable, and the absence of amino

Table 2. Probabilistic contribution of amino acids and their positions on the stability of the DRNA pentapeptide

Amino acid		position	Probabilistic contribution
presence	absence		
D		1	0.048
R		2	0.018
	A	1	0.010
A		4	0.0040
A		5	–0.0083
N		3	–0.0096

acid A in the first position increases the probability that the pentapeptide is stable by only 1%. The presence of amino acid A at the last position reduces the probability of pentapeptide stability by 0.8%. It is assumed that the features influence the stability of the pentapeptide independently of each other.

If a similar probabilistic analysis is carried out for a variety of pentapeptides, the overall result can be presented in the form of a diagram of the probabilistic contributions of amino acids and their positions to stability. Figure 1 shows a diagram for the most significant features. Due to the great computational difficulties associated with the need to solve the classification problem for all possible sets of features, calculations were carried out for 1000 randomly selected pentapeptides. In the diagram, a single point corresponds to the result of an analysis of a single pentapeptide. The presence of a feature (the presence of an amino acid at a specified location) is represented by an open symbol, and its absence is represented by a closed symbol.

The figure shows that if the pentapeptide contains amino acid K at the fifth position, it has the greatest positive effect on its stability. The opposite effect is that amino acid A in the first position has a negative effect on stability.

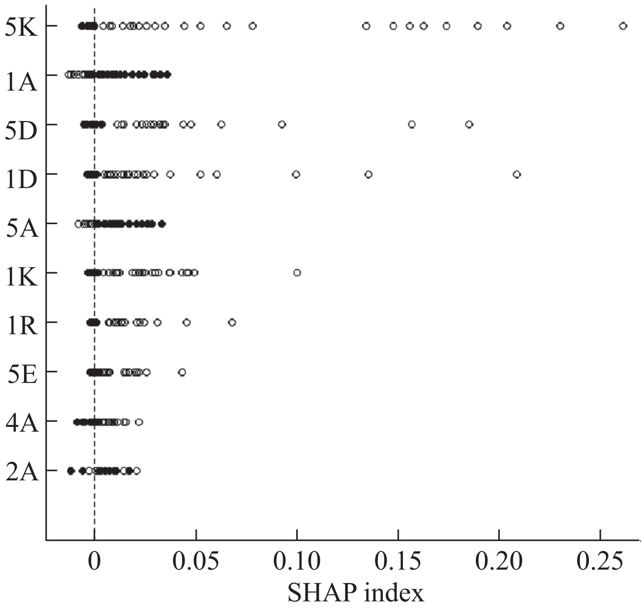


Fig. 1. Diagram of the probabilistic contributions of features to the stability of 1000 randomly selected pentapeptides under OHE encoding, constructed using the SHAP algorithm.

5.2. Continuous String Representation

When encoding using n-grams, the estimate of the probabilistic contribution to the stability of an individual feature turns out to be less than when encoding OHE. This is a consequence of the fact that when using n-grams up to the third order, the number of features is 256 times greater than with ONE encoding. Table 3 shows examples of estimates of the probabilistic contribution to the stability of the DRNAA peptide.

In Table 3 rows are ordered in order of decreasing probabilistic contribution of amino acids and their positions to the stability of the pentapeptide. The table shows that when encoding using n-grams, the joint contribution of amino acids D and R located in the first and second positions to the stability of the pentapeptide is estimated at about 0.5%, whereas with OHE coding the estimate is 6%.

Table 3. Examples of assessing the probabilistic contribution to the stability of the DRNAA pentapeptide when encoded using n-grams

Combination of amino acids		position	Probabilistic contribution
presence	absence		
D		1	0.0030
R		2	0.0022
	K	2	–0.00004
	EK	1	–0.00008
	T	5	–0.000028
	R	5	–0.000029
A		5	–0.0001

Figure 2 shows an example diagram of the probabilistic contributions of amino acids and their positions to the stability of 1000 randomly selected pentapeptides. The figure shows that single combinations of amino acids have the greatest significance; amino acid K in the fifth position has the most powerful positive effect, and amino acid A in the first position has the most negative effect.

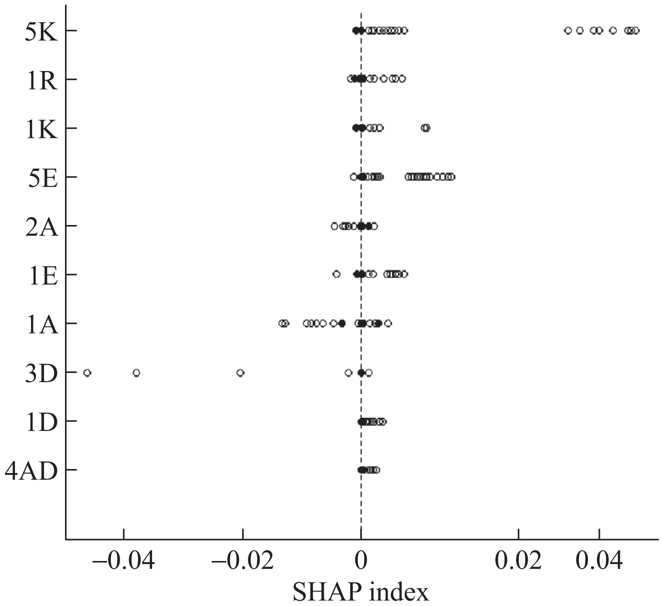


Fig. 2. Diagram of the probabilistic contributions of features to the stability of 1000 randomly selected pentapeptides when encoded using n-grams, constructed using the SHAP algorithm.

5.3. Discontinuous String Representation

Table 4 presents the result of estimating the probabilistic contribution to the stability of an individual feature using the example of the DRNAA pentapeptide when encoded with a discontinuous n-gram. Figure 3 shows an example of a diagram of the probabilistic contributions of amino acids and their positions to the stability of 1000 randomly selected pentapeptides with the same encoding.

Table 4. Examples of assessing the probabilistic contribution to the stability of the DRNAA pentapeptide when encoded with a broken n-grams

Combination of amino acids		position	Probabilistic contribution
presence	absence		
R–A		2	0.0093
R–A		2	0.0041
	A–A	1	0.0031
D–A		1	0.0020
	A–A	1	0.0013
	A–K	2	–0.0014
D–N		1	–0.0030

In Table 4 rows are ordered as the probabilistic contribution of the combination of amino acids and their positions to the stability of the pentapeptide decreases. It follows from the table that the greatest effect on the stability of the DRNAA pentapeptide is exerted by the combination of amino acids R in the second position and A in the fourth or fifth position. The absence of amino acid A in the first position and simultaneously in the fourth or fifth positions also increases the likelihood of stability of the DRNAA pentapeptide, but to a lesser extent.

Figure 3 shows that combinations with amino acid A in the second and K in the fifth position have the greatest significance for stability. The presence of two amino acids A in a pentapeptide with two or three gaps between them, on the contrary, is a sign of its instability.

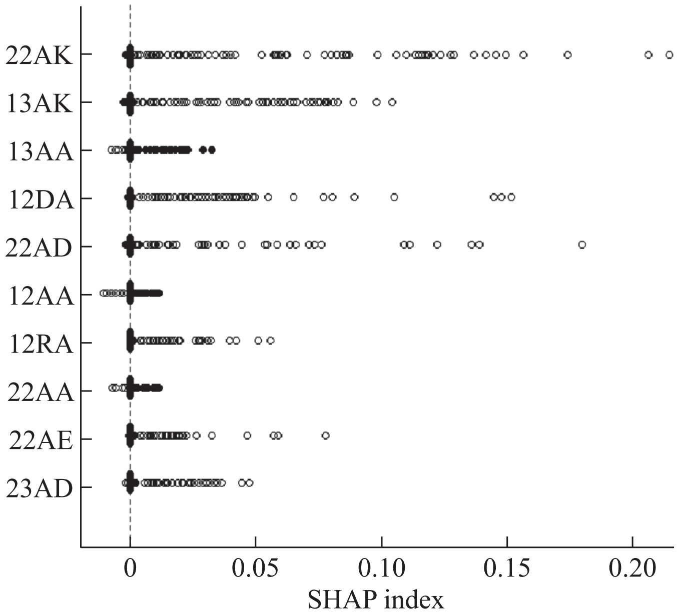


Fig. 3. Diagram of the probabilistic contributions of features to the stability of 1000 randomly selected pentapeptides when encoded with a broken n-gram, constructed using the SHAP algorithm.

6. CONCLUSION

The article discusses the result of using three different encodings of the pentapeptide structure when predicting its stability through solving the classification problem. Binary encoding (One Hot Encoding), continuous string representation (n-gram), and discontinuous string representation (broken n-gram) were considered. Each of the encodings generates feature spaces of different dimensions: 100 when using the OHE binary encoding, 25 600 when encoding using n-grams no higher than the third order, and 10 400 when using a discontinuous n-gram. This creates varying degrees of data sparsity. The problem of classifying pentapeptides into stable and unstable was solved by the gradient boosting method (LGBM). The study used a set of 49 745 pentapeptides, of which 3.43% were stable. The data was randomly divided into "training", "testing" and "validation" sets in proportions while maintaining the original class balance. After training, the results of testing on the control sample for each of the encodings showed approximately the same value for the quality metric F_1, equal to 0.45 for binary encoding and 0.40 when using different n-grams.

Assessing the importance of features for predicting the stability of pentapeptides highlighted the most important features. Each encoding method has its own characteristics. OHE coding evaluates the importance of the location of a particular amino acid at a particular position. When using n-grams, encoding allows one to evaluate the importance of the combination of amino acids at adjacent positions, and when using broken n-grams, the importance of the arrangement of amino acids at positions distant from each other is assessed. Encoding using broken n-grams makes it possible to highlight the effect of the influence of a combination of amino acids located in different positions of the pentapeptide molecule.

The question of the structural stability of pentapeptides was considered in [8]. In this work, a problem dimensionality reduction method based on the calculation of mutual information between the stability feature and the pentapeptide description was used in the binary OHE encoding. It turned out that dimensionality reduction using mutual information allows one to use the "simple" K-nearest neighbors classification method for stability prediction. At the same time, the quality of the result in terms of the "accuracy" and "completeness" metrics practically coincides with the result of using the "random forest" method, which requires significantly greater computational and time costs. A probabilistic assessment of the effect of pentapeptide composition on its stability was not performed in that study. In this work, the emphasis was placed on assessing the influence of the pentapeptide composition and the results of such an assessment are presented for 1000 randomly selected pentapeptides, which is associated with large requirements for the necessary computing power.

REFERENCES

1. Senior, A.W., Evans, R., Jumper J., et al., Improved Protein Structure Prediction Using Potentials from Deep Learning, *Nature*, 2020, vol. 577, pp. 706–710.

2. Pereira, J., Simpkin, A.J., Hartmann, M.D., et al., High Accuracy Protein Structure Prediction in CASP14, *Proteins Structure Function and Bioinformatics*, 2021, vol. 89, no. 12, pp. 1687–1699. https://doi.org/10.1002/prot.26171

3. Nekrasov, A.N., Kozmin, Yu.P., Kozyrev, S.V., et al., Hierarchical Structure of Protein Sequence, *Int. J. Mol. Sci.*, 2021, vol. 22, no. 15, 8339. https://doi.org/10.3390/ijms22158339

4. Anashkina, A.A., Nekrasov, A.N., Alekseeva, L.G., et al., A Minimum Set of Stable Blocks for Rational Design of Polypeptide Chains, *Biochimie*, 2019, vol. 160, pp. 88–92.

5. Ke, G., Meng, Q., Finley, T., Wang, T., et al., A Highly Efficient Gradient Boosting Decision Tree, *Proc. 31st Conference on Neural Information Processing Systems (NIPS). Long Beach*, 2017, pp. 3149–3157.

6. Bergstra, J., Yamins, D., and Cox, D.D., Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures, *Proc. of the 30th International Conference on Machine Learning (ICML)*, 2013. pp. 115–123.

7. Lundberg, S.M. and Lee, S.I., A Unified Approach to Interpreting Model Predictions, *Proc. 31st Conference on Neural Information Processing Systems (NIPS). Long Beach*, 2017, pp. 4765–4774.

8. Mikhalskii, A.I., Petrov, I.V., Tsurko, V.V., Anashkina, A.A., et al., Application of Mutual Information Estimation for Prediction the Structural Stability of Pentapeptides, *Rus. J. Numer. Anal. Math. Model.*, 2020, vol. 35, no. 5, pp. 263–271.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1437–1448.

Models and Methods for Checking the Attainability of Goals and Feasibility of Plans in Large-Scale Systems Using the Example of Goals and Plans for Elimination of the Consequences of Flood

A. D. Tsvirkun[*,a], **A. F. Rezchikov**[*,b], **V. A. Kushnikov**[**,c], **O. I. Dranko**[*,d], **A. S. Bogomolov**[**,e], **and A. D. Selyutin**[**,f]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
[**] *Federal State Budgetary Institution Federal Research Centre*
"Saratov Scientific Centre of the Russian Academy of Sciences", Saratov, Russia
e-mail: [a] *tsvirkun@ipu.ru,* [b] *rw4cy@mail.ru,* [c] *kushnikoff@yandex.ru,* [d] *olegdranko@gmail.com,*
[e] *alexbogomolov@yandex.ru,* [f] *aseliutin@ya99.ru*

Received May 26, 2023
Revised September 18, 2023
Accepted September 30, 2023

Abstract—Models and methods have been developed to verify the achievability of goals and the feasibility of plans implemented when managing large-scale systems in their development. An algorithm for analyzing the achievability of a set of goals and plans implemented when managing these systems is proposed and justified. Statements and hypotheses that make it possible to machine-check the feasibility of plans have been generated. A model example is given that confirms the possibility of checking the feasibility of plans for eliminating the consequences of a flood using the developed models and methods. In managing large-scale systems development, it is advisable to use control loops that check the achievability of set goals and the feasibility of plans over a selected time interval. In the absence of this verification, the chosen trajectory of development of a large-scale system at specific points in time may turn out to be unrealizable, which will lead to disruption of the work being carried out, as well as to significant costs of the human, financial, technical and other types of resources for the implementation of obviously impracticable plans.

Keywords: autonomous goal setting, achievability of goals, feasibility of structurally complex plans, system dynamics

DOI: 10.25728/arcRAS.2024.59.53.001

1. INTRODUCTION

Considerable attention is paid to the study of models and methods for forming and testing the achievability of set goals when managing complex human-machine, social, economic, and biological systems. Currently, checking the achievability of a set of goals carried out when planning, designing, and managing large-scale systems needs to be formalized more and is carried out mainly using the intuition and experience of decision-makers. Characteristics of targets at various levels of the hierarchy, as well as indicators of their implementation, can change significantly over time intervals for achieving goals, which complicates the activities of decision-makers in designing and managing large-scale systems and planning the results of their activities. Issues of formalizing the goal-setting procedure were discussed in the works of domestic and foreign researchers [1, p. 5]. The results obtained in this area of research, however, have not yet led to the creation of a holistic set of models and methods that make it possible to verify the achievability of the goals of large-scale systems. The lack of these developments, as well as specialized mathematical and information software designed to verify the achievability of the goals of large-scale systems, as well as plans for their creation and development, understood as a series of actions combined sequentially to achieve a goal with possible deadlines, causes difficulties in the development and management of human-machine, economic, social objects [2, p. 252].

The article is devoted to developing of new tasks, models, and methods for checking the achievability of goals and the feasibility of plans in large-scale systems.

2. MAIN AREAS OF RESEARCH

We develop the following models and methods for testing the achievability of goals and the feasibility of plans in large-scale systems:

1. To develop a methodological basis for creating an intelligent system that allows anyone to predict, identify, and prevent events that lead to the impracticability of action plans based on the mathematical apparatus of system dynamics, probabilistic safety analysis, and the theory of Bayesian networks [3, p. 168].

2. To formulate and justify a general approach to checking the feasibility of structurally complex plans, which involves analyzing feasibility using the apparatus of Boolean functions, Bayesian networks, and knowledge representation models of intelligent systems, as well as using the system dynamic approach and system dynamics equations [4, p. 21].

3. To develop models and methods for an intelligent decision support system designed to analyze the feasibility of structurally complex action plans using logical probabilistic models, Bayesian networks, a system dynamic approach, the mathematical apparatus of probabilistic safety analysis, and the theory of deep neural networks.

4. Develop methods that allow anyone to present the plan being tested in the form of a hierarchical cause-and-effect model and generate indicators of its feasibility.

5. To create models and methods for checking quickly the achievability of goals and the feasibility of plans using the apparatus of dynamic graphs and knowledge representation models, characterized by the ability to analyze plans over long time intervals in dynamics, which will allow timely changes to plans for large-scale systems when their impracticability occurs.

6. To develop problem statements, models, and methods to test the feasibility of a structurally complex action plan using a system dynamic approach. The action plan is presented as a cause-and-effect network of events. The modeled variables are indicators characterizing the implementation of individual plan activities. The arcs are the cause-and-effect relationships that exist between these indicators. A system of differential equations is written, and the initial conditions corresponding to the desired values of the indicated indicators are determined. A verifiable structurally complex plan is feasible if the generated system of equations under the selected initial conditions has the solution in a given range.

7. To propose and justify methods for conducting computational experiments that characterize the possibilities of using the developed mathematical software when testing the feasibility of a structurally complex action plan for domestic energy development.

8. To create and test a problem-oriented intelligent decision support system that implements the main results of this research.

3. FORMULATION OF THE PROBLEM

The formulation of the problem of checking the feasibility of plans for the functioning of industrial enterprises and organizations in a substantive and formal form is given in [5–8]. In those papers, this statement is extended to the goals of large-scale systems and plans for their implementation.

It has the following formulation: models and methods for checking the achievability of goals and the feasibility of plans used in large-scale systems at various time intervals during their creation and operation; identify possible reasons that impede the solution of this problem and suggest ways to eliminate them. Solving this problem will make it possible to create a methodological basis for the development of intelligent goal-setting systems, the use of which in managing large-scale complexes will significantly increase the efficiency of their functioning. Some approaches to managing development plans and programs are shown in [9–12].

4. GENERAL APPROACH TO THE SOLUTION

Imagine the plan being tested as the tree containing conjunctive and disjunctive vertices. The possibility of such a transformation follows from its hierarchical structure (conjunctive vertices) and the conditions for implementing individual activities $M_i \in \{M_1, \ldots, M_n\}$. Each vertex of the graph G lets us match the variable g_i, $i = \overline{1,n}$, which takes the value 1 when the event is performed M_i and the value is 0 if it is not executed. The developed solution method is based on the following statements.

Statement 1. *At the moment in time $t_0 \in [t_h, t_k]$, plan $P(\vec{x}, \vec{u}) \in \{P(\vec{x}, \vec{u})\}$ is impossible if the output is at least one chain of conjunctive and/or disjunctive vertices connecting any terminal vertex connecting any terminal vertex of the tree G with its root vertex, so the relation is valid $g_i = 0$.*

The description of this statement follows from the fact that if at least one plan event that is part of the conjunctive chain is not completed, then the entire plan will not be completed since otherwise, the vertex corresponding to this event must be excluded from the graph as not affecting the execution plan. This statement is a necessary condition for satisfiability $P(\vec{x}, \vec{u}) \in \{P(\vec{x}, \vec{u})\}$, asserting that for the plan to be feasible, each activity must be completed $M_j \in \{M_1, \ldots, M_n\}$, without which the corresponding event $M_j \in \{M_1, \ldots, M_n\}$ is impossible.

Statement 2. *Let us assume that at the moment of time $t_0 \in [t_h, t_k]$ there are unfulfilled activities $M_j \in \{M_1, \ldots, M_n\}$, included in tree chain composition G, connecting the root vertex to the terminal ones. Then, the plan $P(\vec{x}, \vec{u}) \in \{P(\vec{x}, \vec{u})\}$ will not be feasible at a given time if there is at least one tree section G, with the output of a conjunctive-disjunctive chain $g_i = 0$.*

In the design and management of large-scale systems, a goal or a plan developed to achieve it is considered achieved if the requirements of the above statements are met. The problem with using this approach to checking the feasibility of goals is that the boundaries of the numerical range are determined by the decision maker, usually based on experience and intuition based on opportunistic considerations using largely incomplete and subjective ideas about the system, time-varying cause-and-effect relationships, existing between individual indicators, with insufficient consideration of the influence of environmental disturbances, etc. All this leads to the fact that plan $P(\vec{x}, \vec{u}) \in \{P(\vec{x}, \vec{u})\}$ may not be feasible at specific points in time, the occurrence of which is complicated to predict in advance.

The hypothesis forms a sufficient condition for checking the plan when using a system of indicators. It allows anyone to present the main stages of checking the achievability of goals and the feasibility of plans in the diagram (Fig. 1).

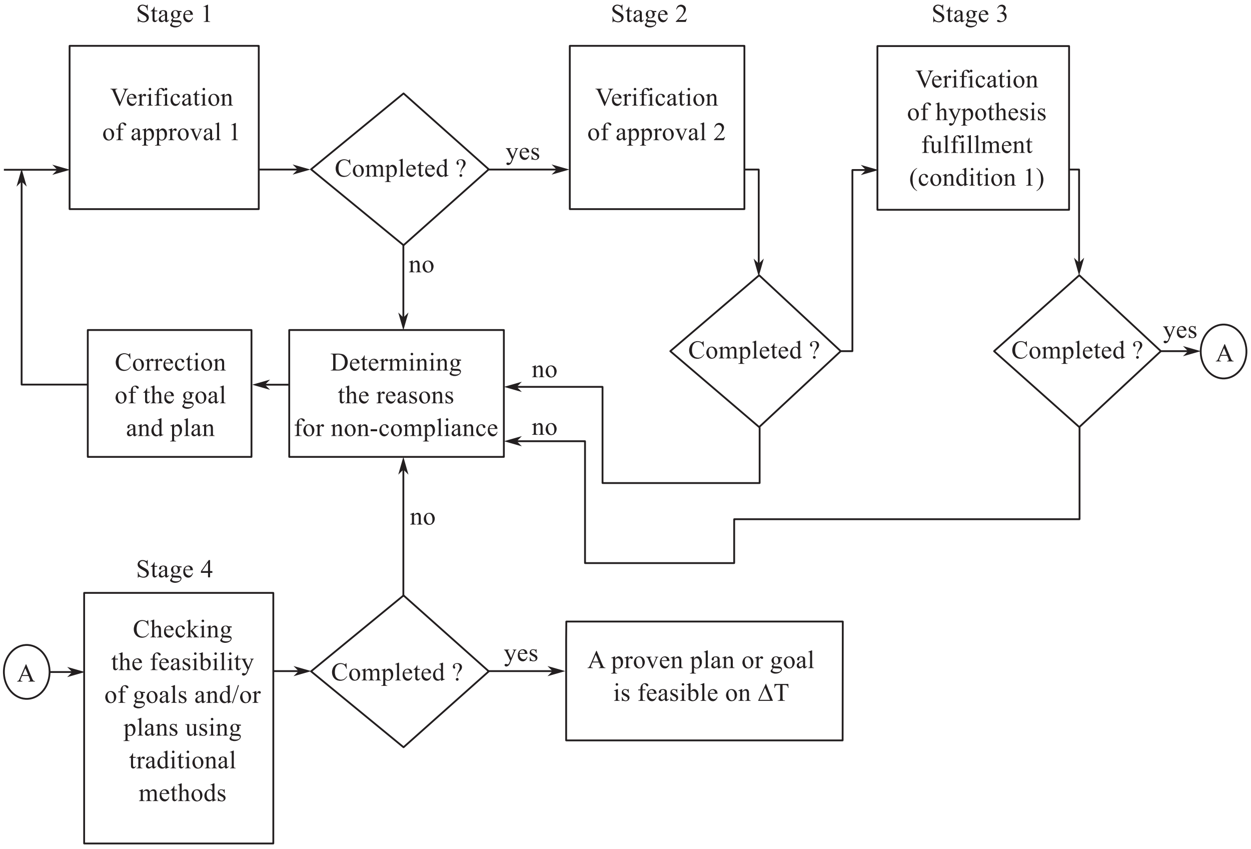

Stage 1
Stage 2
Stage 3
Verification of approval 1
Completed ?
yes
Verification of approval 2
Verification of hypothesis fulfillment (condition 1)
no
Correction of the goal and plan
Determining the reasons for non-compliance
no
no
Completed ?
Completed ?
yes
A
no
Stage 4
A
Checking the feasibility of goals and/or plans using traditional methods
Completed ?
yes
A proven plan or goal is feasible on ΔT

Fig. 1. Main stages of checking the achievability of goals and feasibility of plans.

Proposition 1. *Plan $P(\vec{x},\vec{u}) \in \{P(\vec{x},\vec{u})\}$ will be achieved within a given time interval ΔT, if a system of indicators of its achievability is known $l_1,\ldots,l_m$, for which the following is performed:*

$$\exists t_i \in \Delta T : \min l_i \leqslant l_i \leqslant \max l_i, \quad i = 1,\ldots,m, \tag{1}$$

where $\min l_i$, $\max l_i$ *are the lower and upper limits of indicator changes* l_i*;* m *is a known constant.*

Decision makers widely use this hypothesis when checking the achievability of goals at various levels of the management hierarchy of large-scale systems. This circumstance confirms the possibility of its use in the development of a computer system for verifying the achievability of goals and the feasibility of plans.

In the first stage, the fulfillment of statement (1) is checked; if in the conjunctive chains connecting the terminal vertices with the terminal one, at least one event has not been completed, then the goal or plan is impossible and requires correction. In the second stage, the requirements for goals are checked and plans, excluding the possibility of approving a plan that is not feasible due to an unfavorable combination of events (Statement 1). In the third stage of verification, it is determined whether all indicator values $t_0 \in [t_h, t_k]$ are within the acceptable range for all periods all $l_1(t_0),\ldots,l_m(t_0)$. When checking this condition, the apparatus of system dynamics is used since the indicators are influenced by a large number of linear and nonlinear feedbacks, as well as time-varying environmental disturbances. In the fourth stage, mathematical tools are used to check plans' achievability and goals' feasibility.

5. FEASIBILITY CHECKING ALGORITHM OF THE ACTION PLAN

Algorithm 1.

1. The beginning of the algorithm.

2. Set the vertex u^* on the graph $G^*(U, E)$, having zero half-degree of approach. This vertex corresponds to the vertices M_1 or Z_1, characterizing the implementation of the action plan or the achievement of the general goal, respectively.

3. Set all vertices $u_{m_0}, u_{k_0}, \ldots, u_{l_0} \in U$ on the graph $G^*(U, E)$ with incidental u^*. Add the first condition into the emerging product model F: Plan M or goal Z_1 will be carried out when carrying out activities or goals corresponding to the vertices $u_{m_0}, u_{k_0}, \ldots, u_{l_0}$.

4. Continue step 2 until the vertices reach $G^*(U, E)$ with zero half-degree of outcome. Thus, build the production model F ultimately.

5. Match the production system with a logical function $f(u_{1k}, \ldots, u_{vk})$, which takes the value 0 if the plan has not been completed, or 1 otherwise.

6. Construct a digital discrete device circuit DU, to determine the values $f(u_{1k}, \ldots, u_{vk})$, and the function indicators $f_{ind}(C, C_1)$, which characterize the degree of implementation of the action plan or achievement of the general goal.

7. Submit at the entrance DU binary signals characterizing the fulfillment or non-fulfillment of individual activities of the plan under review or the goals of the analyzed target structure. At $f_{ind}(C, C_1) = 1$ the plan is feasible or the goal is achievable; if $f_{ind}(C, C_1) = 0$, then it is necessary to correct them.

8. Moving along the branches of propagation of zero signals of the device DU, determine the reasons for the plan's impracticability or the goal's unattainability and report them to the decision maker.

9. Determine whether the conditions for using the Kolmogorov–Chapman equations are met to calculate the probability of failure to complete the action plan or the unattainability of the goal. If not, then go to step 10.

10. Determine the minimum sections L_i, $i = 1, \ldots, m$, characterizing failure to implement a plan or achieve a goal due to a combination of unfavorable circumstances. Each of the minimum sections characterizes one of the combinations of relatively unimportant events, which in their totality lead to the goal's unattainability and the plan's impracticability.

11. Solving the system of linear homogeneous Kolmogorov–Chapman equations for each minimal section, determine the probability of an unfavorable combination of circumstances occurring P_i, $i = 1, \ldots, m$.

12. If $P_i \geqslant \varepsilon$, $i = 1, \ldots, m$, then issue a message about the high probability of unattainability of the goal and impracticability of the plan due to an unfavorable combination of event L_i, $i = 1, \ldots, m$, issue recommendations to the decision maker, change the goal or plan being checked and proceed to step 7.

13. Select an indicator system I_i, $i = 1, \ldots, h$, which characterizes the feasibility of the plan being tested or the achievability of the goal. Identify relevant relationships between indicators, which can be linear or nonlinear. Determine environmental disturbances affecting the indicators.

14. Determine the limit values of indicators I_i^*, $i = 1, \ldots, h$, the achievement of which means the feasibility of the plan being verified or the feasibility of the set goal.

15. Create a system of nonlinear differential equations in the normal Cauchy form, characterizing the change in the system of indicators over time, considering their mutual influence and the impact of environmental disturbances.

16. Solve a system of equations using one of the numerical methods under given initial conditions. If the solutions obtained go beyond the area limited I_i^*, $i = 1, \dots, h$, then issue a message to the decision maker, recommend actions to eliminate the discrepancy, and proceed to step 10.

17. Message the decision maker that the check did not reveal the goal's unattainability or the plan's impracticability.

18. End of the algorithm.

6. GOAL ACHIEVABILITY CHECKING AND THE ACTION PLAN FEASIBILITY

Let us consider the features of implementing individual stages of checking the achievability of goals and the feasibility of plans for a large-scale system using a plan to eliminate the consequences of floods and floods [13–16]. The problem statement has the following formulation:

Task 1. Develop formal models and algorithms that allow, on a time interval $t \in [t_0; t_N]$ determine whether goal attainability indicators are missing $X_i(t, a(t), p(t))$, $i = 1, \dots, n$ beyond specified limits: $X_i(t, a(t), p(t)) \geqslant X_i^{\min}$, $i = 1, \dots, n$. If this condition is not satisfied for at least one indicator, then the plan is considered unfeasible due to the inability to achieve the required value of this indicator.

The indicator values are determined by solving a system of nonlinear differential equations $\frac{dX_i(t,p(t),a(t))}{dt} = f(t, a(t), X_1(t, p(t)), \dots, X_n(t, p(t)))$, $i = 1, \dots, n$ at $t > 0$, $0 < X_i(t, a(t), p(t)) \leqslant M_{\max}^{X_i}$, $i = 1, \dots, n$, where X_i^* are recommended values for characteristics of flood consequences, $X_i(t, a(t), p(t))$, $i = 1, \dots, n$ are characteristics of the consequences of flooding, affecting the amount of damage, γ_i are the weight coefficient of the characteristic, $a(t)$ is the vector of environmental parameters.

6.1. Mathematical Model

The following system of first-order nonlinear differential equations describes the mathematical model.

$$\frac{dX_i(t, a(t), p(t))}{dt} = f_i^+(F_1, \dots, F_m) - f_i^-(F_1, \dots, F_m), \quad i = 1, \dots, n, \tag{2}$$

where f_i^+, f_i^-, $i = 1, \dots, n$ – rates, continuous or piecewise continuous functions that determine the positive and negative rate of change in the value of a system variable $X_i(t, a(t), p(t))$, $i = 1, \dots, n$. Functions f_i^+, f_i^-, $i = 1, \dots, n$ are functions of factors F_j, $j = 1, \dots, m$, wherein F_j may be system variables or environmental parameters.

The directed cause-and-effect graph shows the relationships between model variables (Fig. 2).

The functions on the right side of (3) have the form

$$f_i^{+/-}(F_1, \dots, F_n) = \sum_{l=1}^{n} k_{i,l}^{+/-} \prod_{j=1}^{n} f_{i,l}^{F_j}(F_j),$$

where coefficients $k_{i,l}^{+/-}$, $i = 1, \dots, 12$ are determined at the stage of adapting the model to the object of study. Let us also assume that the coefficients $k_{i,l} = 0$, $l = 1, \dots, m-1$, $k_{i,l} \neq 0$, $l = m$, $k_{i,l} = 0$, $l = m+1, \dots, n$. Then this expression will take the form $f_i^{+/-}(F_1, \dots, F_n) = k_i^{+/-} \prod\limits_{j=1}^{n} f_i^{F_j}(F_j)$.

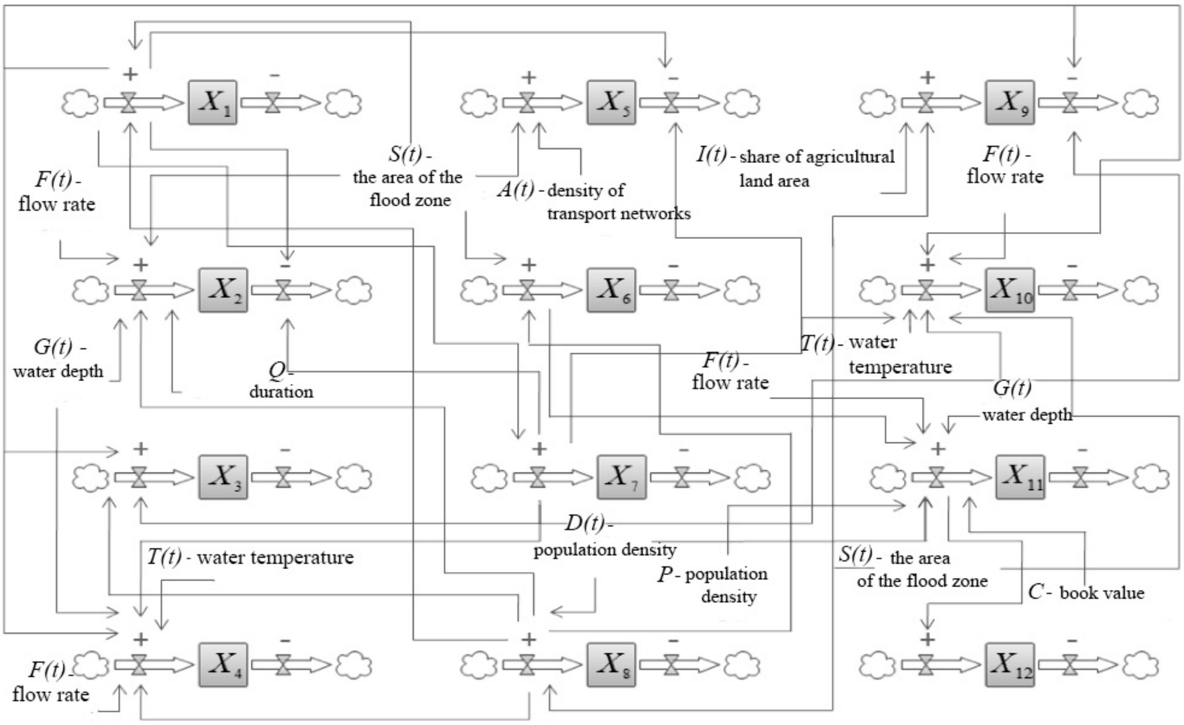


Fig. 2. Cause-and-effect relationships between model variables.

The developed mathematical model will have a general form based on the analysis of the graph of cause-and-effect relationships:

$$\begin{cases}
\dfrac{dX_1(t)}{dt} = k_1^+ f_1^S(S(t)) f_1^{X_8}(X_8(t)), \\
\dfrac{dX_2(t)}{dt} = k_2^+ F(t)G(t)t f_2^S(S(t)) f_2^{X_8}(X_8(t)) - k_2^- f_2^{X_1}(X_1(t)) f_2^{X_7}(X_7(t)), \\
\dfrac{dX_3(t)}{dt} = k_3^+ f_3^{X_8}(X_8(t)) f_3^{X_1}(X_1(t)) f_3^{X_7}(X_7(t)), \\
\dfrac{dX_4(t)}{dt} = k_4^+ F(t)G(t)T(t) f_4^{X_8}(X_8(t)) f_4^{X_7}(X_7(t)) f_4^{X_1}(X_1(t)), X_1(t)), \\
\dfrac{dX_5(t)}{dt} = k_5^+ A(t) f_5^S(S(t)) - k_5^- f_5^{X_1}(X_1(t)) f_5^{X_7}(X_7(t)), \\
\dfrac{dX_6(t)}{dt} = k_6^+ f_6^S(S(t)) f_6^{X_8}(X_8(t)), \\
\dfrac{dX_7(t)}{dt} = k_7^+ f_7^{X_1}(X_1(t)), \\
\dfrac{dX_8(t)}{dt} = k_8^+ D(t) f_8^S(S(t)) - k_8^- f_8^{X_4}(X_4), \\
\dfrac{dX_9(t)}{dt} = k_9^+ I(t) f_9^S(S(t)) - k_9^- f_9^{X_1}(X_1(t)) f_9^{X_7}(X_7(t)), \\
\dfrac{dX_{10}(t)}{dt} = k_{10}^+ F(t)G(t)T(t) f_{10}^S(S(t)) f_{10}^{X_1}(X_1(t)) f_{10}^{X_7}(X_7(t)), \\
\dfrac{dX_{11}(t)}{dt} = k_{11}^+ PCF(t)G(t)D(t) f_{11}^S(S(t)) f_{11}^{X_6}(X_6(t)), \\
\dfrac{dX_{12}(t)}{dt} = k_{12}^+ f_{12}^{X_{11}}(X_{11}(t)),
\end{cases} \tag{3}$$

where $f_j^{X_i}$ – functional dependence of the system variable on $X_j(t)$ from X_i, and f_j^S – dependence of the X_j from $S(t)$, $i, j = 1, \ldots, 12$. If such dependencies are unknown, they can be determined based on statistical data by experts or software developers.

The mathematical model will take the following form, taking into account the polynomials of auxiliary dependencies:

$$\left\{\begin{aligned}
&\frac{dX_1(t)}{dt} = \frac{1}{X_1^{\max}}(k_1^+(0.001S^3(t) - 0.04S^2(t) + 0.6S(t) - 2.1)\\
&\times(54X_8^4(t) - 137X_8^3(t) + 103.4X_8^2(t) - 20.7X_8(t) + 1.2)),\\
&\frac{dX_2(t)}{dt} = \frac{1}{X_2^{\max}}(kt(-0.02S^3(t) + 0.64S^2(t) - 6.4S(t) + 21)\\
&\times(-14.5X_8^2(t) + 22.5X_8(t) - 3.3) - k_2^-(0.57X_1^2(t) + 0.276X_1(t) + 0.05)\\
&\times(-3.3X_7^2(t) + 5.6X_7(t) - 0.13)),\\
&\frac{dX_3(t)}{dt} = \frac{1}{X_3^{\max}}(k_3^+(3.28X_8^2(t) - 23.31X_8(t) + 12.3)\\
&\times(-1.26X_1^2(t) + 10.1X_1(t) - 17.8)(-0.33X_7^2 + 2.2X_7 - 0.26)),\\
&\frac{dX_4(t)}{dt} = \frac{1}{X_4^{\max}}(k_4^+F(t)G(t)T(t)(-1.3X_8^4(t) + 1.92X_8^3(t) - 0.95X_8^2(t)\\
&+0.3X_8(t) + 0.7)(-0.42X_7^4(t) - 7.19X_7^3(t) + 19.34X_7^2(t) - 15.1X_7(t) + 4.435)\\
&\times(X_1^3(t) - X_1^2(t) + 1.5X_1(t) + 0.02)),\\
&\frac{dX_5(t)}{dt} = \frac{1}{X_5^{\max}}(k_5^+A(t)(0.01S^2(t) - 0.1S(t) + 0.5) - k_5^-(0.217X_1^2(t)\\
&-0.505X_1(t) + 0.3(-0.304X_7^2(t) + 1.1X_7(t) + 0.26)),\\
&\frac{dX_6(t)}{dt} = \frac{1}{X_6^{\max}}(k_6^+(0.002S^2(t) + 0.056S(t) + 0.48)(-0.05X_8^3(t)\\
&+0.9X_8^2(t) - 0.02X_8(t) + 0.23)),\\
&\frac{dX_7(t)}{dt} = \frac{1}{X_7^{\max}}(k_7^+(3.5X_1^3(t) - 5.3X_1^2(t) + 3.27X_1(t) + 0.0003)),\\
&\frac{dX_8(t)}{dt} = \frac{1}{X_8^{\max}}(k_8^+D(t)(0.18S^3(t) - 0.06S^2(t) + 0.77S(t) - 1.77)\\
&-k_8^-(2.17X_4^2(t) - 0.0024X_4(t) + 0.16)),\\
&\frac{dX_9(t)}{dt} = \frac{1}{X_9^{\max}}(k_9^+I(t)(0.002S^2(t) + 0.07S(t) + 0.5) - k_9^-(0.43X_1^3(t) - 2.3X_1^2(t)\\
&+3.2X_1(t) - 0.07)(1.15X_7^3(t) - 1.78X_7^2(t) + 0.93X_7(t) - 0.024)),\\
&\frac{dX_{10}(t)}{dt} = \frac{1}{X_{10}^{\max}}(k_{10}^+F(t)G(t)T(t)(-0.0007S^4(t) + 0.03S^3(t) - 0.46S^2(t)\\
&+2S(t) - 0.4)(0.25X_1^3(t) - 1.24X_1^2(t) + 2.04X_1(t) - 0.049)\\
&\times(10.9X_7^3(t) - 26.57X_7^2(t) + 16.7X_7(t) - 0.515)),\\
&\frac{dX_{11}(t)}{dt} = \frac{1}{X_{11}^{\max}}(k_{11}^+PCF(t)G(t)D(t)(-0.0005S^3(t) + 0.02S^2(t) - 0.01S(t) + 0.4)\\
&\times(-3.5X_6^3(t) + 7.8X_6^2(t) - 2.7X_6(t) + 0.25)),\\
&\frac{dX_{12}(t)}{dt} = \frac{1}{X_{12}^{\max}}(k_{12}^+(-45.3X_{11}^4(t) + 111.95X_{11}^3(t) - 84.07X_{11}^2(t) + 20.04)),
\end{aligned}\right. \tag{4}$$

where $t_0 = 1,\ X_i(t_0) = X_{i0},\ i = 1, \ldots, 12.$

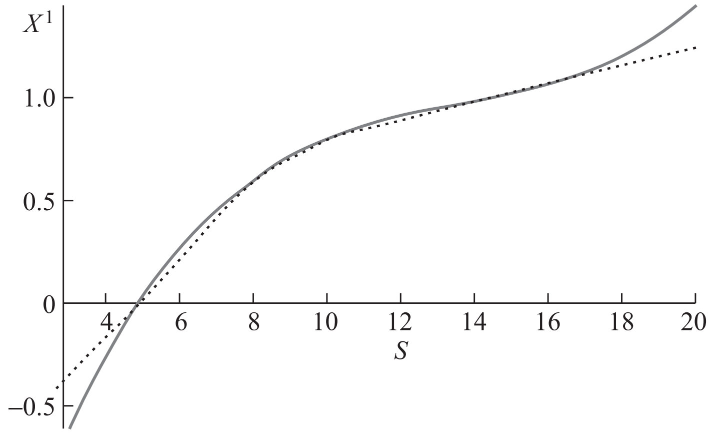


Fig. 3. Graphs of piecewise linear function and polynomial for f_1^S.

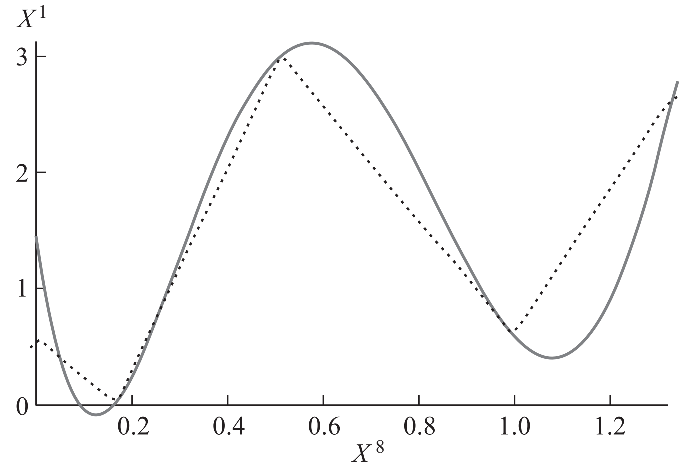


Fig. 4. Graphs of piecewise linear function and polynomial for $f_1^{X^8}$.

The system of differential equations (4) is a Cauchy problem; it can be solved by one of the numerical methods. The simulated characteristics of the system were normalized relative to the maximum values for the convenience of presenting the obtained results.

In particular, auxiliary dependencies $f_j^{X_i}$ and f_j^S take the following form for the case of the flood in the Primorsky region in 2001.

On Figs. 3 and 4 the constructed polynomials are presented $f_1^S = 0.001S^3(t) - 0.04S^2(t)+ 0.6S(t) - 2.1$ and $f_1^{X_8} = 54X_8^4(t) - 137X_8^3(t) + 103.4X_8^2(t) - 20.7X_8(t) + 1.9$ for functional dependencies f_1^S and $f_1^{X^8}$.

The developed mathematical models make it possible to solve problem (1), as shown in the description of the model example.

7. MODEL EXAMPLE

Let us check the constructed model to check the plan's feasibility for eliminating the consequences of floods and floods.

The plan has been developed to reduce losses from their occurrence in various regions of Russia, the upper level of which is shown in Fig. 5.

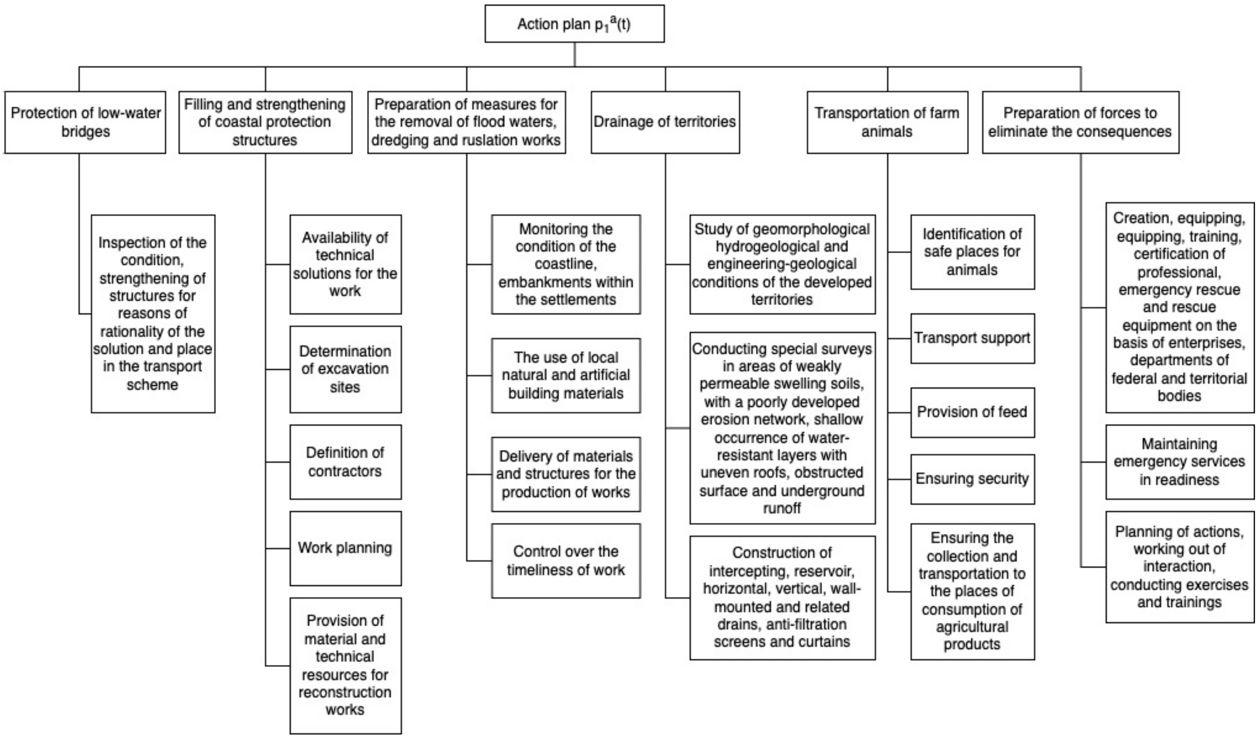

Action plan p1a(t)
Protection of low-water bridges
Inspection of the condition, strengthening of structures for reasons of rationality of the solution and place in the transport scheme
Filling and strengthening of coastal protection structures
Availability of technical solutions for the work
Determination of excavation sites
Definition of contractors
Work planning
Provision of material and technical resources for reconstruction works
Preparation of measures for the removal of flood waters, dredging and ruslation works
Monitoring the condition of the coastline, embankments within the settlements
The use of local natural and artificial building materials
Delivery of materials and structures for the production of works
Control over the timeliness of work
Drainage of territories
Study of geomorphological hydrogeological and engineering-geological conditions of the developed territories
Conducting special surveys in areas of weakly permeable swelling soils, with a poorly developed erosion network, shallow occurrence of water-resistant layers with uneven roofs, obstructed surface and underground runoff
Construction of intercepting, reservoir, horizontal, vertical, wall-mounted and related drains, anti-filtration screens and curtains
Transportation of farm animals
Identification of safe places for animals
Transport support
Provision of feed
Ensuring security
Ensuring the collection and transportation to the places of consumption of agricultural products
Preparation of forces to eliminate the consequences
Creation, equipping, equipping, training, certification of professional, emergency rescue and rescue equipment on the basis of enterprises, departments of federal and territorial bodies
Maintaining emergency services in readiness
Planning of actions, working out of interaction, conducting exercises and trainings

Fig. 5. Flood consequences liquidation plan.

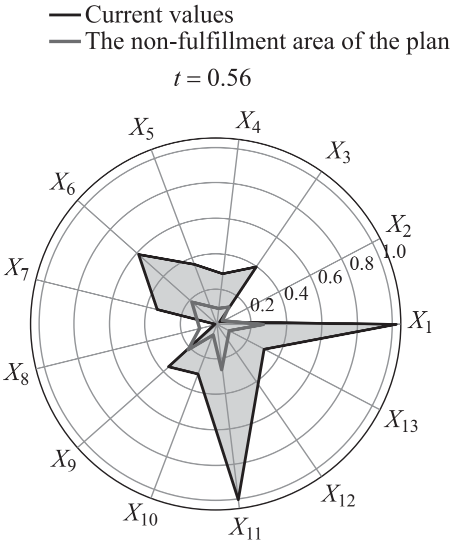


Fig. 6. Feasibility of the action plan at various points in computer time.

To achieve greater clarity, we assume that the plan feasibility check is carried out over a computer time interval $[0;1]$, normalized values of model variables were selected as indicators of feasibility $X_i(t),\ i = 1,\dots,13$. Results of solving a system of equations under initial conditions $X_i(t_0) = 0.5;\ i = 1,\dots,13$ are shown in Fig. 6. They show that the action plan turns out to be impracticable at the moment of computer time $t = 0.56$. For the rest of the interval, the tested plan is feasible. The minimum values of the indicators at which the plan will be feasible are shown in Fig. 6 in gray lines, and the current values of the indicators are shown in black lines.

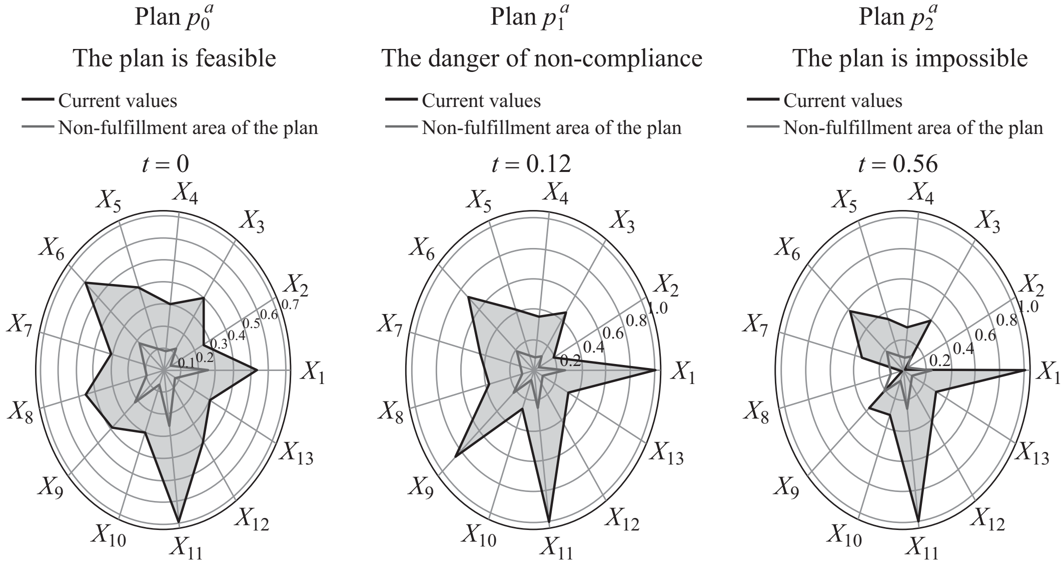


Fig. 7. Checking the feasibility of flood mitigation plans.

Figure 7 shows the results of a feasibility test of three alternative flood response plans. The first indicates the points in time when these plans are feasible, the second when there is a risk of failure, and the third when they are not feasible.

The feasibility of various flood mitigation plans shows it is necessary to check for the feasibility of the plan that implements the most preferred management strategy when solving problems of managing large-scale systems. Solving control problems for large-scale systems shows that the plan implementing the most preferred control strategy must be checked for feasibility. This plan must be feasible at any point in the time interval under consideration. Suppose the plan is not feasible, at least at one point. In that case, preference should be given to a control strategy whose implementation plan will be feasible over the entire control time interval.

8. CONCLUSIONS

Models and methods for checking the achievability of goals and the feasibility of plans implemented when managing large-scale systems in their development are considered. An algorithm for analyzing the achievability of a set of goals and plans implemented when managing these systems is proposed and justified. Statements and hypotheses that make it possible to machine-check the feasibility of plans have been generated.

A model example is given that confirms the possibility of checking the feasibility of plans for eliminating the consequences of a flood using the developed models and methods. In managing large-scale systems development, it is advisable to use control loops that check the achievability of set goals and the feasibility of plans over a selected time interval. In the absence of this verification, the chosen trajectory of development of a large-scale system at specific points in time may turn out to be unrealizable, which will lead to disruption of the work being carried out, as well as to significant costs of the human, financial, technical and other types of resources for the implementation of obviously impracticable plans.

REFERENCES

1. Vasiliev, S.N., From Classical Regulation Problems to Intelligent Control. I, *Proceedings of the Academy of Sciences. Theory and Control Systems*, 2001, no. 1, pp. 5–22.

2. Pupkov, K.A., Voronov, E.M., and Konkov, V.G., *Metody klassicheskoi i sovremennoi teorii avtomaticheskogo upravleniya* (Methods of Classical and Modern Theory of Automatic Control), Moscow: MSTU im. N.E. Bauman, 2004, p. 252.

3. Kushnikov, V.A., Rezchikov, A.F., and Tsvirkun, A.D., Control in Human-Machine Systems with an Automated Procedure for Target Correction, *Autom. Remote Control*, 1978, no. 7, pp. 168–175.

4. Panov, A.I., Goal Setting and Synthesis of a Behavior Plan by a Cognitive Agent, *Iskusstvennyi intellekt i prinyatie reshenii*, 2018, no. 2, pp. 21–35.

5. Sklemin, A.A. and Kushnikov, V.A., Analysis of the Feasibility of Action Plans in the Management of an Industrial Enterprise, *Izv. Vyssh. Uchebn. Zaved., Povolzhskii region. Tekhnicheskie nauki*, 2012, no. 4 (24), p. 18.

6. Sklemin, A.A., Kushnikov, V.A., and Rezchikov, A.F., Models and Algorithms for Checking the Feasibility of Action Plans in the Management of an Industrial Enterprise, *Vestnik SGTU*, 2012, vol. 3, no. 1, (67), p. 145.

7. Pshenichnikov, I.S., Kushnikov, V.A., Shlychkov, E.I., and Rezchikov, A.F., Analysis of the Feasibility of Action Plans in the Automated Control System of a Bridge Construction Organization, *Mekhatronika, Avtomatiz., Upravl.*, 2006, no. 11, p. 45.

8. Shcherbakov, M.A. and Kushnikov, V.A., Analysis of the Feasibility of Goal Trees When Managing Information Security of Enterprises and Organizations, *Vestnik SGTU*, 2013, no. 4 (73), p. 136.

9. Novikov, D.A., *Teoriya upravleniya organizatsionnymi sistemami* (Theory of Management of Organizational Systems), 4th ed., Moscow: Fizmatlit, 2021, 636 p.

10. Dranko, O.I., Novikov, D.A., Raikov, A.N., and Chernov, I.V., *Managing regional development: modeling opportunities*, Moscow: URSS, LLC “LENAND”, 2023, 432 p.

11. Tsvirkun, A.D., Management of the Process of Eliminating the Consequences of Floods at Industrial Facilities and Territories, *Upravlen. Bol'shimi Sist.: Sb. Tr.*, 2020, no. 83, pp. 75–106.

12. Brodsky, Yu.I., *Lektsii po matematicheskomu i imitatsionnomu modelirovaniyu* (Lectures on Mathematical and Simulation Modeling), M.-Berlin: Direct-Media, 2015, p. 240.

13. Geyda, A.S. and Lysenko, I.V., Assessing Indicators of Operational Properties of Systems and Processes of Their Functioning, *Tr. SPII Ross. Akad. Nauk*, 2013, no. 2 (25), pp. 317–337.

14. Davtyan, A.G. et al., Modeling Narrative Management in Socio-Economic Systems, *Vestnik MGTU im. N.E. Bauman, Priborostroenie*, 2022, no. 1 (138), pp. 85–99.

15. Yakhnin, E.D., The Problem of Goal Setting, *Vopr. Filosofii*, 2020, no. 5, pp. 5–11.

16. Kostyuk, F.V., From Calculating Combat Effectiveness Indicators to the Theory of Operations Research and Non-Antagonistic Games: The Scientific Legacy of Professor Yuri Borisovich Germeyer, *Izv. Ross. Akad. Nauk, Teor. Sist. Upravlen.*, 2019, no. 2, pp. 30–40.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1449–1456.

Optimization of Group Incentive Schemes

V. N. Burkov[*,a], **I. V. Burkova**[*,b], **and A. R. Kashenkov**[**,c]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
[**] *Vologda State Technical University, Vologda, Russia*
e-mail: [a]*vlab17@bk.ru,* [b]*irbur27@gmail.com,* [c]*alex27k@mail.ru*

Received May 30, 2023
Revised September 10, 2023
Accepted September 30, 2023

Abstract—This paper considers the problem of motivating the reduction of project duration. The duration cuts of project works and the corresponding costs are given. A group incentive scheme is used to compensate for the costs. In this scheme, all works are partitioned into groups and a unified incentive scheme is applied for each group. Two types of unified incentive schemes are studied for groups, namely, linear and jump ones. The problem is to partition all project works into groups and choose an appropriate incentive scheme for each group by minimizing the total incentive fund. Solution algorithms are proposed based on determining the shortest path in the network. Special cases are also analyzed (partition with the minimum number of groups and partition with the maximum number of groups).

Keywords: group incentive scheme, linear unified incentive scheme, jump unified incentive scheme, shortest path in the network

DOI: 10.25728/arcRAS.2024.15.74.001

1. INTRODUCTION

Consider a project consisting of n works. For each work, the duration cut and the costs of this cut are given. An incentive scheme is defined to compensate for the costs.

The design problems of optimal incentive schemes have been considered by many researchers; for example, see the books [1–3] and references therein. As a rule, two types of such schemes are considered, namely, individual incentive schemes and unified incentive schemes. In individual incentive schemes, a particular scheme is defined for each group from a given class (linear, jump, rank, etc [1]). In unified incentive schemes, the same scheme is defined for all agents. Compared to unified incentive counterparts, the advantage of individual incentive schemes is an appreciably smaller incentive fund (in several cases), and the drawbacks are disinterest in reducing costs and rather high opportunities for manipulation (strategic behavior). The benefits of unified incentive schemes are smaller opportunities for manipulation and significantly greater interest in reducing costs, and the disadvantage is an appreciably larger incentive fund (in many cases). Group incentive schemes occupy an intermediate position. In such schemes, the set of all agents is partitioned into groups and a unified incentive scheme is applied for each group. Group incentive systems retain to some extent the advantages of unified and individual incentive systems and, at the same time, diminish their drawbacks.

In this paper, design problems are formulated for optimal group incentive schemes and methods for solving them are considered.

2. PROBLEM STATEMENT

Consider a project consisting of n works. There is a given plan for reducing project duration: according to this plan, the duration cut of work i is specified by a value y_i. The costs of project contractors to reduce the duration, z_i, $i = \overline{1, n}$, are also given.

To compensate for the costs, it is necessary to determine a group incentive scheme (GIS). Consider a GIS in which all works are partitioned into $1 < m < n$ groups and a certain unified incentive scheme (UIS) is defined for each group. Further analysis deals with two classes of UISs, namely, linear incentive schemes (LISs) and jump incentive schemes (JISs). We denote by R_j the set of works belonging to group j :

$$\bigcup_j R_j = R, \quad R_i \cap R_j = \varnothing \tag{1}$$

for all i and j, where R is the set of all works. If an LIS is chosen for group j, all contractors of this group will be compensated for their costs using the minimum incentive fund

$$S_j = a_j T_j, \tag{2}$$

where

$$a_j = \max_{i \in R_j} k_i, \quad T_j = \sum_{i \in R_j} y_i, \quad k_i = \frac{z_i}{y_i}, \quad i = \overline{1, n}.$$

If a JIS is chosen for group j, all contractors of this group will be compensated for their costs using the minimum incentive fund

$$S_j = n_j \max_{i \in R_j} z_i, \tag{3}$$

where n_j stands for the number of works in group j.

Problem 1. Find a partition R_j, $j = \overline{1, m}$, and choose an appropriate incentive scheme for each group by minimizing the incentive fund. This problem will be considered in three modifications as follows: in the first, only LISs are used for all groups; in the second, only JISs; in the third, both classes of the incentive schemes mentioned (further called mixed incentive schemes, MISs).

Now we describe the formal problem statement. Let $x_{ij} = 1$ if work i belongs to group j, and $x_{ij} = 0$ otherwise. In the case of LISs, the incentive fund of group j is given by

$$S_{1j} = \left(\sum_i x_{ij} y_i \right) \max_i k_i x_{ij}. \tag{4}$$

In the case of JISs, the incentive fund of group j is given by

$$S_{2j} = \left(\sum_i x_{ij} \right) \max_i k_i y_i x_{ij}. \tag{5}$$

Finally, in the case of MISs, the incentive fund of group j is given by

$$S_{3j} = \min \left(S_{1j}, S_{2j} \right). \tag{6}$$

Consequently, the total incentive fund constitutes

$$S_k = \sum_j S_{kj}, \quad k = \overline{1, 3}, \tag{7}$$

depending on the chosen incentive scheme k.

Problem constraints may have different forms. For example, given n_j works in each group j, the constraints are

$$\sum_i x_{ij} = n_j, \quad j = \overline{1,m}. \tag{8}$$

If the number of works in a group must be within given bounds, the constraints take the form

$$l_1 \leqslant \sum_i x_{ij} \leqslant l_2. \tag{9}$$

Other constraints are possible as well.

Problem 2. Find (x_{ij}), $i = \overline{1,n}$, $j = \overline{1,m}$, to minimize (7) subject to the constraints (8), (9) or others.

Methods for solving these problems are presented below.

3. LINEAR INCENTIVE SCHEMES WITH $y_i = y$

In this section, we investigate the case $y_i = y$, $i = \overline{1,n}$. Let all works be numbered in ascending (nondescending) order of k_i, i.e.,

$$k_1 \leqslant k_2 \leqslant \ldots \leqslant k_n.$$

This sequence will be called original.

Definition 1. A fragment of the original sequence is its part between some works i and $j > i$.

Theorem 1. *The optimal partition of works into groups is the set of fragments of original sequences.*

Proof. Assume first that all values k_i differ. Let P be an optimal partition. Consider the group with the maximum value k_n. If this group is not a fragment, then there exists a maximum number s of the original sequence such that the corresponding work is absent from the group with k_n but present in another group, where it has the maximum value k_s. Let us swap work s with any work from the group with k_n that does not belong to the fragment. Obviously, the incentive costs in the group with the maximum value k_n will not change; at the same time, the incentive costs in the group with i_s will decrease because the maximum value k_i in the group with i_s is smaller than k_s, which contradicts the optimality of the partition P. Thus, the group with work n is a fragment. The next group with the maximum value k_i is considered by analogy, and the procedure continues for all groups.

To proceed, we reject the assumption that all k_i are different. In this case, there exist several original sequences; for any optimal partition, however, it is possible to find an original sequence such that the partition will form the set of fragments of this sequence. The proof of Theorem 1 is complete.

Let all works be numbered in ascending (nondescending) order of k_i, i.e.,

$$k_1 \leqslant k_2 \leqslant \ldots \leqslant k_n.$$

We construct a network of admissible partitions (NAP) of works into groups. This network consists of an input, an output, and $(m-1)$ layers. Each vertex i of layer p shows the total number of works Q_{ip} in the first p groups. Note that the minimum number of works is 2 and the maximum number is $n - 2\,(m-p)$ since each group includes at least two works. Therefore, layer p contains

$$a = n - 2\,(m-p) - 2p + 1 = n - 2m + 1$$

vertices, and this number is independent of p.

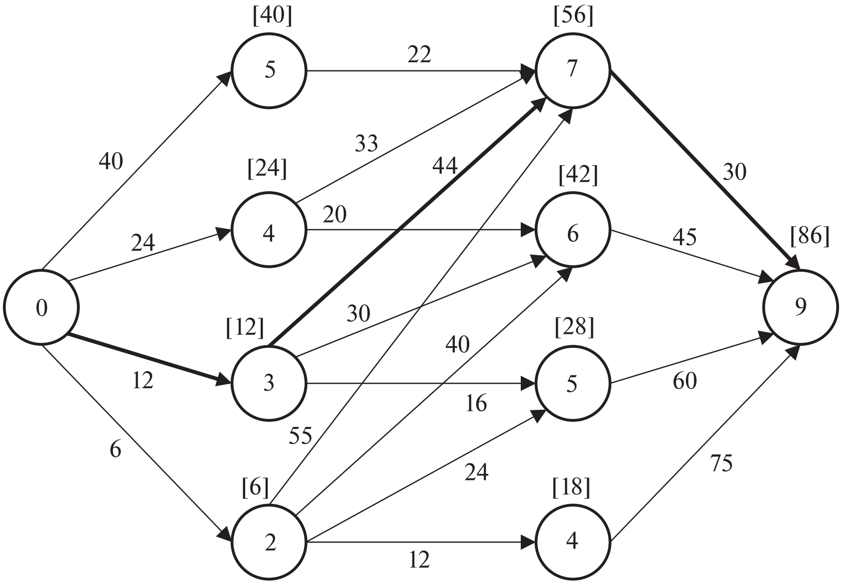


Fig. 1.

We connect vertex i of layer p to vertex j of layer $(p+1)$ by an arc if

$$Q_{jp+1} - Q_{ip} \geqslant 2.$$

Also, we connect the network input 0 to each vertex of layer 1 and each vertex of layer $(m-1)$ to the network output by an arc.

Theorem 2. *A unique path in the NAP corresponds to each admissible partition of works into groups and, conversely, a unique partition of works into groups corresponds to each path in the NAP.*

Proof. For each admissible partition $(n_1, \ldots, n_m)$, there is a sequence of values Q_{ip}, $p = \overline{1, m-1}$, such that the difference of the values of neighbor layers exceeds or equal 2. By the construction of the NAP, an arc connects the corresponding vertices. Conversely, for each path in the NAP, there is a sequence $(n_1, \ldots, n_m)$, where n_k equals the difference $(Q_{jk+1} - Q_{ik})$ of the corresponding adjacent vertices. This sequence defines a unique partition of works into groups. The proof of Theorem 2 is complete.

Example 1. Consider a project of nine works and let $m = 3$. Then we have

$$q = 9 - 6 + 1 = 4.$$

The corresponding network is demonstrated in Fig. 1. Table 1 provides the data of works. Assume that $y_i = 1$ for all i, i.e., $z_i = k_i$. The lengths of all arcs are specified in Fig. 1. The shortest path (0,3,7,9) has a length of 86. The optimal partition into three groups is given by $R_1 = (1, 2, 3)$, $R_2 = (4, 5, 6, 7)$, and $R_3 = (8, 9)$.

Table 1

i	1	2	3	4	5	6	7	8	9
k_i	1	3	4	6	8	10	11	12	15
z_i	1	15	12	12	8	40	33	24	15

Next, we solve the problem with the maximum number of groups $m = [n/2] = 4$. The corresponding NAP is presented in Fig. 2 $(q = 2)$.

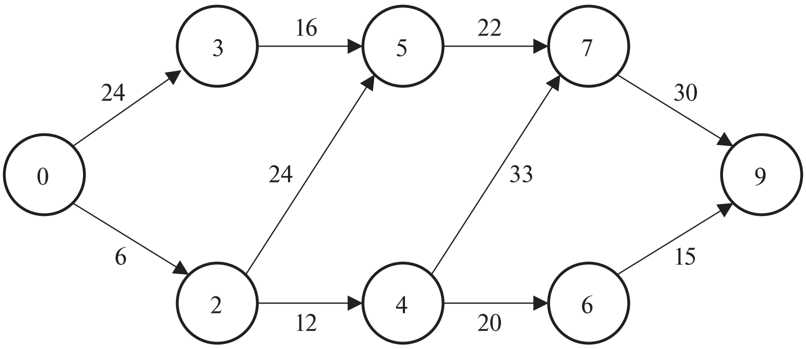


Fig. 2.

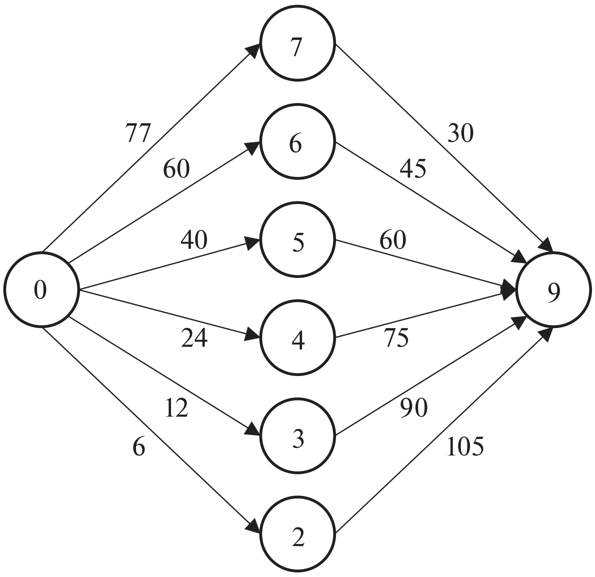


Fig. 3.

We calculate the vertices λ_{ip}:

$$\lambda_{\text{in}} = 0, \quad \lambda_{11} = 6, \quad \lambda_{21} = 24,$$
$$\lambda_{12} = \lambda_{11} + 12 = 18,$$
$$\lambda_{22} = \min\,[\lambda_{11} + 24, \lambda_{21} + 16] = 30,$$
$$\lambda_{13} = \lambda_{12} + 20 = 38,$$
$$\lambda_{23} = \min\,[\lambda_{12} + 33, \lambda_{22} + 22] = 51,$$
$$\lambda_{\text{out}} = \min\,[\lambda_{13} + 45, \lambda_{23} + 30] = 81.$$

The optimal partition is the one with four groups: (1, 2), (3, 4), (5, 6, 7), and (8, 9).

Finally, we solve the problem with the minimum number of groups $m = 2$, $q = 6$. The corresponding NAP is presented in Fig. 3.

We calculate the vertices:

$$\lambda_{\text{in}} = 0, \quad \lambda_{11} = 6, \quad \lambda_{21} = 12,$$
$$\lambda_{31} = 24, \quad \lambda_{41} = 40, \quad \lambda_{51} = 60, \quad \lambda_{61} = 77,$$
$$\lambda_{\text{out}} = \min\,[\lambda_{11} + 105, \lambda_{21} + 90, \lambda_{31} + 75, \lambda_{41} + 60, \lambda_{51} + 45, \lambda_{61} + 30] = 99.$$

The optimal partition is given by (1, 2, 3, 4) and (5, 6, 7, 8, 9).

Note that incentive costs decrease as the number of groups grows, which is fairly obvious.

4. AN HEURISTIC ALGORITHM

The algorithm described above can be applied to the general case of different values y_i as a heuristic. It can be justified as follows. If there exists a partition into groups such that the coefficients k_i are the same for each group, then this partition is optimal. Therefore, a reasonable assumption is that the closer the coefficients k_i in the groups are, the closer the partition will be to the optimal one.

Example 2. Consider the problem with three groups and the data of Table 1. The corresponding network with the arc lengths (5) is shown in Fig. 4. Its structure coincides with that of the network in Fig. 1.

We calculate the vertices:

$$\lambda_{\text{in}} = 0, \quad \lambda_{11} = 18, \quad \lambda_{21} = 36, \quad \lambda_{31} = 66, \quad \lambda_{41} = 96,$$
$$\lambda_{12} = 48, \quad \lambda_{22} = \min[18 + 48,\ 36 + 24] = 60, \quad \lambda_{32} = 106, \quad \lambda_{42} = 146,$$
$$\lambda_{\text{out}} = \min[48 + 165,\ 60 + 150,\ 106 + 90,\ 146 + 45] = 191.$$

The optimal partition is given by (1, 2, 3), (4, 5, 6, 7), and (8, 9).

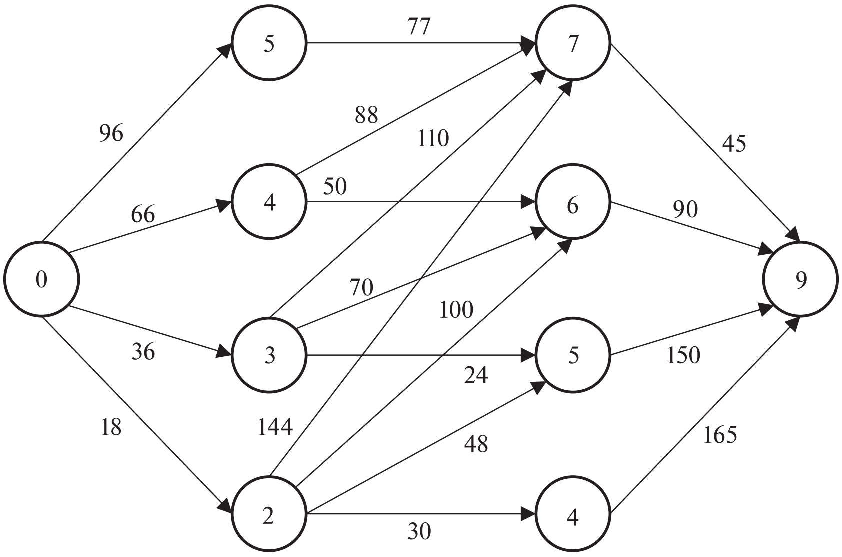


Fig. 4.

5. JUMP INCENTIVE SCHEMES

In this section, we study jump incentive schemes (JISs). For such schemes, an analog of Theorem 1 holds. Let all works be numbered in ascending (nondescending) order of z_i, i.e.,

$$z_1 \leqslant z_2 \leqslant \ldots \leqslant z_n.$$

This sequence will also be called the original sequence. The definition of its fragment is the same as the one for linear incentive schemes. (In other words, a fragment is some part of the original sequence.)

Theorem 3. *The optimal partition is the set of fragments of the original sequence (one of them if there are several original sequences).*

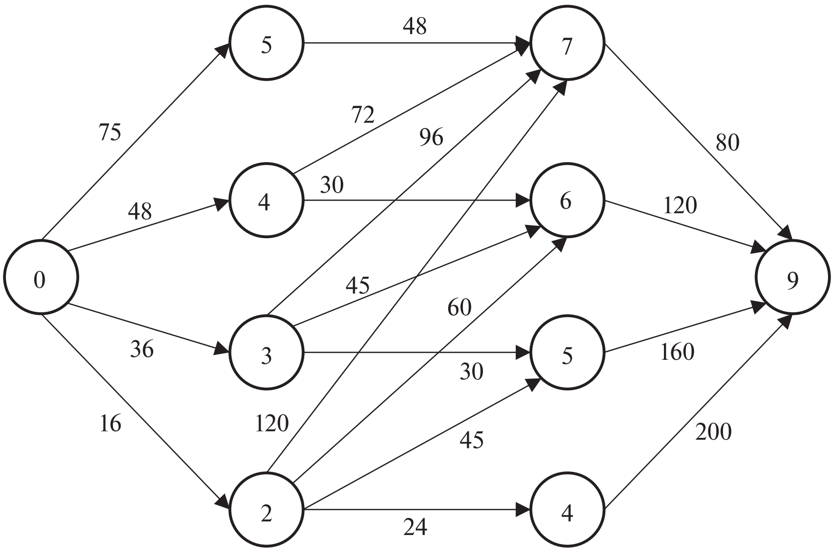


Fig. 5.

Proof. This result is established similarly to Theorem 1. In the optimal partition, we take the group with the maximum value z_n and show that it is a fragment. Assume on the contrary that it is not. We find the work with the maximum value z_s in a fragment that is absent from this group but present in another group. Let us swap work s with any work from the group with work n that does not belong to the fragment. Obviously, the incentive costs will decrease. Thus, the group with work n is a fragment. Then we eliminate the works of this fragment and consider the next group with the maximum value z. The procedure continues for all groups by analogy.

Example 3. Consider the data of Table 1. We renumber the works appropriately to obtain an original sequence.

Table 2

i	1	2	3	4	5	6	7	8	9
K_i	1	8	4	6	15	3	12	11	10
y_i	1	1	3	2	1	5	2	3	4
z_i	1	8	12	12	15	15	24	33	40

Let us find the optimal GIS for three groups. Note that the corresponding network will have the same structure as in Fig. 1 but with other arc lengths (see Fig. 5).

We calculate the vertices:

$$\lambda_{\text{in}} = 0, \quad \lambda_{11} = 16, \quad \lambda_{21} = 36, \quad \lambda_{31} = 48, \quad \lambda_{41} = 75,$$
$$\lambda_{12} = 40, \quad \lambda_{22} = \min\,[16+45,\ 36+30] = 61,$$
$$\lambda_{32} = \min\,[48+30,\ 36+45,\ 16+60] = 76,$$
$$\lambda_{42} = \min\,[75+48,\ 48+72,\ 36+96,\ 16+120] = 120,$$
$$\lambda_{\text{out}} = \min\,[40+200,\ 61+160,\ 76+120,\ 120+80] = 196.$$

The optimal partition is given by (1, 2), (3, 4, 5, 6), and (7, 8, 9).

6. TWO-GROUP PARTITION FOR LINEAR INCENTIVE SCHEMES

This section is devoted to a special case of partitions into two groups. Let the maximum coefficient k_j be given for the second group. The resulting problem is easy to solve. If $k_j < k_{n-1}$, then the first group includes all works with $k_i > k_j$ whereas the second group all works with $k_i \leqslant k_j$. Indeed, any transfer of work with $k_i \leqslant k_j$ to the first group increases the incentive fund by $(k_{n-1} - k_j)y_i > 0$. If $k_j = k_{n-1} < k_n$, we add a work to the first group for the number of works to exceed one. The matter concerns the work with the minimum value y.

Example 4. Consider the data of Table 1. We perform the calculations:

1. $k_j = 12$. We add work 1 with the minimal duration $y_1 = 1$ in the first group. The incentive fund is

$$\Phi_1 = 15 \times 2 + 12 \times 20 = 270.$$

2. $k_j = 11$. The first group contains works 8 and 9.

$$\Phi_2 = 45 + 209 = 254.$$

3. $k_j = 10$. The first group contains works 7, 8, and 9. The incentive fund is

$$\Phi_3 = 90 + 160 = 250.$$

4. $k_j = 8$. The first group contains works 6, 7, 8, and 9. The incentive fund is

$$\Phi_4 = 150 + 96 = 246.$$

5. $k_j = 6$. The first group contains works 5, 6, 7, 8, and 9. The incentive fund is

$$\Phi_5 = 165 + 66 = 231.$$

6. $k_j = 4$. The first group contains works from 4 to 9. The incentive fund is

$$\Phi_6 = 195 + 36 = 231.$$

7. $k_j = 3$. The first group contains works from 3 to 9. The incentive fund is

$$\Phi_4 = 240 + 18 = 258.$$

The optimal partition of works into groups is given by (1, 2, 3) and (4, 5, 6, 7, 8, 9).

7. CONCLUSIONS

This paper has considered the design of group incentive schemes for linear and jump incentive schemes. Note that for the two-group partition with linear incentive schemes, the heuristic algorithm yields an optimal solution in many cases. It seems interesting to justify this conclusion rigorously. Another promising line is to consider other incentive schemes (basic and combined). As for mixed incentive systems, we emphasize that any linear or jump incentive scheme can be turned into a mixed one by recalculating the arc lengths of the corresponding network using formula (6). However, generally speaking, the resulting solution will be nonoptimal. The problem of an optimal mixed incentive scheme has not been solved yet. All these problems require further research.

REFERENCES

1. Novikov, D.A., *Stimulirovanie v organizatsionnykh sistemakh* (Incentives in Organizational Systems), Moscow: Sinteg, 2003.
2. Novikov, D.A. and Tsvetkov, A.V., *Mekhanizmy stimulirovaniya v mnogoelementnykh organizatsionnykh sistemakh* (Incentive Mechanisms in Multielement Organizational Systems), Moscow: Apostrof, 2000.
3. Burkov, V.N., Burkova, I.V., Goubko, M.V., et al., *Mekhanizmy upravleniya* (Control Mechanisms), Novikov, D.A., Ed., Moscow: LENAND, 2013.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1457–1467.

Comparison of Distribution Procedures in Blended Finance

A. V. Shchepkin

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
e-mail: av_shch@mail.ru

Received May 31, 2023

Revised September 9, 2023

Accepted September 30, 2023

Abstract—This paper is devoted to the blended (joint) finance mechanism of a megaproject consisting of several projects. One part of the megaproject budget comes from the megaproject manager and the other part from project contractors. When distributing this budget, the megaproject manager considers information about the amount of the contractor's internal funds allocated to project implementation. Project contractors seek to get more funds from the megaproject manager; in turn, the megaproject manager is interested in attracting more funds from project contractors. To achieve this goal, the megaproject manager applies different procedures to distribute the budget. Project contractors use the information reported to the megaproject manager to increase the funds allocated to them. Straight and reverse priority distribution procedures in the blended finance mechanism are analyzed. A distribution procedure is determined that stimulates project contractors to allocate more of their internal funds to the project in a Nash equilibrium.

Keywords: blended finance, straight priorities, reverse priorities, planned profit, factual profit

DOI: 10.25728/arcRAS.2024.82.33.001

1. INTRODUCTION

Megaprojects are often financed by several sources jointly. In this case, a typical situation is that one source is the megaproject manager and the other is the contractors of the individual projects making up the megaproject. In other words, the blended finance mechanism is implemented; see [1–3]. (Such a mechanism is also called joint finance.) As a rule, the budget of the entire megaproject is limited and turns out to be insufficient to implement the required projects. According to [4], the idea of blended finance is that funds from the megaproject budget are allocated on the condition that the contractor of each project commits to allocate its internal funds to its project.

Blended finance implies that it is profitable for project contractors to invest their internal funds. However, the megaproject manager faces the problem of distributing the budget among project contractors. Traditionally, in the theory of active systems [5, 6], the megaproject manager requests information about the necessary funds from project contractors to implement the corresponding distribution mechanisms. The amount of funds received by the project contractors significantly depends on the information reported, the megaproject budget, and its distribution procedure. At the same time, the amount of funds for each contractor depends on its information and the information of all project contractors.

In the studies of foreign researchers, the consideration of blended finance mechanisms is associated with the evaluation of specific instruments, such as equity capital, guarantees, loans, etc. [7, 8]. Blended finance is treated as the use of capital from public or philanthropic sources to augment private-sector investment [9]. Special attention is paid to the issues of investment under which blended finance increases the potential return on investments or reduces risk factors, making them more attractive to investors [10, 11].

In this paper, we analyze straight and reverse priority distribution procedures applied by the megaproject manager in the blended finance mechanism. We determine a distribution procedure that stimulates project contractors to allocate more of their internal funds to the project in a Nash equilibrium [5].

2. AGENTS FUNDING UNDER PRINCIPAL'S COMPLETE AWARENESS

Consider a two-level system consisting of a Principal (the megaproject manager, the upper level), which distributes a budget for project implementation, and agents (project contractors, the lower level). The megaproject consists of n projects and is implemented by n contractors (agents). Each agent knows the factual costs z_i of implementing project i, where $i = 1, \ldots, n$. The Principal has funds in an amount R, which are distributed among project contractors. The Principal's complete awareness implies that the Principal knows the factual costs of each project.

The game-theoretic statement of the problem is as follows.

1. Each agent reports to the Principal the value w, which is some part of the factual project implementation costs allocated by the agent from its internal funds. For agent i, the planned amount $u_i^{(\mathrm{p})}$ of its internal funds for project implementation is therefore given by

$$u_i^{(\mathrm{p})} = w_i z_i, \quad i = 1, \ldots, n.$$

It follows that the finance request of agent i is given by

$$s_i = (1 - w_i) z_i, \quad i = 1, \ldots, n.$$

2. The Principal determines the amount of funds c_i, $i = 1, \ldots, n$, for all projects based on the information received. If $c_i < s_i$, the factual amount $u_i^{(\mathrm{f})}$, $i = 1, \ldots, n$, of agent's internal funds for project implementation is given by

$$u_i^{(\mathrm{f})} = z_i - c_i, \quad i = 1, \ldots, n.$$

3. The agents and the Principal determine their payoffs. The agent's payoff is its profit. The Principal's payoff function may have different forms. It does not matter here: this paper aims to establish conditions ensuring the allocation of more agents' internal funds to the project.

Let agent i gain an effect E_i from the project implemented. Assume also that the project will be implemented (and the agent will gain the effect) only if $c_i + u_i^{(\mathrm{f})} \geqslant z_i$. In this case, the profit of agent i can be written as

$$f_i = \mathrm{E}_i + c_i - u_i^{(\mathrm{f})}, \quad i = 1, \ldots, n. \tag{1}$$

We begin with the case when the Principal can distribute the requested funds in full to all agents. Then, obviously, $c_i = s_i$, $i = 1, \ldots, n$, and

$$f_i = \mathrm{E}_i + (1 - 2w_i) z_i, \quad i = 1, \ldots, n. \tag{2}$$

According to (2), to increase their profits, agents are interested in reducing their internal funds for project implementation and maximizing their finance requests. To eliminate this interest, the Principal introduces an additional condition. For receiving funds from the Principal, agents should allocate their internal funds in an amount not less than dz_i, where d is the share of factual costs set by the Principal. In this case, the request of agent i is given by

$$s_i = (1 - d) z_i, \quad i = 1, \ldots, n.$$

Consequently, the profit of agent i is given by

$$f_i = \mathrm{E}_i + (1-2d)z_i \geqslant 0, \quad i = 1, \ldots, n.$$

Due to the latter expression, the agent can affect the amount of profit only when setting $w_i > d$. However, see the discussion above, agents are interested in reducing their internal funds for project implementation; this corresponds to $w_i = d$.

If the Principal's funds are limited, priority distribution procedures [6] are used to determine the amount of funds c_i for project i, $i = 1, \ldots, n$. These procedures have the following form:

$$c_i^{(\mathrm{sp})} = \min\left\{ s_i;\ \frac{A_i s_i}{\sum\limits_{q=1}^{n} A_q s_q} R \right\}, \quad i = 1, \ldots, n$$

(straight priorities) and

$$c_i^{(\mathrm{rp})} = \min\left\{ s_i;\ \frac{A_i}{s_i \sum\limits_{q\in N} A_q/s_q} R \right\}, \quad i = 1, \ldots, n \tag{3}$$

(reverse priorities). Here, A_i denotes the project priority set by the Principal for agent i.

First, we study the straight priority procedure. Since the agents are financed under the Principal's complete awareness, the amount of funds allocated to project i is given by

$$c_i^{(\mathrm{sp})} = \frac{A_i(1-w_i)z_i}{\sum\limits_{q=1}^{n} A_q(1-w_q)z_q} R, \quad i = 1, \ldots, n. \tag{4}$$

The condition $c_i^{(\mathrm{sp})} + u_i^{(\mathrm{f})} \geqslant z_i$ must hold for agent i to gain the effect E_i. Hence,

$$\sum_{q=1}^{n} u_q^{(\mathrm{f})} \geqslant \sum_{q=1}^{n} z_q - R. \tag{5}$$

This conclusion seems obvious enough. If the implementation of all projects requires the amount of funds $\sum\limits_{q=1}^{n} z_q$, and the Principal allocates the amount of funds R, then the expression (5) exactly determines the amount of agent's internal funds.

Given (4), the goal function (1) of agent i takes the form

$$f_i = \mathrm{E}_i + \frac{2A_i(1-w_i)z_i}{\sum\limits_{q=1}^{n} A_q(1-w_q)z_q} R - z_i, \quad i = 1, \ldots, n.$$

Obviously,

$$\frac{\partial f_i}{\partial w_i} = -2A_i z_i \frac{\sum\limits_{q=1}^{n} A_q(1-w_q)z_q - A_i z_i(1-w_i)}{\left(\sum\limits_{q=1}^{n} A_q(1-w_q)z_q\right)^2} R < 0.$$

Therefore, the agents have an interest in reducing their internal funds for project implementation and maximizing their finance requests.

On the other hand, the Principal seeks to attract more of the agents' internal funds for project implementation. Accordingly, the Principal sets the priority of agent i so that it increases with the growing amount of the agent's internal funds allocated to the project. For example, the priority can be defined as

$$A_i = \frac{a_i}{1 - w_i}. \tag{6}$$

This priority has a peculiarity as follows. The more internal funds the agent allocates to the project, the higher its priority will be.

In this case, formula (4) can be written as

$$c_i^{(\mathrm{sp})} = \frac{a_i z_i}{\sum\limits_{q=1}^{n} a_q z_q} R, \quad i = 1, \ldots, n. \tag{7}$$

Accordingly, agent i, $i = 1, \ldots, n$, allocates the following factual amount of its internal funds for project implementation:

$$u_i^{(\mathrm{f,dp})} = z_i - c_i^{(\mathrm{sp})}, \quad i = 1, \ldots, n.$$

Let all projects being implemented satisfy the following requirement.

Condition 1. All projects are divided into two groups. The projects of the first group, those with the numbers $i = 1, \ldots, m$, have the high priorities $a_i = b^3 > 1$. The projects of the second group, those with the numbers $i = m + 1, \ldots, n$, have the low priorities $a_i = 1$.

In this case, formula (7) can be written as

$$c_i^{(\mathrm{sp})} = \begin{cases} \dfrac{b^3 z_i}{b^3 \sum\limits_{q=1}^{m} z_q + \sum\limits_{q=m+1}^{n} z_q} R, & i = 1, \ldots, m, \\[2ex] \dfrac{z_i}{b^3 \sum\limits_{q=1}^{m} z_q + \sum\limits_{q=m+1}^{n} z_q} R, & i = m + 1, \ldots, n. \end{cases} \tag{8}$$

Let $z_1 = z_n$, i.e., the costs of project 1 with the high priority coincide with those of project n with the low priority. In this case, due to (8), the agent whose project has the low priority receives fewer funds from the Principal and, accordingly, allocates more of its internal funds to the project.

In addition, according to (7), the amount of agents' funds is independent of the information reported by the agents.

Now, we analyze the reverse priority procedure. The procedure (3) can be represented as

$$c_i^{(\mathrm{rp})} = \min \left\{ (1 - w_i) z_i ; \ \frac{A_i}{(1 - w_i) z_i \sum\limits_{q \in N} A_q \Big/ [(1 - w_q) z_q]} R \right\}, \quad i = 1, \ldots, n.$$

The agent receives the maximum amount of funds under the condition

$$(1 - w_i) z_i = \frac{A_i}{(1 - w_i) z_i \sum\limits_{q \in N} A_q \Big/ [(1 - w_q) z_q]} R, \quad i = 1, \ldots, n.$$

After straightforward calculations, we obtain

$$(1 - w_i) z_i = \frac{\sqrt{A_i}}{\sum\limits_{q \in N} \sqrt{A_q}} R, \quad i = 1, \ldots, n. \tag{9}$$

If the Principal sets the priorities (6), then the relation (9) can be written as

$$c_i^{(\mathrm{rp})} = (1-w_i)z_i = \frac{\sqrt[3]{a_i z_i}}{\sum\limits_{q\in N}\sqrt[3]{a_q z_q}}R, \quad i=1,\ldots,n, \tag{10}$$

and, accordingly,

$$u_i^{(\mathrm{f,rp})} = z_i - c_i^{(\mathrm{rp})}, \quad i=1,\ldots,n.$$

Under Condition 1, formula (10) reduces to

$$c_i^{(\mathrm{rp})} = \begin{cases} \dfrac{b\sqrt[3]{z_i}}{b\sum\limits_{q=1}^{m}\sqrt[3]{z_q} + \sum\limits_{q=m+1}^{n}\sqrt[3]{z_q}}R, & i=1,\ldots,m, \\[2ex] \dfrac{\sqrt[3]{z_i}}{b\sum\limits_{q=1}^{m}\sqrt[3]{z_q} + \sum\limits_{q=m+1}^{n}\sqrt[3]{z_q}}R, & i=m+1,\ldots,n. \end{cases} \tag{11}$$

Assuming $z_1 = z_n$ and considering (11), we obtain $u_1^{(\mathrm{f,rp})} = z_1 - c_1^{(\mathrm{f,rp})}$ and $u_n^{(\mathrm{f,rp})} = z_n - c_n^{(\mathrm{rp})}$. Direct comparison of $u_1^{(\mathrm{f,rp})}$ and $u_n^{(\mathrm{f,rp})}$ gives a result similar to the one established for the straight priority principle.

3. AGENTS FUNDING UNDER PRINCIPAL'S INCOMPLETE AWARENESS

Under incomplete awareness, the Principal does not know the factual costs z_i, $i=1,\ldots,n$, of each project and receives information about the planned costs Z_i, $i=1,\ldots,n$, of projects from the agents.

In this case, each agent reports to the Principal the planned costs Z_i, $i=1,\ldots,n$, and the value w_i, which is some part of the planned costs covered by the agent from its internal funds. Hence,

$$u_i = w_i Z_i, \quad i=1,\ldots,n.$$

Accordingly, the finance request of agent i is given by

$$s_i = (1-w_i)Z_i, \quad i=1,\ldots,n. \tag{12}$$

The factual profit of agent i is given by

$$f_i^{(\mathrm{f})} = \mathrm{E}_i + c_i - z_i, \quad i=1,\ldots,n, \tag{13}$$

and its planned profit can be written as

$$f_i^{(\mathrm{p})} = \mathrm{E}_i + c_i - Z_i, \quad i=1,\ldots,n.$$

Let us represent (13) in the form

$$f_i^{(\mathrm{f})} = \mathrm{E}_i + c_i - z_i = \mathrm{E}_i + c_i - (z_i - Z_i + Z_i) = f_i^{(\mathrm{p})} + Z_i - z_i, \quad i=1,\ldots,n.$$

The factual costs z_i are known to the agents, and the agent cannot receive more funds from the Principal than it plans to spend. Therefore, by a natural assumption, the planned costs Z_i exceed the factual ones. In this case, the difference $(Z_i - z_i) > 0$ can be treated as the excess planned profit $f_i^{(\mathrm{ep})} = Z_i - z_i$. In the sequel, the factual profit of agent i is calculated as

$$\begin{aligned} f_i^{(\mathrm{f})} = f_i^{(\mathrm{p})} + q f_i^{(\mathrm{ep})} &= \mathrm{E}_i + c_i - Z_i + q(Z_i - z_i) \\ &= \mathrm{E}_i + c_i - (1-q)Z_i - q z_i, \quad i = 1, \ldots, n, \end{aligned} \tag{14}$$

where $q \leqslant 1$. If $q \in (0, 1]$, the Principal leaves some of the excess profit at the agent's disposal. Accordingly, q is the norm determining the amount of excess profit left to the agent. If $q \leqslant 0$, then q is the penalty coefficient for manipulating the agent's information about the project implementation costs; see [12].

As before, we begin with the case where the Principal can distribute the requested funds to all agents in full. Then, obviously, $c_i = s_i$, $i = 1, \ldots, n$, and

$$f_i^{(\mathrm{f})} = \mathrm{E}_i + s_i - (1-q)Z_i - q z_i, \quad i = 1, \ldots, n. \tag{15}$$

In view of (12), the expression (15) can be written as

$$f_i^{(\mathrm{f})} = \mathrm{E}_i + (1 - w_i + q)Z_i - q z_i, \quad i = 1, \ldots, n. \tag{16}$$

According to (16), the agents always benefit by overestimating their planned costs: for $q \in (0, 1]$,

$$1 - w_i + q > 0.$$

If the Principal's funds are limited, then (similar to the case of complete information) the Principal uses priority distribution procedures [6] to determine the amount of funds c_i for project i, $i = 1, \ldots, n$, and the agent's goal function has the form (14).

First, we consider the straight priority procedure. The amount of funds allocated by the Principal for implementing project i is given by

$$c_i^{(\mathrm{sp})} = \frac{A_i(1-w_i)Z_i}{\sum\limits_{q=1}^{n} A_q(1-w_q)Z_q} R, \quad i = 1, \ldots, n.$$

If the Principal sets the priorities (6), then

$$c_i^{(\mathrm{sp})} = \frac{a_i Z_i}{\sum\limits_{q=1}^{n} a_q Z_q} R, \quad i = 1, \ldots, n.$$

In this case, the goal function (14) takes the form

$$f_i^{(\mathrm{f})} = \mathrm{E}_i + \frac{a_i Z_i}{\sum\limits_{q=1}^{n} a_q Z_q} R - (1-q)Z_i - q z_i, \quad i = 1, \ldots, n.$$

To find the planned costs Z_i^* in a Nash equilibrium, we solve the system of equations

$$\frac{\partial f_i^{(\mathrm{f})}}{\partial Z_i} = a_i \frac{\sum\limits_{q=1}^{n} a_q Z_q - a_i Z_i}{\left(\sum\limits_{q=1}^{n} a_q Z_q\right)^2} R - (1-q) = 0, \quad i = 1, \ldots, n. \tag{17}$$

The solution of (17) is

$$Z_i^* = \frac{n-1}{(1-q)a_i \sum\limits_{q=1}^{n} \dfrac{1}{a_q}} R \left(1 - \frac{n-1}{a_i \sum\limits_{q=1}^{n} \dfrac{1}{a_q}} \right), \quad i = 1, \ldots, n. \tag{18}$$

In the Nash equilibrium, agent i receives the amount of funds

$$c_i^{*(\mathrm{sp})} = \left(1 - \frac{n-1}{a_i \sum\limits_{q=1}^{n} \dfrac{1}{a_q}} \right) R, \quad i = 1, \ldots, n. \tag{19}$$

Accordingly, $u_i^{*(\mathrm{sp})} = z_i - c_i^{*(\mathrm{sp})}$, $i = 1, \ldots, n$.

Under Condition 1, the expression (18) can be written as

$$Z_i^{*(\mathrm{sp})} = \begin{cases} \dfrac{(n-1)[(b^3-1)(n-m)+1]}{(1-q)[m+b^3(n-m)]^2} R, & i = 1, \ldots, m, \\[2ex] \dfrac{(n-1)b^3[b^3 - m(b^3-1)]}{(1-q)[m+b^3(n-m)]^2} R, & i = m+1, \ldots, n. \end{cases} \tag{20}$$

In this Nash equilibrium, agent i receives the amount of funds

$$c_i^{*(\mathrm{sp})} = \begin{cases} \dfrac{(b^3-1)(n-m)+1}{m+b^3(n-m)} R, & i = 1, \ldots, m, \\[2ex] \dfrac{b^3 - m(b^3-1)}{m+b^3(n-m)} R, & i = m+1, \ldots, n. \end{cases}$$

The positivity requirement of project funding leads to the inequality

$$b^3 - (b^3-1)m > 0. \tag{21}$$

From (21) we arrive at

$$m < \frac{b^3}{b^3-1}. \tag{22}$$

Due to (22), the greater ratio of the maximum priority to the minimum one is, the smaller the number of projects with the maximum priority should be.

In the case $m = n$ (all projects are equally important for the Principal), the expressions (18) and (19) reduce to

$$\begin{cases} \hat{Z}_i^* = \dfrac{n-1}{(1-q)n^2} R \\[2ex] \hat{c}_i^{*(\mathrm{sp})} = R/n, \end{cases} \quad i = 1, \ldots, n. \tag{23}$$

Using (23), we can express the amount of agents' internal funds allocated to projects in the Nash equilibrium: $\hat{u}_i^{*(\mathrm{sp})} = z_i - \hat{c}_i^{*(\mathrm{sp})}$, $i = 1, \ldots, n$.

Under the assumption $z_1 = z_n$, direct comparison of $\hat{u}_1^{*(\mathrm{sp})}$ and $\hat{u}_n^{*(\mathrm{sp})}$ indicates the following: the agent whose project has the low priority allocates more of its internal funds to project implementation.

Indeed, this result is immediate from the inequality

$$z_n - \frac{b^3 - m(b^3-1)}{m + b^3(n-m)} R > z_1 - \frac{(b^3-1)(n-m)+1}{m+b^3(n-m)} R.$$

Now, we take the reverse priority procedure. In this case, the Principal distributes to the implementation of project i the amount of funds

$$c_i^{(\mathrm{rp})} = \min\left\{(1-w_i)Z_i;\ \frac{A_i}{(1-w_i)Z_i \sum\limits_{q\in N} A_q \Big/ [(1-w_q)Z_q]} R\right\}, \quad i = 1,\ldots,n.$$

The agent receives the maximum amount under the condition

$$(1-w_i)Z_i = \frac{A_i}{(1-w_i)Z_i \sum\limits_{q\in N} A_q \Big/ [(1-w_q)Z_q]} R, \quad i = 1,\ldots,n.$$

Hence, it follows that

$$c_i^{(\mathrm{rp})} = \frac{\sqrt[3]{a_i Z_i}}{\sum\limits_{q\in N} \sqrt[3]{a_q Z_q}} R, \quad i = 1,\ldots,n. \tag{24}$$

Let Condition 1 be valid; then (24) can be written as

$$c_i^{(\mathrm{rp})} = \begin{cases} \dfrac{b\sqrt[3]{Z_i}}{b\sum\limits_{q=1}^{m} \sqrt[3]{Z_q} + \sum\limits_{q=m+1}^{n} \sqrt[3]{Z_q}} R, & i = 1,\ldots,m, \\[2ex] \dfrac{\sqrt[3]{Z_i}}{b\sum\limits_{q=1}^{m} \sqrt[3]{Z_q} + \sum\limits_{q=m+1}^{n} \sqrt[3]{Z_q}} R, & i = m+1,\ldots,n. \end{cases} \tag{25}$$

In this case, the goal function (14) takes the form

$$f_i^{(\mathrm{f})} = \begin{cases} \mathrm{E}_i + \dfrac{b\sqrt[3]{Z_i}}{b\sum\limits_{q=1}^{m} \sqrt[3]{Z_q} + \sum\limits_{q=m+1}^{n} \sqrt[3]{Z_q}} R - (1-q)Z_i - qz_i, & i = 1,\ldots,m, \\[2ex] \mathrm{E}_i + \dfrac{\sqrt[3]{Z_i}}{b\sum\limits_{q=1}^{m} \sqrt[3]{Z_q} + \sum\limits_{q=m+1}^{n} \sqrt[3]{Z_q}} R(1-q)Z_i - qz_i, & i = m+1,\ldots,n. \end{cases}$$

Under the weak contagion condition [5]

$$\frac{\partial}{\partial Z_i} \frac{1}{b\sum\limits_{q=1}^{m} \sqrt[3]{Z_q} + \sum\limits_{q=m+1}^{n} \sqrt[3]{Z_q}} = 0,$$

the planned costs Z_i^* in a Nash equilibrium are found by solving the system of equations

$$\frac{\partial f_i^{(\mathrm{f})}}{\partial Z_i} = \begin{cases} \dfrac{b}{3Z_i^{2/3}\left[be\sum\limits_{q=1}^{m}\sqrt[3]{Z_q} + \sum\limits_{q=m+1}^{n}\sqrt[3]{Z_q}\right]}R - (1-q) = 0, \quad i = 1, \ldots, m, \\ \dfrac{1}{3Z_i^{2/3}\left[be\sum\limits_{q=1}^{m}\sqrt[3]{Z_q} + \sum\limits_{q=m+1}^{n}\sqrt[3]{Z_q}\right]}R - (1-q) = 0, \quad i = m+1, \ldots, n. \end{cases} \tag{26}$$

From (26) we obtain

$$Z_i^* = \begin{cases} \dfrac{b\sqrt{b}}{3(1-q)\left[mb\sqrt{b} + (n-m)\right]}R, \quad i = 1, \ldots, m, \\ \dfrac{1}{3(1-q)\left[mb\sqrt{b} + (n-m)\right]}R, \quad i = m+1, \ldots, n. \end{cases} \tag{27}$$

According to (27), the planned costs of the agent with the high priority exceed in the Nash equilibrium those of the agent with the low priority.

Using (25), we calculate the amount of agent funds in the Nash equilibrium:

$$c_i^{*(\mathrm{rp})} = \begin{cases} \dfrac{b^{3/2}}{b^{3/2}m + (n-m)}R, \quad i = 1, \ldots, m, \\ \dfrac{1}{b^{3/2}m + (n-m)}R, \quad i = m+1, \ldots, n. \end{cases} \tag{28}$$

According to (28), the funding of the agent with the high priority exceeds in the Nash equilibrium that of the agent with the low priority.

Using (28), we can express the amount of agents' internal funds allocated to projects in the Nash equilibrium: $u_i^{*(\mathrm{rp})} = z_i - c_i^{*(\mathrm{rp})}$.

Under the assumption $z_1 = z_n$, direct comparison of $u_1^{*(\mathrm{sp})}$ and $u_n^{*(\mathrm{sp})}$ shows the following: the agent whose project has the low priority allocates more of its internal funds to project implementation.

Indeed, this result is immediate from the inequality

$$z_n - \frac{1}{b^{3/2}m + (n-m)}R > z_1 - \frac{b^{3/2}}{b^{3/2}m + (n-m)}R.$$

In the case $m = n$ (all projects are equally important for the Principal), the expressions (27) and (28) take the form

$$Z_i^* = \frac{R}{3(1-q)}, \quad i = 1, \ldots, n,$$

and, consequently, $c_i^{*(\mathrm{rp})} = R/n$, $i = 1, \ldots, n$.

Let us demonstrate that

$$c_i^{*(\mathrm{sp})} < c_i^{*(\mathrm{rp})}, \quad i = m+1, \ldots, n. \tag{29}$$

Inequality (29) can be written as

$$\frac{b^3 - m(b^3-1)}{m + b^3(n-m)}R < \frac{1}{b^{3/2}m + (n-m)}R.$$

Trivial transformations yield

$$\left[b^3 - m(b^3 - 1)\right]\left(b^{3/2} - 1\right) - (n-1)(b^3 - 1) < 0. \tag{30}$$

Since $m \geqslant 1$, inequality (30) holds if

$$\left(b^{3/2} - 1\right) - (n-1)(b^3 - 1) < 0.$$

With this condition written as

$$1 - (n-1)\left(b^{3/2} + 1\right) < 0,$$

inequality (29) obviously holds. In fact, we have established the following result: in a Nash equilibrium, using the inverse priority principle, the Principal distributes more funds to the agents with the low priority than in the case of the straight priority principle. Therefore, $u_i^{*(\mathrm{f,rp})} < u_i^{*(\mathrm{f,dp})}$, $i = m+1, \ldots, n$.

4. CONCLUSIONS

According to the above analysis of the blended finance mechanism model, we draw the following conclusions. In the case of the Principal's complete awareness and the straight (or reverse) priority distribution procedure, agents allocate different amounts of their internal funds for projects with different priorities but the same factual costs. In addition, the agent whose project has a low priority receives less funds from the Principal and, accordingly, allocates more of its internal funds for project implementation.

In the case of the Principal's incomplete awareness and the straight (or reverse priority) distribution procedure, the agent with the high priority receives more funds in a Nash equilibrium than that with a low priority. Under the same factual costs, the agent whose project has a low priority allocates more of its internal funds to the project. Note that in a Nash equilibrium, the Principal's reverse priority distribution procedure provides the agents with a low priority with more funds compared to the case of straight priorities.

REFERENCES

1. Burkov, V.N. and Novikov, D.A., *Kak upravlyat' proektami* (Project Management), Moscow: Sinteg, 1997.

2. Novikov, D.A., Puzyrev, S.A., and Khorokhordina, N.V., Joint Financing Mechanisms, *Sist. Upravlen. Inform. Tekhn.*, 2009, no. 2 (20), pp. 71–72.

3. Burkov, V.N., Burkova, I.V., Goubko, M.V., et al., *Mekhanizmy upravleniya* (Control Mechanisms), Novikov, D.A., Ed., Moscow: LENAND, 2013.

4. Ivashchenko, A.A., Kolobov, D.V., and Novikov, D.A., *Mekhanizmy finansirovaniya innovatsionnogo razvitiya firmy* (Financing Mechanisms for Firm's Innovative Development), Moscow: Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2005.

5. Burkov, V.N., *Osnovy matematicheskoi teorii aktivnykh sistem* (Foundations of the Mathematical Theory of Active Systems), Moscow: Nauka, 1977.

6. Burkov, V., Goubko, M., Korgin, N., and Novikov, D., *Introduction to Theory of Control in Organizations*, Boca Raton: CRC Press, 2015.

7. Habbel, V., Jackson, E., Orth, M., et al., Evaluating Blended Finance Instruments and Mechanisms: Approaches and Methods, *OECD Development Co-operation Working Papers*, no. 101, Paris: OECD Publishing, 2021.

8. Andersen, W.O., Basile, I., Gotz, G., et al., Blended Finance Evaluation: Governance and Methodological Challenges, *OECD Development Co-operation Working Papers*, no. 51, Paris: OECD Publishing, 2019. https://dx.doi.org/10.1787/4c1fc76e-en

9. Pereira, J., Blended Finance: What Is It, How It Works and How It Is Used, *Oxfam International*, February 13, 2017. https://www.oxfam.org/en/research/blended-finance-what-it-how-it-works-and-how-it-used

10. Chainz, Ch. and Hakenes, H., The Politician and His Banker—How to Efficiently Grant State Aid, *Journal of Public Economics*, 2012, vol. 96, pp. 218–225.

11. Chen, J., Risk-Adjusted Return, *Investopedia*, December 20, 2018. https://www.investopedia.com/terms/r/riskadjustedreturn.asp

12. Burkov, V.N. and Shchepkin, A.V., Pricing Mechanisms for Cost Reduction under Budget Constraints, *Control Sciences*, 2021, vol. 3, pp. 37–43. http://doi.org/10.25728/cs.2021.3.5

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1468–1477.

An Attracting Cycle in a Coupled Mechanical System with Phase Shifts in Subsystem Oscillations

V. N. Tkhai

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
e-mail: tkhai@ipu.ru

Received April 24, 2023
Revised September 9, 2023
Accepted September 30, 2023

Abstract—This paper considers the set of reversible mechanical systems with single-period oscillations and individual phase shifts in them. The problem of aggregating a coupled system with an attracting cycle is solved. The approach developed below is to choose a leader (control) system that acts on the other (follower) systems through one-way coupling control: in an aggregated system, there are no links between follower systems. Universal coupling controls are used. Particular attention is paid to conservative systems. Possible scenarios for the operation of the aggregated system are presented.

Keywords: reversible mechanical system, symmetric periodic motions, coupling controls, leader system, follower system, attracting cycle, stabilization

DOI: 10.25728/arcRAS.2024.26.46.001

1. PRELIMINARIES

Models containing coupled subsystems are studied in various fields of knowledge. In mechanics, A. Sommerfeld's sympathetic pendulums have become such a (classical) model. Other examples were given, e.g., in [1–5].

Aggregation consists in constructing a coupled system from a given set of systems so that the resulting whole will possess a desired dynamic property. In oscillation stabilization, this property is achieved, in particular, in an attracting cycle of the system. Aggregation occurs by finding appropriate coupling controls between the systems.

In the paper [6], the aggregation problem was solved for a set of conservative systems. According to [6, Lemma 1], there exists a cycle in the system only if all mechanical systems, possibly except one system with a degenerate family of oscillations, contain nondegenerate families of oscillations. Aggregation was carried out for systems containing nondegenerate families of oscillations that also form a nondegenerate family in an uncoupled system as a whole. The case of phase-synchronized oscillations in systems was considered. The universal control from [7] was applied.

At the same time, it is of definite interest to study oscillation modes in which, e.g., the phases in system oscillations are equidistant from each other or neighbor systems oscillate in antiphase. Therefore, a common problem is to find coupling controls for implementing an attracting cycle of the coupled mechanical system with phase shifts in the oscillations of its constituent systems. Also, it seems interesting to aggregate a coupled system containing one or more mechanical systems with degenerate families of oscillations. Thus, we arrive at the following general problem statement: aggregate a coupled system with an attracting cycle on a set of mechanical systems admitting oscillations.

Note that some approaches to aggregating a general-form autonomous system with an attracting cycle were proposed in [8]; Lyapunov method-based procedures for aggregating a complex system were described in [9].

2. PROBLEM STATEMENT

Consider the set Ξ of n smooth reversible mechanical systems with one degree of freedom

$$\ddot{q}_s + f_s(q_s, \dot{q}_s) = 0, \quad f_s(q_s, -\dot{q}_s) = f_s(q_s, \dot{q}_s), \quad s = 1, \ldots, n. \tag{1}$$

The phase portrait of the sth system is symmetric with respect to the fixed set $M_s = \{q_s, \dot{q}_s \colon \dot{q}_s = 0\}$, where q_s denotes the generalized coordinate. By assumption, each system of the set Ξ admits a single-frequency oscillation. It will be symmetric with respect to the set M_s and represents a symmetric periodic motion (SPM). The SPM is described by the formula

$$q_s = \varphi_s(h_s, t + \gamma_s), \quad s = 1, \ldots, n,$$

where the period $T_s(h_s)$ depends on the parameter h_s and the parameter γ_s specifies the time shift of the initial point: for $\gamma_s = 0$, the initial point belongs to the fixed set M_s. In this case, the SPM is described by a function even in the variable t. SPMs always form families. In a conservative system, the function f_s does not depend on the velocity $\dot{q}_s$.

Further considerations involve a definition from [6].

Definition 1. A family of SPMs in the parameter h is said to be nondegenerate if the derivative of the period $T(h)$ with respect to the variable h differs from zero on this family. An SPM of a nondegenerate family is called nondegenerate as well.

The period $T(h)$ on the family of nondegenerate SPMs can increase or decrease. For example, the period of oscillations of a mathematical pendulum monotonically increases with the energy of the pendulum, and the oscillations are nondegenerate. The solutions of the equation $\ddot{x} + x^3 = 0$ belong to the family of SPMs with a decreasing period.

The oscillations of a linear oscillator are isochronous and form a degenerate family of SPMs. As a rule, a degenerate SPM of a nonlinear system is on the boundary of the family of its nondegenerate SPMs. In a conservative system, the parameter h is usually the constant of the energy integral.

The general problem statement involves the set Ξ of reversible mechanical systems containing nondegenerate (and/or degenerate) families of SPMs with increasing (and/or decreasing) periods on the family. In this case, if the set Ξ simultaneously includes a system with an increasing period ($dT_1/dh_1 > 0$) and a system with a decreasing period ($dT_2/dh_2 < 0$), the period curves will intersect at one point where $T_1(h_1^*) = T_2(h_2^*) = T^*$. The phases of oscillations generally differ. If the set Ξ also includes a system with a degenerate family, the period on it will equal T^* as well. The set of three equations in Ξ leads in the coupled system to two arbitrary phases of oscillations in the systems. Given an arbitrary number n of equations in Ξ, it is assumed that $\gamma_s = \gamma_1 + \delta_s$, $s = 2, \ldots, n$. Therefore, we pose the problem of aggregating a coupled system with an attracting cycle for all possible vectors $\delta = (\delta_2, \ldots, \delta_n)$.

Further analysis focuses on an autonomous coupled mechanical system of the form

$$\ddot{q}_s + f_s(q_s, \dot{q}_s) = \varepsilon \sigma_s u_s(q, \dot{q}), \quad s = 1, \ldots, n, \tag{2}$$

where the coupling control

$$u(q, \dot{q}) = (u_1(q, \dot{q}), \ldots, u_n(q, \dot{q})) \tag{3}$$

acts with a small gain ε: the switches σ_s are $+1$ or (-1). By assumption, for $\varepsilon = 0$, system (2), treated as a whole, admits a T^*-periodic SPM. The problem is to find the coupling control (3) ensuring the existence of an attracting cycle with the period T^* in system (2).

This problem covers the following special cases:

1) All reversible mechanical systems in the set Ξ admit a family of nondegenerate SPMs with an increasing (decreasing) period.

2) The monotonicity of the period in the systems differs by character.

3) The set of mechanical systems contains nondegenerate and degenerate families of SPMs.

In [6], case 1) was investigated for conservative systems under the additional assumption that the set of uncoupled systems as a whole admits a nondegenerate family of SPMs.

3. UNIVERSAL COUPLING CONTROLS UNDER $\delta \neq 0$

For the vector $\delta \neq 0$, we find universal coupling controls ensuring the existence and orbital asymptotic stability of the cycle of system (2). Such coupling controls can be treated as a generalization of the couplings from [7].

The proposed coupling controls have the form

$$\begin{aligned} u_1 &= [1 - K_1(h_1)q_1^2]\dot{q}_1, \\ u_j &= [1 - K_j(h_j, \delta_j)q_1^2]\dot{q}_j, \quad j = 2, \dots, n. \end{aligned} \tag{4}$$

The functions $K_1(h_1)$ and $K_j(h_j, \delta_j)$ are calculated below.

By assumption, for $\varepsilon = 0$ system (2) admits a T^*-periodic SPM and the corresponding values in the subsystems are $h_s = h_s^*$, $s = 1, \dots, n$. According to formulas (4), the equations in (2) become unequal: we construct a controlled coupled system in which the system with number $s = 1$ is the leader and the other systems are followers. Another feature of the controls (4) is that the subsystems with numbers $s = 2, \dots, n$ have no direct influence on each other. Due to these remarks, we analyze $(n - 1)$ independent subsystems of the same type:

$$\begin{aligned} \ddot{q}_1 + f_1(q_1, \dot{q}_1) &= \varepsilon\sigma_1[1 - K_1(h_1^*)q_1^2]\dot{q}_1, \\ \ddot{q}_j + f_j(q_j, \dot{q}_j) &= \varepsilon\sigma_j[1 - K_j(h_j^*, \delta_j)q_1^2]\dot{q}_j, \quad j = 2, \dots, n. \end{aligned} \tag{5}$$

For the subsystem with number j in (5), we solve the cycle problem under the condition $\varepsilon \neq 0$. Then, applying the obtained result to all subsystems with numbers $j = 2, \dots, n$, we come to the solution of the cycle problem for the coupled system. In system (5), $K_1(h_1^*)$ and $K_j(h_j^*)$ denote values. In addition, h_1^* and h_j^* mean that the controls are chosen for the SPM with the period T^* and the corresponding values $h_1 = h_1^*$ and $h_j = h_j^*$. On the other hand, when solving the control problem in (5) for another pair (h_1, h_j), a different pair of the coefficients $(K_1(h_1), K_j(h_j, \delta))$ is chosen: in (6), the control is applied with some changed coefficients K_1 and K_j. Hence, we design an adaptive control system in (6).

Thus, for the adaptive control system (6), it is required to find the relationships $K_1(h_1)$ and $K_j(h_j, \delta)$ (the second with the parameter δ) ensuring the existence of an attracting cycle.

For the subsystem with number j, we write the system of amplitude equations

$$\begin{aligned} I_1(h_1) &\equiv \int_0^{T^*} [1 - K_1(h_1^*)\varphi_1^2(h_1, t)]\dot{\varphi}_1(h_1, t)\psi_1(h_1, t)dt = 0, \\ I_j(h_1, h_j, \delta_j) &\equiv \int_0^{T^*} [1 - K_j(h_j^*, \delta_j)\varphi_1^2(h_1, t)]\dot{\varphi}_j(h_j, t + \delta_j)\psi_j(h_j, t + \delta_j)dt = 0. \end{aligned} \tag{6}$$

These equations are used to find $h_1 = h_1^*$ and $h_j = h_j^*$ that meet the necessary conditions for the existence of a cycle with the period T^* in the controlled system (5). In (6), $(\psi_1(h_1, t), \psi_j(h_j, t + \delta_j))$

denotes the solution of the adjoint equation for $q_1 = \varphi_1(h_1, t)$ and $q_j = \varphi_j(h_j, t + \delta_j)$. This solution is calculated in the Appendix.

The first equation in system (6) is the same for all numbers j. It can be analyzed independently of the second one.

We begin with the first equation of (5). The necessary conditions for the existence of a cycle must hold for all values of the parameter h_1 and the corresponding values of the period $T_1(h_1)$. Therefore,

$$\int\limits_0^{T_1(h_1)} [1 - K_1(h_1)\varphi_1^2(h_1, t)]\dot{\varphi}_1(h_1, t)\psi_1(h_1, t)dt \equiv 0, \tag{7}$$

which gives

$$K_1(h_1) = \frac{\int\limits_0^{T_1(h_1)} \dot{\varphi}_1(h_1, t)\psi_1(h_1, t)dt}{\int\limits_0^{T_1(h_1)} \varphi_1^2(h_1, t)\dot{\varphi}_1(h_1, t)\psi_1(h_1, t)dt}.$$

The denominator of this expression does not vanish; for the case of a conservative system, see Section 4. In the general case of a reversible mechanical system, this result follows from the solution of the adjoint equation calculated in the Appendix.

In view of the odd function $\dot{\varphi}_1(h_1, t)$ and the equality $T_1(h_1^*) = T^*$, we determine the derivative of the function $I_1(h_1)$ at the point $h_1 = h_1^*$ from identity (7):

$$\frac{dI_1(h_1^*)}{dh_1} = \chi_1\nu_1,$$

$$\chi_1 = \frac{dK_1(h_1^*)}{dh_1}, \quad \nu_1 = \int\limits_0^{T^*} \varphi_1(h_1^*, t)^2\dot{\varphi}_1(h_1^*, t)\psi_1(h_1^*, t)dt.$$

The equality $I(h_1^*) = 0$ means that the necessary condition for the existence of a T^*-periodic solution holds in the first equation of system (5). Due to the inequality $\chi_1\nu_1 \neq 0$, this solution is a cycle, which becomes attracting under an appropriately chosen sign of σ_1 (see [7]).

The second equation of system (6) is considered by analogy. We define the function

$$K_j(h_j, \delta_j) = \frac{\int\limits_0^{T_j(h_j)} \dot{\varphi}_j(h_j, t + \delta_j)\psi_j(h_j, t + \delta_j)dt}{\int\limits_0^{T_j(h_j)} \varphi_1^2(h_1^*, t)\dot{\varphi}_j(h_j, t + \delta_j)\psi_j(h_j, t + \delta_j)dt}$$

and calculate the derivative

$$\frac{dI_j(h_1^*, h_j^*, \delta_j)}{dh_j} = \chi_j\nu_j,$$

$$\chi_j = \frac{dK_j(h_j^*, \delta_j)}{dh_j}, \quad \nu_j = \int\limits_0^{T^*} \varphi_1(h_1^*, t)^2\dot{\varphi}_j(h_j^*, t + \delta_j)\psi_j(h_j^*, t + \delta_j)dt$$

for $h_j = h_j^*$ $(h_1 = h_1^*)$.

The conditions $\chi_1\nu_1 \neq 0$ and $\chi_j\nu_j \neq 0$ are now sufficient for the existence of a simple root (h_1^*, h_j^*) of the system of amplitude equations (6) with a fixed number j. Then the simplicity of this root ensures the existence of a cycle in system (5) with a fixed number j. The cycle will be attracting if the switches are chosen from the conditions $\sigma_1\chi_1\nu_1 < 0$ and $\sigma_j\chi_j\nu_j < 0$.

Consider the systems of amplitude equations (6) for all numbers $j = 2, \ldots, n$. Then, under the inequalities $\chi_s\nu_s \neq 0$, $s = 1, \ldots, n$, a cycle is implemented in the coupled system (5). Given the additional condition $\sigma_s\chi_s\nu_s < 0$, $s = 1, \ldots, n$, the cycle becomes attracting.

Thus, the following result is true.

Theorem 1. *Assume that the set of reversible mechanical systems with one degree of freedom admits a T^*-periodic motion. Then the coupled mechanical system* (5)*, where $j = 2, \ldots, n$, has a unique cycle of the period T^* if $\chi_s\nu_s \neq 0$, $s = 1, \ldots, n$. Under the additional conditions $\sigma_s\chi_s\nu_s < 0$, $s = 1, \ldots, n$, the cycle becomes attracting.*

Remark 1. The cycle of the coupled system (5) is determined within an arbitrary shift on the trajectory. The cycle-generating oscillations have the phase shifts $\delta_2, \ldots, \delta_n$ with respect to the phase of the oscillation in the first equation of system (5).

Remark 2. In system (6), the integral

$$\kappa_j = \int_0^{T^*} \dot{\varphi}_j(h_j^*, \tau + \delta_j)\psi_j(h_j^*, \tau + \delta_j)d\tau$$

does not depend on δ_j on the period. Therefore, for $\kappa_j \neq 0$, we define the function

$$K_j(h_j^*, \delta_j) = \frac{\kappa_j}{\int_0^{T^*} \varphi_1^2(h_1^*, \tau - \delta_j)\dot{\varphi}_j(h_j^*, \tau)\psi_j(h_j^*, \tau)d\tau}, \tag{8}$$

which will be T^*-periodic in δ_j.

Remark 3. In formula (8), the nonzero denominator defines the admissible range of the phase shift δ_j in the jth subsystem of system (5).

Remark 4. Theorem 1 designs the piecewise continuous system (5). Since the amplitude equations (6) are independent of σ_j, there exists a cycle in every smooth switchless system. The attraction conditions ($\chi_s\nu_s \neq 0$) must hold in the subsystem on the trajectories with both $h_s > h_s^*$ and $h_s < h_s^*$. Therefore, the signs of σ_s for these trajectories usually differ. An example of a switch control law was provided in [10].

4. CONSERVATIVE SYSTEMS

For the set of conservative systems, the functions f_s in (1) are independent of the velocities $\dot{q}_s$, and each system admits an energy integral under $\varepsilon = 0$. The variational equations for SPMs contain a symmetric matrix; therefore, the one-degree-of-freedom system under consideration satisfies the equations

$$\psi_s(h_s^*, \tau + \delta_s) = -\dot{\varphi}_s(h_s^*, \tau + \delta_s), \quad s = 1, \ldots, n \quad (\delta_1 = 0).$$

Consequently, $\nu_s > 0, \kappa_s < 0,\ s = 1, \ldots, n$.

The integrand in (8) is $(T^*/2)$-periodic on δ and two points symmetric with respect to the fixed set correspond to each value $K_j(h_j^*, \delta^*)$. In turn, these points implement one cycle.

Thus, we arrive at the following result.

Theorem 2. *For the set of conservative systems with one degree of freedom that admits a T^*-periodic SPM, the coupled system* (5) *has a unique attracting cycle under the conditions $\sigma_s\chi_s < 0$, $s = 1, \ldots, n$.*

Example 1. In the coupled system

$$\begin{aligned}\ddot{x} + \sin x &= \varepsilon(1 - K_x(h_x)x^2)\dot{x},\\ \ddot{y} + y^3/4 &= \sigma\varepsilon(1 - K_y(h_y,\delta)x^2)\dot{y}\end{aligned} \tag{9}$$

with $\varepsilon = 0$, the first equation describes a mathematical pendulum. Starting at 2π, the period $T_x(h_x)$ grows monotonically with the pendulum energy h_x on the family of oscillations, and the function $K_x(h_x)$ is monotonically decreasing (see [11]). The solutions of the second equation form a family of oscillations with the period $T_y(h_y)$ representing a decreasing function of the constant energy h_y.

Indeed, the period $T_y(h_y)$ is given by

$$T_y(h_y) = 2\int\limits_{-y(0)}^{y(0)} \frac{dy}{\sqrt{h - y^4}},$$

where $y(0)$ denotes the initial value of the variable y. Passing to the variable $z = y/h_y^{1/4}$ yields

$$T_y(h_y) = \frac{2}{h_y^{1/4}}\int\limits_{1}^{-1} \frac{dz}{\sqrt{1 - z^4}} = \frac{a}{h_y^{1/4}}, \quad a = 4 * 1.3\ldots,$$

an explicit-form relationship between the period and the system energy.

According to the analysis above, for any h_x^* and $T_x(h_x^*) > 2\pi$, there exists a value h_y^* such that $T_x(h_x^*) = T_y(h_y^*)$ (the equality of periods in (9)) with an increasing function f, i.e., $h_y^* = f(h_x^*)$. Hence, system (9) with $\varepsilon = 0$ admits a one-parameter family of SPMs with the parameter h_x^*.

The function $K_x(h_x)$ is monotonically decreasing. Therefore, for $h_x = h_x^*$, the first equation in (9) has an attracting cycle. By Theorem 2, the additional condition $dK_y(h_y^*,\delta)/dh_y \neq 0$ leads to an attracting cycle of the coupled system (9).

Thus, for any oscillation of the mathematical pendulum corresponding to the energy value h_x^*, there exists an energy value h_y^* of the second equation in (9) such that an attracting cycle is implemented in the coupled system. Moreover, the phases in the oscillations of the equations differ by the desired value δ.

5. THE CASE OF A DEGENERATE FAMILY OF SPMS

Under identical phases in the oscillations of subsystems, a cycle in the coupled system exists only if all mechanical systems, possibly except one system with a degenerate family of SPMs, contain nondegenerate families of SPMs. This result was established in [6, Lemma 1]. In what follows, we investigate in detail the case where one of the families is degenerate. Assume that the oscillations in the systems are not phase-synchronized and $n = 2$ in system (5).

Consider the system

$$\begin{aligned}\ddot{x} + x &= \varepsilon(1 - K_x(h_x)x^2)\dot{x},\\ \ddot{y} + f(y) &= \varepsilon\sigma(1 - K_y(h_y,\delta)x^2)\dot{y},\end{aligned} \tag{10}$$

in which the first equation contains a degenerate family of oscillations under $\varepsilon = 0$ and the period of oscillations in the second equation monotonically depends on the energy h_y. The solution of the uncoupled system is described by the formulas $x = A_x \cos t$ and $y = \varphi(h_y, t + \delta)$. On the generating solution, we have $A_x = 2/\sqrt{K_x}$, and the period of oscillations 2π corresponds to the constant h_y^* in the second equation. Let us find the relationship between $K_y(h_y,\delta)$ and K_x.

For system (10), formula (8) is written as

$$K_y(h_y^*,\delta)=-\frac{\kappa}{\int\limits_0^{T^*}A_x^2\cos^2 t\dot\varphi^2(h_y^*,t+\delta)dt},$$
$$\kappa=-\int\limits_0^{T^*}\dot\varphi^2(h_y^*,t)dt,\quad T^*=2\pi. \tag{11}$$

The integral in the denominator can be transformed as follows:

$$\frac{1}{2}\int\limits_0^{T^*}(1+\cos 2t)\dot\varphi^2(h_y^*,t+\delta)dt=\frac{1}{2}\int\limits_{\delta}^{T^*+\delta}\dot\varphi^2(h_y^*,\tau)d\tau$$
$$+\frac{1}{2}\left(\cos 2\delta\int\limits_{\delta}^{T^*+\delta}\cos 2\tau\dot\varphi^2(h_y^*,\tau)d\tau+\sin 2\delta\int\limits_{\delta}^{T^*+\delta}\sin 2\tau\dot\varphi^2(h_y^*,\tau)d\tau\right).$$

Consider the bracketed expression above; here, the first integral of a 2π-periodic function does not depend on δ on the period, and the second integral is taken for an odd function (and vanishes accordingly). Hence, we obtain a linear function of $\cos 2\varphi$, and $K_y(h_y^*,\delta)$ is given by an even π-periodic function of δ.

Due to the equality $K_x=4/A_x^2$, formula (11) reduces to

$$K_y(h_y^*,\delta)=\frac{K_x\int\limits_0^{T^*}\dot\varphi^2(h_y^*,t)dt}{2\int\limits_0^{T^*}(1+\cos 2\delta\cos 2t)\dot\varphi^2(h_y^*,t)dt}. \tag{12}$$

The derivative of (12) is an odd π-periodic function of δ. On the interval $\delta\in(-\pi/2,\pi/2)$, this derivative vanishes for $\delta=0$.

The cycle of the coupled system (2) can be constructed using Theorem (10). The characteristic $K_y(h_y,\delta)$ is calculated for a given function $\varphi(h_y,t)$. For a mathematical pendulum, the function $K_y(h_y,\delta)$ is monotonically decreasing under $\delta=0$ (see [11]).

Example 2. Consider system (10) in which the function $K_y(h_y,\delta)$ in the second equation is independent of δ and coincides with K_x. Let this system be applied in the mechatronic oscillation stabilization scheme proposed in [12]. More precisely put, the amplitude A_x at the point $\delta=\delta^*$ is selected to adjust the mode of satisfying the equality $K_y(h_y^*,\delta)=K_x=4/A_x^2$. As a result, we obtain a possible scenario for the birth of a cycle described in [10]. Note that the existence of the scenario was proved by analyzing the second equation in (10) by substituting the generating solution of the first equation. In the mechatronic stabilization scheme, the time shift δ^* between the oscillations of the van der Pol oscillator and the mechanical system is given by (12).

6. THE CASE OF TWO DEGENERATE FAMILIES

In system (10), the van der Pol oscillator is used to generate signals for a mechanical system admitting a nondegenerate family of oscillations. The system is designed to stabilize mechanical oscillations. For $K_y=K_x$, the shift δ in the solutions of the equations of system (10) is given by formula (12); see Section 5.

It seems interesting to analyze how the amplitude and phase of oscillations in the leader and follower systems are synchronized in the cycle of the coupled system. We consider this problem for equal systems in Ξ, on an example of two identical linear oscillators. Then the coupled system has the form

$$\begin{aligned}\ddot{x}+x&=\varepsilon(1-K_x(h_x)x^2)\dot{x},\\ \ddot{y}+y&=\varepsilon\sigma(1-K_y(h_y,\delta)x^2)\dot{y},\end{aligned}\tag{13}$$

where the first equation describes the van der Pol oscillator and the second equation becomes the follower for this oscillator.

Under $\varepsilon=0$, system (13) oscillates in each coordinate with a frequency of 1: the oscillations are isochronous. The generating oscillations are given by

$$x=A_x\cos t,\quad A_x=2/\sqrt{K_x},\quad y=A_y\cos(t+\delta).$$

For the second equation in (13), we calculate $\kappa=-\int\limits_0^{2\pi}A_y^2\sin^2 t dt=-\pi A_y^2$.

In the coupled system, $K_y=K_y(h_y,\delta)$; therefore, formula (8) yields

$$K_y(h_y,\delta)=-\frac{4\kappa}{A_x^2A_y^2\pi(2-\cos 2\delta)}=\frac{4}{A_x^2(2-\cos 2\delta)}=\frac{K_x}{2-\cos 2\delta}.$$

Hence, in the cycle of the coupled system, the amplitudes of oscillations in the leader and follower systems are synchronized ($K_y=K_x$) only under $\delta=0$; phase synchronization also occurs under $\delta=0$.

According to the formula $K_y(h_y,\delta)=2/(h_y(2-\cos 2\delta))$, the conditions for the existence of a cycle in the coupled system (10) hold everywhere in δ. The control law $\sigma=1$ is chosen for the attracting cycle.

Note that the amplitudes of oscillations in the systems of the coupled system (13) are close to linear oscillations. Therefore, the oscillations of the systems will appear to be synchronized in δ in the cycle (operating mode) of the coupled system (13) under consideration regardless of the shift δ.

7. CONCLUSIONS

This paper has proposed an approach to aggregating a coupled system with an attracting cycle on a given set n of reversible mechanical systems with oscillations. Within the approach, a leader (control) system is selected to act on the other (follower) systems through one-way coupling control: in an aggregated system, there are no links between follower systems. The coupled system oscillates as $(n-1)$ independent subsystems controlled by the leader system. In addition, the oscillation of each system may have an individual phase shift with respect to the phase of the oscillation in the leader system.

Different control scenarios are possible for the aggregated system. If the subsystems have no phase shift, the *simultaneous control* scenario is implemented for $(n-1)$ mechanical systems; see [6]. The *conveyor* scenario is implemented in the controlled coupled system when specifying a shift change law for $(n-1)$ mechanical systems: for example, the maximum amplitude of oscillations in the follower systems is achieved at different time instants. For $n=2$, the *leader–follower* scenario is implemented, a common one described, e.g., in [12] for a mechatronic oscillation stabilization scheme.

The aggregation approach has been presented on an example of reversible mechanical systems in the plane. It remains valid for a set of mechanical systems of arbitrary dimension. The constructed coupled system represents one level of the hierarchy of a multilevel aggregated system with an attracting cycle (for details, see [8]).

APPENDIX

The adjoint solution can be calculated using Lemma 1.

Consider a smooth reversible mechanical system of the second order:

$$\dot{u} = U(u,v), \quad \dot{v} = V(u,v), \quad U(u,-v) = -U(u,v), \quad V(u,-v) = V(u,v).$$

Let this system admit an SPM described by the functions

$$u = \varphi(t), \quad v = \theta(t), \quad \varphi(-t) = \varphi(t), \quad \theta(-t) = -\theta(t).$$

The variational equations for the SPM have the form

$$\begin{aligned} \dot{x} &= a_-(t)x + a_+(t)y, \\ \dot{y} &= b_+(t)x + b_-(t)y, \end{aligned} \tag{A.1}$$

where $a_\pm(t), b_\pm(t)$ denote even (+) and odd (–) periodic functions. They have the solution $x = \dot{\varphi}(t), y = \dot{\theta}(t)$.

Lemma 1. *For a given SPM, the solution of the system adjoint to* (A.1) *is calculated by constructive formulas.*

Proof. Let us apply the transformation

$$x = \xi_+(t)\tilde{x}, \quad y = \eta_+(t)\tilde{y}$$

with even periodic functions $\xi_+(t)$ and $\eta_+(t)$ with nonzero means. As a result,

$$\begin{aligned} \xi_+(t)\dot{\tilde{x}} + \dot{\xi}_+(t)\tilde{x} &= a_-(t)\xi_+(t)\tilde{x} + a_+(t)\eta_+(t)\tilde{y}, \\ \eta_+(t)\dot{\tilde{y}} + \dot{\eta}_+(t)\tilde{y} &= b_+(t)\xi_+(t)\tilde{x} + b_-(t)\eta_+(t)\tilde{y}. \end{aligned}$$

The functions $\xi_+(t)$ and $\eta_+(t)$ are appropriately chosen to satisfy the equalities

$$\dot{\xi}_+ = a_-(t)\xi_+, \quad \dot{\eta}_+ = b_-(t)\eta_+.$$

Then the transformed system

$$\dot{\tilde{x}} = \tilde{a}_+(t)\tilde{y}, \quad \dot{\tilde{y}} = \tilde{b}_+(t)\tilde{x} \tag{A.2}$$

contains no odd functions of t.

The adjoint system of

$$x_1 = \xi_{1+}(t)\tilde{x}_1, \quad y_1 = \eta_{1+}(t)\tilde{y}_1$$

is transformed by analogy. We obtain

$$\dot{\tilde{x}}_1 = -\tilde{b}_+(t)\tilde{y}_1, \quad \dot{\tilde{y}}_1 = -\tilde{a}_+(t)\tilde{x}_1. \tag{A.3}$$

In the variables $\tilde{x}_1 = -\tilde{y}$ and $\tilde{y}_1 = \tilde{x}$, the resulting system (A.3) coincides with (A.2). Hence, its solution is given by $\tilde{x}_1 = -\xi_+(t)^{-1}\dot{\theta}(t)$, $\tilde{y}_1 = \eta_+(t)^{-1}\dot{\varphi}(t)$. Therefore, the solution of the adjoint system can be written as

$$x_1 = -\xi_{1+}(t)\xi_+(t)^{-1}\dot{\theta}(t), \quad y_1 = \eta_{1+}(t)\eta_+(t)^{-1}\dot{\varphi}(t).$$

The proof of Lemma 1 is complete.

REFERENCES

1. Morozov, N.F. and Tovstik, P.E., Transverse Rod Vibrations under a Short-Term Longitudinal Impact, *Doklady Physics*, 2013, vol. 58, no. 9, pp. 387–391.

2. Kovaleva, A. and Manevitch, L.I., Autoresonance Versus Localization in Weakly Coupled Oscillators, *Physica D: Nonlinear Phenomena*, 2016, vol. 320, pp. 1–8.

3. Kuznetsov, A.P., Sataev, I.R., and Turukina, L.V., Forced Synchronization of Two Coupled Van der Pol Self-Oscillators, *Rus. J. Nonlin. Dyn.*, 2011, vol. 7, no. 3, pp. 411–425.

4. Rompala, K., Rand, R., and Howland, H., Dynamics of Three Coupled Van der Pol Oscillators with Application to Circadian Rhythms, *Communicat. Nonlin. Sci. Numerical Simulation*, 2007, vol. 12, no. 5, pp. 794–803.

5. Yakushevich, L.V., Gapa, S., and Awrejcewicz, J., Mechanical Analog of the DNA Base Pair Oscillations, *Proc. 10th Conf. on Dynamical Systems Theory and Applications*, Lodz: Left Grupa, 2009, pp. 879–886.

6. Barabanov, I.N. and Tkhai, V.N., Stabilization of a Cycle in a Coupled Mechanical System, *Autom. Remote Control*, 2022, vol. 83, no. 1, pp. 54–61.

7. Tkhai, V.N., Stabilizing the Oscillations of a Controlled Mechanical System, *Autom. Remote Control*, 2019, vol. 80, no. 11, pp. 1996–2004.

8. Tkhai, V.N., Aggregation of an Autonomous System with an Attracting Cycle, *Autom. Remote Control*, 2022, vol. 83, no. 3, pp. 332–342.

9. Aleksandrov, A.Yu. and Platonov, A.V., *Metod sravneniya i ustoichivost' dvizhenii nelineinykh sistem* (A Comparison Method and Stability of Motions of Nonlinear Systems), St. Petersburg: St. Petersburg State University, 2012.

10. Tkhai, V.N., Cycle Mode in a Coupled Conservative System, *Autom. Remote Control*, 2022, vol. 83, no. 2, pp. 237–251.

11. Tkhai, V.N., On Stabilization of Pendulum Type Oscillations of a Rigid Body, *Proc. 2018 14th Int. Conf. on Stability and Oscillations of Nonlinear Control Systems (STAB'2018, Pyatnitskiy's Conference).* https://ieeexplore.ieee.org/document/8408408. https://doi.org/10.1109/STAB.2018.8408408.

12. Tkhai, V.N., A Mechatronic Scheme to Stabilize Oscillations, *J. Comput. Syst. Sci. Int.*, 2022, vol. 61, no. 1, pp. 9–15.

This paper was recommended for publication by A.M. Krasnosel'skii, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 12, pp. 1478–1488.
© The Author(s), 2023 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2023.
Russian Text © The Author(s), 2023, published in Avtomatika i Telemekhanika, 2023, No. 12, pp. 133–145.

An Interval Observer-Based Method to Diagnose Discrete-Time Systems

A. N. Zhirabok[*,**,a] **and A. V. Zuev**[*,**,b]

[*]*Far Eastern Federal University, Vladivostok, Russia*
[**]*Institute of Marine Technology Problems, Far Eastern Branch,*
Russian Academy of Sciences, Vladivostok, Russia
e-mail: [a]*zhirabok@mail.ru,* [b]*alvzuev@yandex.ru*

Received August 18, 2023
Revised October 23, 2023
Accepted November 1, 2023

Abstract—This paper proposes a method for diagnosing linear dynamic systems described by discrete-time models with exogenous disturbances based on interval observers. Formulas are derived to construct an interval observer producing two values of the residual as follows: if zero is between these values, then the system has no faults to be detected by the observer. The case where zero does not belong to the interval between these values is qualified as the occurrence of a fault. The theoretical results are illustrated by an example.

Keywords: linear systems, discrete-time models, interval observers, diagnosis, faults

DOI: 10.25728/arcRAS.2024.53.21.001

1. INTRODUCTION AND PROBLEM STATEMENT

This paper further develops the works [1, 2], devoted to the design of interval observers for systems described by linear models with exogenous disturbances. The corresponding problem has been actively studied in recent years. Overviews of the current results can be found in [3, 4]; the solutions for various classes of systems and related applications, in [5–10]. Characteristic features of the cited research are as follows. First, the interval observer has a dimension coinciding with that of the original system; second, the set of admissible values of the full state vector is estimated. At the same time, it may be of theoretical and practical interest to obtain such an estimate only for a given linear function of the state vector. The corresponding interval observer may be considerably simpler than the full-dimensional observer and the resulting interval may have an appreciably smaller width.

In [11–16], interval observers were used to perform functional diagnosis. The authors [11–13, 15] designed the observer based on the original system, which led to cumbersome constructs and complicated methods for minimizing the influence of exogenous disturbances on the diagnosis process. The paper [14] considered the diagnosis problem in a family of coupled subsystems: for each subsystem, a particular interval observer of full dimension was constructed. In [16], a practical problem was solved based on a special-form interval observer.

As is known, an adaptive threshold is traditionally used to reduce the probabilities of false alarms and fault omissions during diagnosis. This threshold sets lower and upper bounds for the residual in the absence of faults. Although the concept of an adaptive threshold appeared more than 30 years ago, it has been developed for various classes of systems up to the present time; for example, see [17, 18]. In these works, the residual was generated by a diagnostic observer whereas the adaptive threshold was formed separately. Such an approach leads to rather complicated diagnostic schemes.

In contrast, due to its specifics, an interval diagnostic observer produces only two values of the residual, which significantly simplifies the scheme. In addition, the residuals are formed so that in the absence of faults, the values of one residual are nonpositive and those of the other are nonnegative. In other words, if zero lies between these values, then the system has no faults to be detected by the observer. The case where zero does not belong to the interval between these values is qualified as the occurrence of a fault. In addition, unlike traditional adaptive threshold schemes, the values of residuals produced by the interval observer are independent of the control and output signals of the diagnosed system. This property also simplifies the decision process based on the diagnosis results.

In this paper, we construct minimal-dimension interval observers for discrete time-invariant systems described by linear dynamic models operating under exogenous disturbances in order to solve the problems of functional diagnosis (fault detection and isolation). In [1, 2], interval observers were used to estimate the values of a given linear function of the state vector of the original system. In contrast to [1, 2], in accordance with the diagnosis task, we change the observer structure and also consider several related issues: methods to maximize sensitivity to faults and isolate them.

Consider a class of systems with the linear discrete-time model

$$\begin{aligned} x(t+1) &= Fx(t) + Gu(t) + Dd(t) + L\rho(t), \\ y(t) &= Hx(t), \end{aligned} \tag{1.1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $y \in \mathbb{R}^l$ denote the state, control, and output vectors, respectively; F, G, H, L, and D are given constant matrices; $\rho(t) \in \mathbb{R}^q$ is an unknown bounded time-varying function describing exogenous disturbances of the system, i.e., $\underline{\rho} \leqslant \rho(t) \leqslant \overline{\rho}$ with given values $\underline{\rho}$ and $\overline{\rho}$. In many cases, system faults occur due to unacceptable changes in system parameters. Therefore, we assume that the variations of the function $d(t) \in \mathbb{R}^p$ within $\underline{d} \leqslant d(t) \leqslant \overline{d}$ with given values $\underline{d}$ and $\overline{d}$ are admissible, being not treated as a fault; leaving the interval $[\underline{d}, \overline{d}]$ is qualified as a fault to be detected. As in the paper [3], for arbitrary vectors x^1, x^2 and matrices A^1, A^2, the relations $x^1 \leqslant x^2$ and $A^1 \leqslant A^2$ are understood elementwise.

2. THE MAIN RESULT

The problem under consideration will be solved using the minimal-dimension model of system (1.1). In the general case, this model is described by the equation

$$\begin{aligned} x_*(t+1) &= F_*x_*(t) + G_*u(t) + J_*y(t) + D_*d(t) + L_*\rho(t), \\ y_*(t) &= H_*x_*(t), \end{aligned} \tag{2.1}$$

where $x_*(t) \in \mathbb{R}^k$ and $k < n$ denotes the model dimension; $y_* \in \mathbb{R}$; F_*, G_*, J_*, H_*, D_*, and L_* are the matrices to be determined. By assumption, the relations $x_*(t) = \Phi x(t)$ and $y_*(t) = R_*y(t)$ with some matrices Φ and R_* hold in the absence of faults and exogenous disturbances. The rules for building this model are presented in Section 3.

According to [1, 2], the model matrices satisfy the conditions

$$\begin{gathered} \Phi F = F_*\Phi + J_*H, \quad R_*H = H_*\Phi, \\ \Phi G = G_*, \quad \Phi D = D_*, \quad \Phi L = L_*. \end{gathered} \tag{2.2}$$

As was demonstrated in [1], the matrices F_* and H_* can be written in the canonical form

$$F_* = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad H_* = (\ 1 \ \ 0 \ \ 0 \ \ \dots \ \ 0\ ). \tag{2.3}$$

From the standpoint of the problem solved, this form seems ideal since the matrix F_* is stable for discrete systems and nonnegative (a necessary property to construct an interval observer) and the matrix H_* is nonnegative (a property simplifying the observer's form).

The desired interval observer is constructed based on model (2.1). By analogy with [12, 13], we find it in the form

$$\begin{aligned} \underline{x}_*(t+1) &= F_*\underline{x}_*(t) + G_*u(t) + J_*y(t) + D_*^+\underline{d} - D_*^-\overline{d} + L_*^+\underline{\rho} - L_*^-\overline{\rho}, \\ \overline{x}_*(t+1) &= F_*\overline{x}_*(t) + G_*u(t) + J_*y(t) + D_*^+\overline{d} - D_*^-\underline{d} + L_*^+\overline{\rho} - L_*^-\underline{\rho}, \\ \underline{y}_*(t) &= H_*\underline{x}_*(t), \\ \overline{y}_*(t) &= H_*\overline{x}_*(t), \\ \underline{r}(t) &= R_*y(t) - \overline{y}_*(t), \\ \overline{r}(t) &= R_*y(t) - \underline{y}_*(t), \end{aligned} \tag{2.4}$$

where $A^+ = \max\{0, A\}$ and $A^- = A^+ - A$ for an arbitrary matrix A. Obviously, $A^+ \geqslant 0$ and $A^- \geqslant 0$.

Theorem 1. *If $\underline{x}_*(0) \leqslant x_*(0) \leqslant \overline{x}_*(0)$, then the relation $0 \in [\underline{r}(t), \overline{r}(t)]$ holds for all $t \geqslant 0$ in the absence of faults. The case $0 \notin [\underline{r}(t), \overline{r}(t)]$ for some $t > 0$ is qualified as the occurrence of a fault.*

Proof. We introduce the errors $\underline{e}(t) = x_*(t) - \underline{x}_*(t)$ and $\overline{e}(t) = \overline{x}_*(t) - x_*(t)$. In view of the expressions (1.1), (2.1), and (2.2), the equation for the first error can be written and transformed as follows:

$$\begin{aligned} \underline{e}(t+1) &= x_*(t+1) - \underline{x}_*(t+1) \\ &= F_*x_*(t) + G_*u(t) + J_*y(t) + D_*d(t) + L_*\rho(t) \\ &\quad - (F_*\underline{x}_*(t) + G_*u(t) + J_*y(t) + D_*^+\underline{d} - D_*^-\overline{d} + L_*^+\underline{\rho} - L_*^-\overline{\rho}) \\ &= F_*(\underline{e}(t) + \underline{x}_*(t)) - F_*\underline{x}_*(t) + D_*d(t) - (D_*^+\underline{d} - D_*^-\overline{d}) \\ &\quad + L_*\rho(t) - (L_*^+\underline{\rho} - L_*^-\overline{\rho}) \\ &= F_*\underline{e}(t) + D_*d(t) - (D_*^+\underline{d} - D_*^-\overline{d}) + L_*\rho(t) - (L_*^+\underline{\rho} - L_*^-\overline{\rho}). \end{aligned} \tag{2.5}$$

Since $D_* = D_*^+ - D_*^-$,

$$\begin{aligned} D_*d(t) - (D_*^+\underline{d} - D_*^-\overline{d}) &= D_*^+d(t) - D_*^-d(t) - (D_*^+\underline{d} - D_*^-\overline{d}) \\ &= D_*^+(d(t) - \underline{d}) + D_*^-(\overline{d} - d(t)). \end{aligned}$$

In the absence of faults, we have $\underline{d} \leqslant d(t) \leqslant \overline{d}$ and, in addition, $D_*^+ \geqslant 0$ and $D_*^- \geqslant 0$. Consequently,

$$D_*d(t) - (D_*^+\underline{d} - D_*^-\overline{d}) \geqslant 0.$$

Similar considerations are adopted to show that

$$L_*\rho(t) - (L_*^+\underline{\rho} - L_*^-\overline{\rho}) \geqslant 0.$$

Recall that, by assumption, $\underline{e}(0) = x_*(0) - \underline{x}_*(0) \geqslant 0$ and $F_* \geqslant 0$. From (2.5) it therefore follows that $\underline{e}(1) \geqslant 0$. By induction we establish the inequality $\underline{e}(t) \geqslant 0$ for all $t \geqslant 0$. The second inequality $\overline{e}(t) \geqslant 0$ is proved by analogy.

Considering (2.2) and $H_* \geqslant 0$, formula (2.4) implies

$$\begin{aligned}\underline{r}(t) &= R_*y(t) - \overline{y}_*(t) = R_*Hx(t) - H_*\overline{x}_*(t)\\ &= H_*\Phi x(t) - H_*(\overline{e}(t) + x_*(t))\\ &= H_*x_*(t) - H_*(\overline{e}(t) + x_*(t))\\ &= -H_*\overline{e}(t) \leqslant 0\end{aligned}$$

for all $t \geqslant 0$. Similar considerations yield $\overline{r}(t) = R_*y(t) - \underline{y}_*(t) \geqslant 0$. The last two inequalities are equivalent to the required result, which can be written as the implication

$$d(t) \in [\underline{d}, \overline{d}] \Rightarrow 0 \in [\underline{r}(t), \overline{r}(t)]$$

for all $t \geqslant 0$. Then, under the condition $0 \notin [\underline{r}(t), \overline{r}(t)]$ for some $t > 0$, applying the negation operation to this implication gives

$$0 \notin [\underline{r}(t), \overline{r}(t)] \Rightarrow d(t) \notin [\underline{d}, \overline{d}],$$

which corresponds to the occurrence of a fault. The maximal sensitivity to faults is ensured by choosing appropriate matrices of the observer; see Section 3. The proof of Theorem 1 is complete.

Remark 1. In principle, the condition $\underline{x}_*(0) \leqslant x_*(0) \leqslant \overline{x}_*(0)$ can be omitted: due to observer's stability, the requirement $0 \in [\underline{r}(t), \overline{r}(t)]$ will hold for all $t \geqslant t_0$ with some finite time instant t_0.

Remark 2. Since the matrix F_* is stable by construction, the observer (2.4) is stable as well. It seems natural to assume that the original system is also stable and the control action $u(t)$ is finite; in this case, the variables $y(t)$, $\underline{y}_*(t)$, $\overline{y}_*(t)$ and the residuals $\underline{r}(t)$ and $\overline{r}(t)$ will be finite as well.

Thus, the built observer produces the interval $[\underline{r}(t), \overline{r}(t)]$. If zero belongs to this interval, the decision about no system faults is made (see Section 1); otherwise, the occurrence of a fault is concluded. In view of the observer's equations (2.4), the width of the interval $[\underline{r}(t), \overline{r}(t)]$ depends on exogenous disturbances and the admissible range of the variable $d(t)$. The smaller this width is, the more reliably the faults will be detected.

In terms of diagnosis quality, in particular, sensitivity to faults, the best interval observer is the one with the minimal width $[\underline{r}(t), \overline{r}(t)]$. According to (2.4), the corresponding case is when the model has insensitivity to the disturbance, i.e., $L_* = \Phi L = 0$. The method for building such a model was developed in [1, 2]. We briefly describe it below.

3. MODEL BUILDING

3.1. Main Relations

Due to the canonical form (2.3), equations (2.2) can be written as

$$\Phi_1 = R_*H, \quad \Phi_i F = \Phi_{i+1} + J_{*i}H, \quad i = 1, \ldots, k-1, \quad \Phi_k F = J_{*k}H, \tag{3.1}$$

where Φ_i and J_{*i} indicate the ith rows of the matrices Φ and J_*, respectively, $i = 1, \ldots, k$, and k is the dimension of model (2.1). These equations are reduced [1, 2] into the single one

$$(\ R_* \quad -J_{*1} \quad -J_{*2} \quad \ldots \quad -J_{*k}\ )V^{(k)} = 0, \tag{3.2}$$

where

$$V^{(k)} = \begin{pmatrix} HF^k \\ HF^{k-1} \\ \dots \\ H \end{pmatrix}.$$

The condition of insensitivity to disturbances ($\Phi L = 0$) can be represented as

$$(\, R_* \;\; -J_{*1} \;\; -J_{*2} \;\; \dots \;\; -J_{*k} \,)L^{(k)} = 0, \tag{3.3}$$

where

$$L^{(k)} = \begin{pmatrix} HL & HFL & HF^2L & \dots & HF^{k-1}L \\ 0 & HL & HFL & \dots & HF^{k-2}L \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Since the row $(\, R_* \;\; -J_{*1} \;\; -J_{*2} \;\; \dots \;\; -J_{*k} \,)$ satisfies (3.2), from (3.2) and (3.3) we obtain

$$(\, R_* \;\; -J_{*1} \;\; -J_{*2} \;\; \dots \;\; -J_{*k} \,)(V^{(k)} \;\, L^{(k)}) = 0. \tag{3.4}$$

Equation (3.4) has a nontrivial solution if

$$\operatorname{rank}(V^{(k)} \;\, L^{(k)}) < l(k+1).$$

This condition serves to determine the minimal dimension $k \geqslant 1$ under which equation (3.4) is solvable. Then, it is necessary to find the solution of (3.4), obtain the rows of the matrix Φ from (3.1), and let $G_* := \Phi G$ and $D_* := \Phi D$.

3.2. Maximizing Sensitivity to Faults

If equation (3.4) with the minimal dimension k has several solutions, it is possible to choose the one with the maximal contribution of faults to the observer (consequently, the maximal sensitivity to faults, estimated by the norm of the matrix $D_* = \Phi D$). This can be done more efficiently as follows. By analogy with the analysis of the contribution made by exogenous disturbances, we introduce the matrix

$$D^{(k)} = \begin{pmatrix} HD & HFD & HF^2D & \dots & HF^{k-1}D \\ 0 & HD & HFD & \dots & HF^{k-2}D \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Due to (3.1), it can be demonstrated that

$$\|D_*\| = \|(R_* \; - J_1 \; - J_2 \; \dots \; - J_k)D^{(k)}\|.$$

Then the contribution of faults is maximized by maximizing the norm

$$\|(R_* \; - J_1 \; - J_2 \; \dots \; - J_k)D^{(k)}\|$$

subject to condition (3.4).

Here, the idea is to find the minimal dimension k under which equation (3.4) has at least two linearly independent solutions of the form $(\, R_* \;\; -J_{*1} \;\; -J_{*2} \;\; \dots \;\; -J_{*k} \,)$. All these solutions are combined in a matrix W so that each row represents some solution of equation (3.4). According to the aforesaid, another solution is an arbitrary linear combination of the rows of this matrix with a weight vector $w = (w_1, \ldots, w_N)$, where N specifies the number of rows in the matrix W. The problem is to determine the vector w maximizing the norm $\|wWD^{(k)}\|$.

To solve this problem, we calculate the singular-value decomposition of the matrix product $WD^{(k)}$. In other words, the matrix $WD^{(k)}$ is represented as

$$WD^{(k)} = U_D \Sigma_D V_D,$$

where U_D and V_D are orthogonal matrices and the matrix Σ_D has the form

$$\Sigma_D = (\operatorname{diag}(\sigma_1, \ldots, \sigma_s)\ 0) \quad \text{or} \quad \Sigma_D = \begin{pmatrix} \operatorname{diag}(\sigma_1, \ldots, \sigma_s) \\ 0 \end{pmatrix}$$

depending on the number of rows and columns of the matrix $WD^{(k)}$, where $s = \min(N, kp)$ and $0 \leqslant \sigma_1 \leqslant \ldots \leqslant \sigma_s$ denote the singular values of the matrix $WD^{(k)}$ [19, 20]. Choosing the ith transposed column of the matrix U_D as the weight vector $w = (w_1, \ldots, w_N)$ yields $\|wWD^{(k)}\| = \sigma_i$ [19, 20]. In view of the considerations above, the appropriate vector $w = (w_1, \ldots, w_N)$ is the transposed column of the matrix U_D that corresponds to the maximal singular value and $(R\ -J_{*1}\ -J_{*2}\ \ldots\ -J_{*k}) := wW$. Finally, it is necessary to obtain the rows of the matrix Φ from (3.1) and let $G_* := \Phi G$ and $D_* := \Phi D$.

Note that this solution is optimal for the chosen dimension k; increasing the dimension further may give a better solution in terms of the maximum norm of the matrix $(R\ -J_{*1}\ -J_{*2}\ \ldots\ -J_{*k})D^{(k)}$.

3.3. Minimizing the Contribution of Exogenous Disturbances

If for all $k < n$ equation (3.4) is unsolvable, we cannot build the model insensitive to exogenous disturbances. Then it is necessary to employ robust methods minimizing the contribution of exogenous disturbances to the model [19]. Based on the analysis above, this problem obviously reduces to minimizing the norm $\|(R_*\ -J_1\ -J_2\ \ldots\ -J_k)L^{(k)}\|$ subject to condition (3.2).

By analogy with the considerations above, the idea is to find the minimal dimension k under which equation (3.2) has at least two linearly independent solutions of the form $(\ R_*\ \ -J_{*1}\ \ -J_{*2}\ \ \ldots\ \ -J_{*k}\ )$. All these M solutions, are combined in a matrix V so that each row represents some solution of equation (3.2). The problem is to determine a weight vector $v = (v_1, \ldots, v_M)$ mimizing the norm $\|vVL^{(k)}\|$.

Next, we calculate the singular-value decomposition of the matrix product $VL^{(k)}$, i.e., $VL^{(k)} = U_L \Sigma_L V_L$, and take the first transposed column of the matrix U_L as the weight vector $v = (v_1, \ldots, v_M)$. According to the aforesaid, the linear combination of the solutions corresponding to the rows of the matrix V with the weights $v_1, \ldots, v_M$ gives the optimal solution $(\ R_*\ \ -J_{*1}\ \ -J_{*2}\ \ \ldots\ \ -J_{*k}\ ) = vV$. Finally, it is necessary to obtain the rows of the matrix Φ from (3.1) and let $G_* := \Phi G$ and $D_* := \Phi D$. Thus, the robust model has been designed.

Other methods for building robust models, particularly the ones considering the contribution of faults, were discussed in [19].

4. ISOLATING FAULTS

The observer constructed above allows detecting the set of faults defined by the condition $D_* := \Phi D \neq 0$. To isolate faults, i.e., determine where they occur, it is necessary to design a bank of observers in which each observer will be sensitive to a particular set of faults and insensitive to the others. Such a bank can be constructed as follows. Let the set of possible faults in (1.1) be defined by the sum $\sum_{i=1}^{s} D_i d_i(t)$ instead of the term $Dd(t)$. A model insensitive to the first fault is built by solving the equation

$$\begin{pmatrix} R_* & -J_{*1} & -J_{*2} & \dots & -J_{*k} \end{pmatrix}(V^{(k)}\ D_1^{(k)}) = 0, \tag{4.1}$$

where

$$D_1^{(k)} = \begin{pmatrix} HD_1 & HFD_1 & HF^2D_1 & \dots & HF^{k-1}D_1 \\ 0 & HD_1 & HFD_1 & \dots & HF^{k-2}D_1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Its minimal dimension k is determined starting from $k = 1$. Next, the rows of the matrix Φ are obtained, $G_* := \Phi G$ is assigned, and an interval observer is constructed according to the above rules. It will be insensitive to several other faults, particularly to those for which $D_j = D_1 N$ with some matrix N, and sensitive to the faults for which $\Phi D_j \neq 0$. Choosing the first fault among them, we build a model and an observer insensitive to it by analogy. The procedure continues until the consideration of all faults.

The information about the sensitivity and insensitivity of each observer is reflected by the syndrome matrix S, where the rows correspond to observers and the columns to faults. In this matrix, $S(i, j) = 0$ if the ith observer is insensitive to the jth fault, and $S(i, j) = 1$ otherwise. The syndrome matrix may have two or more identical columns, meaning that some system faults are indistinguishable from each other by the described procedure. Therefore, it is necessary to apply more sophisticated approaches.

For fault isolation, the most convenient matrices are

$$S^1 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \quad S^2 = \begin{pmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots & \\ 1 & 1 & 1 & \dots & 0 \end{pmatrix}.$$

The first matrix allows isolating faults of arbitrary multiplicity, but it is rarely implementable in applications due to the very strict requirement of insensitivity to many faults. From this point of view, the second matrix seems more practical, but it is not always implementable as well. The matter is that the elements of the matrix S may have certain relations due to the peculiarities and faults of system (1.1), which make their choice nonarbitrary.

5. A PRACTICAL EXAMPLE

Consider an electric drive whose open circuit is described by the following model with viscous friction:

$$\begin{aligned} x_1(t+1) &= \gamma_1 x_2(t) + x_1(t), \\ x_2(t+1) &= \gamma_2 x_2(t) + \gamma_3 x_3(t) + \rho(t), \\ x_3(t+1) &= \gamma_4 x_2(t) + \gamma_5 x_3(t) + \gamma_6 u(t) + d(t), \\ y_1(t) &= x_2(t), \quad y_2(t) = x_3(t). \end{aligned} \tag{5.1}$$

Here, x_1 is the rotation angle of the gearbox output shaft, x_2 is the angular velocity of the electric motor shaft, and x_3 is the electric motor current. The coefficients γ_1–γ_6 depend on the drive parameters and the sampling interval; in particular, viscous friction is specified by the coefficient γ_2.

These coefficients are given by

$$\gamma_1 = \frac{\Delta t}{i_r}, \quad \gamma_2 = -\frac{\Delta t k_b}{J} + 1, \quad \gamma_3 = \frac{\Delta t k_m}{J},$$
$$\gamma_4 = -\frac{\Delta t k_\omega}{L_m}, \quad \gamma_5 = -\frac{\Delta t R_m}{L_m} + 1, \quad \gamma_6 = \frac{\Delta t k_u}{L_m}$$

with the following notations: Δt is the sampling interval; i_r is is the gear ratio; k_b is the viscous friction coefficient; k_m is the torque coefficient; J is the moment of inertia of the motor rotor and rotating parts of the gearbox reduced to this rotor; k_ω is the counter-emf coefficient; R_m is the rated active resistance of the armature circuit; L_m is the armature circuit inductance; k_u is the power amplifier gain; finally, $u(t)$ is the drive input voltage.

The electric drive is described by the matrices

$$F = \begin{pmatrix} 1 & \gamma_1 & 0 \\ 0 & \gamma_2 & \gamma_3 \\ 0 & \gamma_4 & \gamma_5 \end{pmatrix}, \; G = \begin{pmatrix} 0 \\ 0 \\ \gamma_5 \end{pmatrix}, \; H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; D = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \; L = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

We build a model insensitive to the disturbance. Letting $k = 1$, we calculate the matrices $V^{(1)}$ and $B^{(1)}$:

$$V^{(1)} = \begin{pmatrix} 1 & \gamma_1 & 0 \\ 0 & \gamma_4 & \gamma_5 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B^{(1)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Since rank $(V^{(1)}, B^{(1)}) = 3 < 2(1+1) = 4$, equation (3.4) has a solution with the matrices

$$R_* = (\gamma_4 \quad -\gamma_1), \quad J_* = (\gamma_4 \quad -\gamma_1\gamma_5).$$

As a result, $\Phi = (\gamma_4 \;\; 0 \;\; -\gamma_1)$, $G_* = -\gamma_1\gamma_6$, and $D_* = -\gamma_1$; model (2.1) takes the form

$$\begin{aligned} x_*(t+1) &= \gamma_4 y_1(t) - \gamma_1\gamma_5 y_2(t) - \gamma_1\gamma_6 u(t) - d(t), \\ y_*(t) &= x_*(t), \end{aligned} \tag{5.2}$$

where $x_* = \gamma_4 x_1 - \gamma_1 x_3$. Obviously, $D_*^+ = 0$ and $D_*^- = \gamma_1$.

According to (2.4) and (5.2), the interval observer is described by the equations

$$\begin{aligned} \underline{x}_*(t+1) &= \gamma_4 y_1(t) - \gamma_1\gamma_5 y_2(t) - \gamma_1\gamma_6 u(t) - \gamma_1\overline{d}, \\ \overline{x}_*(t+1) &= \gamma_4 y_1(t) - \gamma_1\gamma_5 y_2(t) - \gamma_1\gamma_6 u(t) - \gamma_1\underline{d}, \\ \underline{y}_*(t) &= \underline{x}_*(t), \quad \overline{y}_*(t) = \overline{x}_*(t), \\ \underline{r}(t) &= \gamma_4 y_1 - \gamma_1 y_3 - \overline{y}_*(t), \quad \overline{r}(t) = \gamma_4 y_1 - \gamma_1 y_3 - \underline{y}_*(t). \end{aligned}$$

For the sake of simplicity in simulation, we choose $\gamma_1 = \gamma_3 = \gamma_6 = 1$, $\gamma_2 = \gamma_4 = \gamma_5 = -1$, and $u(t) = 2 + \sin(t)$; the disturbance $\rho(t)$ is represented by a random variable with the uniform distribution on the interval $[-0.2, 0.2]$; finally, the admissible variations of the function $d(t)$ belong to the interval $[\underline{d}, \overline{d}] = [-0.05, 0.05]$. Figures 1 and 2 show the simulation results for the observer with the initial conditions $x_1(0) = x_2(0) = x_3(0) = 0$, $\underline{x}_*(0) = -0.2$, and $\overline{x}_*(0) = 0.2$.

In Fig. 1, $d(t) = 0$ for $t < 40$ s, and $d(t) = 0.04$ for $t \geqslant 40$ s. Since the value $d(t) = 0.04$ lies within the admissible interval, we have $0 \in [\underline{d}, \overline{d}]$, which is qualified as no faults. In Fig. 2, the function $d(t)$ is represented by a random variable with the uniform distribution on the interval

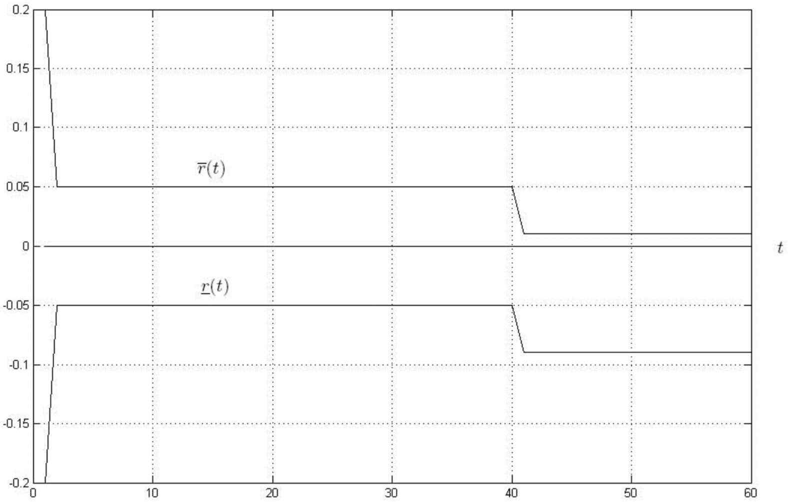


Fig. 1. The residuals $\underline{r}$ and $\overline{r}$ without faults.

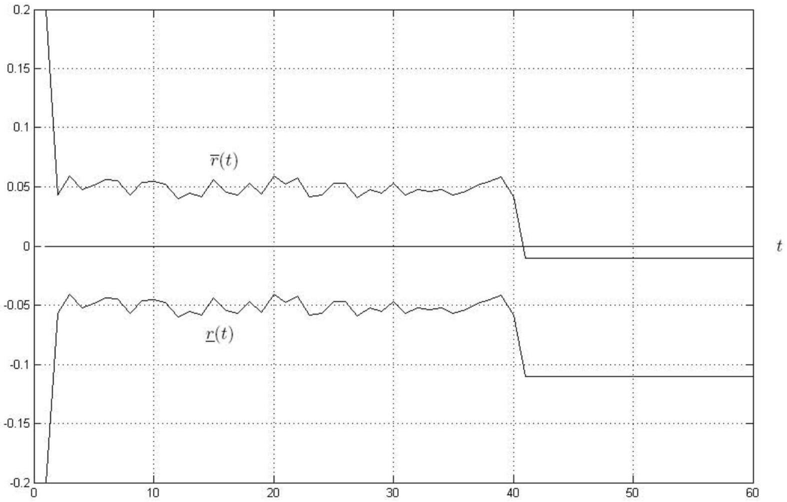


Fig. 2. The residuals $\underline{r}$ and $\overline{r}$ under fault occurrence.

$[-0.01, 0.01]$ for $t < 40$ s, and $d(t) = 0.06$ for $t \geqslant 40$ s. Now, $0 \notin [\underline{r}(t), \overline{r}(t)]$ for $t > 40$ s, and the occurrence of a fault is concluded.

According to Fig. 2, the random variable $d(t)$ affects the behavior of the functions $\underline{r}(t)$ and $\overline{r}(t)$. The disturbance $\rho(t)$ is also represented by a random variable, but the functions $\underline{r}(t)$ and $\overline{r}(t)$ in Fig. 1 are constant (except for a jump due to the change in the function $d(t)$). Therefore, the disturbance has no impact on the result.

6. CONCLUSIONS

This paper has considered interval observer-based functional diagnosis for linear dynamic systems described by discrete-time models with exogenous disturbances. Formulas have been derived to construct an interval observer that is insensitive to disturbances and sensitive to a limited extent. Such an observer produces two values of the residual as follows: if zero is between these values, then the system has no faults to be detected by the observer. The case where zero does not belong to the interval between these values is qualified as the occurrence of a fault. The theoretical results have been illustrated by an example of observer design for a real technical system. The simulation results of this example have confirmed the correctness of theoretical constructs related to fault detection.

FUNDING

This work was supported by the Russian Science Foundation, project no. 23-29-00191, https://rscf.ru/project/23-29-00191/.

REFERENCES

1. Zhirabok, A.N., Zuev, V.V., and Ir, K.C., Method to Design Interval Observers for Linear Time-Invariant Systems, *J. Comput. Syst. Sci. Int.*, 2022, vol. 61, no. 4, pp. 485–495.

2. Zhirabok, A., Zuev, A., Filaretov, V., Shumsky, A., and Kim Chkhun Ir, Jordan Canonical Form in Diagnosis and Estimation Problems, *Autom. Remote Control*, 2022, vol. 83, no. 9, pp. 1355–1370.

3. Efimov, D. and Raissi, T., Design of Interval State Observers for Uncertain Dynamical Systems, *Autom. Remote Control*, 2016, vol. 77, no. 2, pp. 191–225.

4. Khan, A., Xie, W., Zhang, L., and Liu, L., Design and Applications of Interval Observers for Uncertain Dynamical Systems, *IET Circuits Devices Syst.*, 2020, vol. 14, pp. 721–740.

5. Kremlev, A.S. and Chebotarev, S.G., Synthesis of Interval Observer for Linear System with Variable Parameters, *Journal of Instrument Engineering*, 2013, vol. 56, no. 4, pp. 42–46.

6. Efimov, D., Raissi, T., Perruquetti, W., and Zolghadri, A., Estimation and Control of Discrete-Time LPV Systems Using Interval Observers, *Proc. 52nd IEEE Conf. on Decision and Control*, Florence, 2013, pp. 5036–5041.

7. Chebotarev, S., Efimov, D., Raissi, T., and Zolghadri, A., Interval Observers for Continuous-Time LPV Systems with L1/L2 Performance, *Automatica*, 2015, vol. 51, pp. 82–89.

8. Mazenc, F. and Bernard, O., Asymptotically Stable Interval Observers for Planar Systems with Complex Poles, *IEEE Trans. Automatic Control*, 2010, vol. 55, no. 2, pp. 523–527.

9. Zheng, G., Efimov, D., and Perruquetti, W., Interval State Estimation for Uncertain Nonlinear Systems, *Proc. IFAC NOLCOS 2013*, Toulouse, 2013.

10. Zhang, K., Jiang, B., Yan, X., and Edwards, C., Interval Sliding Mode Based Fault Accommodation for Non-Minimal Phase LPV Systems with Online Control Application, *Int. J. Control*, 2019. https://doi.org/10.1080/00207179.2019.1687932

11. Kolesov, N., Gruzlikov, A., and Lukoyanov, E., Using Fuzzy Interacting Observers for Fault Diagnosis in Systems with Parametric Uncertainty, *Proc. 12th Inter. Symp. Intelligent Systems (INTELS'16)*, October 5–7, 2016, Moscow, pp. 499–504.

12. Zhang, Z. and Yang, G., Fault Detection for Discrete-Time LPV Systems Using Interval Observers, *Int. J. Syst. Sci.*, 2017. https://doi.org/10.1080/00207721.2017.1363926

13. Zhang, Z. and Yang, G., Event-Triggered Fault Detection for a Class of Discrete-Time Linear Systems Using Interval Observers, *ISA Transactions*, 2017. https://doi.org/10.1016/j.isatra.2016.11.016

14. Zhang, Z. and Yang, G., Interval Observer-Based Fault Isolation for Discrete-Time Fuzzy Interconnected Systems with Unknown Interconnections, *IEEE Trans. Cybernetics*, 2017. https://doi.org/10.1109/TCYB.2017.2707462

15. Yi, Z., Xie, W., Khan, A., and Xu, B., Fault Detection and Diagnosis for a Class of Linear Time-Varying Discrete-Time Uncertain Systems Using Interval Observers, *Proc. 39th Chinese Control Conf.*, July 27–29, 2020, Shenyang, pp. 4124–4128.

16. Rotondo, D., Fernandez-Cantia, R., Tornil-Sina, S., Blesa, J., and Puig, V., Robust Fault Diagnosis of Proton Exchange Membrane Fuel Cells Using a Takagi–Sugeno Interval Observer Approach, *Int. J. Hydrogen Energy*, 2015, pp. 2875–2886.

17. Saijai, J., Ding, S., Abdo, A., Shen, B., and Damlakhi, W., Threshold Computation for Fault Detection in Linear Discrete-Time Markov Jump Systems, *Int. J. Adapt. Control Signal Process.*, 2014, vol. 28, pp. 1106–1127.

18. Shumskii, A. and Zhirabok, A., Decision Making in Nonlinear Dynamical System Diagnosis by a Nonparametric Method, *Autom. Remote Control*, 2021, vol. 82, no. 2, pp. 278–293.

19. Zhirabok, A., Shumskii, A., Solyanik, S., and Suvorov, A., Design of Nonlinear Robust Diagnostic Observers, *Autom. Remote Control*, 2017, vol. 78, no. 9, pp. 1572–1584.

20. Low, X., Willsky, A., and Verghese, G., Optimally Robust Redundancy Relations for Failure Detection in Uncertain Systems, *Automatica*, 1996, vol. 22, pp. 333–344.

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