Volume 87, Number 1,
January 2026

ISSN 0005-1179
CODEN: AURCAT

AUTOMATION AND REMOTE CONTROL

Automation and Remote Control

Vol. 87, No. 1, January 2026

Editor-in-Chief
Andrey A. Galyaev

http://ait.mtas.ru

Available via license: CC BY 4.0

Automation and Remote Control

ISSN 0005-1179

Editor-in-Chief
Andrey A. Galyaev

Deputy Editors-in-Chief M.V. Khlebnikov and E.Ya. Rubinovich

Coordinating Editor A.S. Samokhin

Editorial Board

F.T. Aleskerov, A.V. Arutyunov, N.N. Bakhtadze, A.A. Bobtsov, P.Yu. Chebotarev, A.G. Chkhartshvili, L.Yu. Filimonyuk, A.L. Fradkov, O.N. Granichin, M.F. Karavai, E.M. Khorov, M.M. Khrustalev, A.I. Kibzun, S.A. Krasnova, A.P. Krishchenko, A.G. Kushner, N.V. Kuznetsov, A.A. Lazarev, A.I. Lyakhov, A.I. Matasov, S.M. Meerkov (USA), R.V. Mescheryakov, A.I. Mikhal'skii, B.M. Miller, O.V. Morzhin, R.A. Munasypov, A.V. Nazin, A.S. Nemirovskii (USA), D.A. Novikov, A.Ya. Oleinikov, P.V. Pakshin, D.E. Pal'chunov, A.E. Polyakov (France), V.Yu. Protasov, L.B. Rapoport, I.V. Rodionov, N.I. Selvesyuk, P.S. Shcherbakov, A.N. Sobolevski, O.A. Stepanov, A.B. Tsybakov (France), D.V. Vinogradov, V.M. Vishnevskii, K.V. Vorontsov, and L.Yu. Zhilyakova

Staff Editor E.A. Martekhina

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

© The Author(s), 2026 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences.

Automation and Remote Control participates in the Copyright Clearance Center (CCC) Transactional Reporting Service.

Available via license: CC BY 4.0

0005-1179/26. *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 87 (12 issues) is published in 2026.

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. 87, No. 1, 2026

Reviews

Linear Systems

Nonlinear Systems

Intellectual Control Systems, Data Analysis

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 1, pp. 1–17.
© The Author(s), 2026 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2026.
Russian Text © The Author(s), 2026, published in Avtomatika i Telemekhanika, 2026, No. 1, pp. 3–27.

REVIEWS

Stochastic Polling Systems: Development and New Applications

O. V. Semenova[*,a] **and V. M. Vishnevsky**[*,b]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a] *olga.semenova@ipu.ru,* [b] *vishn@inbox.ru*

Received October 2, 2025
Revised October 27, 2025
Accepted October 28, 2025

Abstract—This review presents new results obtained in the field of stochastic polling systems research. It systematizes the main directions of practical application of polling models for performance evaluation and design of wireless WMAN networks with centralized control mechanisms, 5G/6G cellular networks, Internet of Things networks, transportation and medical systems, etc. The prospects for the further development of applied research in this important area of queueing theory are discussed.

Keywords: Polling systems, telecommunication networks, 5G/6G, Internet of Things, transport control, healthcare systems

DOI: 10.7868/S1608303226010011

1. INTRODUCTION

Mathematical models of stochastic polling systems have found extensive application in the design of communication, manufacturing, transportation, and medical systems, among others. These models are particularly effective for the evaluation of performance and the design of existing and prospective telecommunication network protocols that employ centralised control mechanisms.

Research into stochastic polling systems and the synthesis of results in this domain of queueing theory up to 1995 are documented in monographs [1, 2]. Because of their broad practical applications, the scholarly interest in polling systems remains undiminished. The past two decades have witnessed the publication of review papers [3–8], primarily focusing on the mathematical methodologies for their analysis. An exception is the 2011 review [9], which systematised the practical applications for stochastic polling models.

The rapid evolution and successive generational shifts in cellular networks, broadband wireless systems, industrial robotics, and various unmanned ground/air systems have necessitated the development of novel polling models and analytical techniques for performance assessment and system design. These include methods for analyzing systems with correlated arrival streams, characteristic of modern computational networks [10, 11], and machine learning approaches applied when traditional queueing-theoretic methods yield intractable or unobtainable analytical/numerical results [12, 13]. Despite a substantial volume of recent publications in the last decade, a comprehensive review systematizing new directions in the practical application of stochastic polling models remains absent. The present survey aims to address this gap.

The paper is structured as follows. Section 2 introduces fundamental concepts and describes the basic stochastic polling model, which constitutes the subject of most theoretical and applied work in the field. Section 3 examines the application of stochastic polling models for performance evaluation in IEEE 802.11 broadband wireless metropolitan area networks with centralized control.

Section 4 systematizes research on the application of centralized polling models in Internet of Things (IoT) and 5G/6G cellular networks. Section 5 discusses how stochastic polling models can help manage vehicular traffic at intersections. Section 6 describes the applications of polling systems in healthcare for the disease prognosis, diagnosis, and the hospital logistics. Section 7 outlines other applications including passive optical networks, energy consumption, and manufacturing processes. The Conclusion identifies prospective directions for further applied research in stochastic polling systems.

2. FUNDAMENTAL CONCEPTS AND THE BASIC STOCHASTIC POLLING MODEL

This section details the primary mathematical model underpinning most applied research on polling systems. Subsequent sections will provide concise descriptions of model variants where they diverge from this foundational framework.

A basic polling system comprises N ($N \geqslant 2$) queues (nodes, stations) with infinite waiting space and a single server. The server is a physical entity processing requests/customers. The service here providing an access to a specific resource (e.g., data channel access) contingent upon the model's context.

Customers arrive at queue Q_i according to a Poisson process with rate λ_i. Service times in Q_i are independent and identically distributed with distribution function $B_i(t)$, first and second moments b_i and $b_i^{(2)}$, and Laplace–Stieltjes Transform (LST) $\tilde{B}_i(x) = \int_0^\infty e^{-xt} dB_i(t)$, for $i = \overline{1, N}$. The server visits queues in a prescribed order, most commonly cyclic. During a visit, customers are served according to a specified discipline: *exhaustive* (server continues until the queue is empty), *gated* (only customers present at the polling moment are served), *globally-gated* (only customers present at the cycle start are served), or k_i-*limited* (at most k_i customers served per visit to Q_i). Comprehensive classifications are available in [4, 6].

The server switchover time to connect to Q_i is a random variable with distribution $S_i(t)$, moments s_i, $s_i^{(2)}$, and LST $\tilde{S}_i(x)$. For cyclic polling systems, the total switchover time per cycle has moments $s = \sum_{j=1}^N s_j$ and $s^{(2)} = s^2 + \sum_{j=1}^N (s_j^{(2)} - s_j^2)$. A *cycle* is the time to poll and serve all queues once. Under cyclic polling, the mean cycle time is $C = \frac{s}{1-\rho}$ where $\rho = \sum_{i=1}^N \rho_i$ is the total load and $\rho_i = \lambda_i b_i$ is the load of Q_i. Stability for exhaustive/gated systems requires $\rho < 1$; for k_i-limited service, this is supplemented by $\lambda_i < \frac{k_i(1-\rho_i)}{s}$ for all i.

The analytical challenge lies in constructing a multidimensional process capturing the polling dynamics, deriving stationary probabilities and mean performance metrics, and obtaining numerical solutions for large-scale systems.

Methodological approaches include mean-value analysis [14], the branching process method [15], and decomposable semi-regenerative process theory [16]. The predominant technique for cyclic/periodic systems is the probability generating function (the PGF) method [17] briefly outlined here for gated and exhaustive service.

Let X_i^j be the number of customers in Q_j when Q_i is polled, and define the PGF $G_i(\mathbf{z}) = \mathbf{E}\left[\prod_{j=1}^N z_j^{X_i^j}\right]$ with $\mathbf{z} = (z_1, \ldots, z_N)$.

For gated service, these functions satisfy

$$G_{i+1}(\mathbf{z}) = G_i\left(z_1, z_2, \ldots, z_{i-1}, \tilde{B}_i\left[\sum_{j=1}^N \lambda_j (1 - z_j)\right], z_{i+1}, \ldots, z_N\right) \tilde{S}_{i+1}\left[\sum_{j=1}^N \lambda_j (1 - z_j)\right], \quad i = \overline{1, N}. \quad (1)$$

The mean queue lengths $f_i(j) = \mathbf{E}[X_i^j]$ are solutions to a linear system yielding:

$$\begin{aligned} f_{i+1}(j) &= f_i(j) + \lambda_j b_i f_i(i) + \lambda_j s_{i+1}, \quad j \neq i, \\ f_{i+1}(i) &= \lambda_i b_i f_i(i) + \lambda_i s_{i+1}, \quad i = \overline{1, N}. \end{aligned} \tag{2}$$

Second moments are obtained via further differentiation, leading to the mean waiting time $W_i = \frac{(1+\rho_i)f_i(i,i)}{2\lambda_i^2 C}$.

For exhaustive service, the relations become:

$$G_{i+1}(\mathbf{z}) = G_i\left(z_1, \ldots, z_{i-1}, \tilde{\theta}_i\left[\sum_{\substack{j=1 \\ j\neq i}}^{N} \lambda_j(1-z_j)\right], z_{i+1}, \ldots, z_N\right) \tilde{S}_{i+1}\left[\sum_{j=1}^{N} \lambda_j(1-z_j)\right],$$

where $\tilde{\theta}_i(w)$, the LST of a busy period in the corresponding $M/G/1$ queue for Q_i, satisfies

$$\tilde{\theta}_i(w) = \tilde{B}_i(w + \lambda_i - \lambda_i \tilde{\theta}_i(w)). \tag{3}$$

The mean waiting time is then $W_i = \frac{\lambda_i b_i^{(2)}}{2(1-\rho_i)} + \frac{f_i(i,i)}{2\lambda_i^2(1-\rho_i)C}$.

Other analytical methods include functional computation [18], matrix-analytic approaches [19], heavy-traffic asymptotics [20, 21], and the pseudo-conservation law [22], detailed in [5, 6].

3. APPLICATION OF STOCHASTIC POLLING MODELS FOR PERFORMANCE EVALUATION AND DESIGN OF BROADBAND WIRELESS METROPOLITAN AREA NETWORKS

A Wireless Metropolitan Area Network (WMAN) is a type of wireless network designed to provide telecommunications coverage over a broad geographical area. WMANs consist of a series of interconnected wireless base stations or access points strategically located throughout a metropolitan area. These base stations communicate with each other and with end-user devices to enable efficient and seamless data transmission.

In wireless metropolitan networks operating under IEEE 802.11 standard protocols, centralized polling mechanisms are widely employed. Among these protocols, which implement cyclic polling of subscriber terminals by a base station (access point server), are the Point Coordination Function (PCF) and its evolution—the Hybrid Coordination Function (HCF) and the HCF Controlled Channel Access (HCCA) protocol [23]. The use of these protocols in metropolitan networks characterized by high-rise buildings and radio interference helps mitigate the hidden terminal problem, efficiently schedules station access to the wireless channel, provides flexible control over the radio cell operation, and allows dynamic adjustment of its parameters according to the current situation.

References [24–26] present a comparative analysis of various polling schemes for evaluating the performance characteristics of IEEE 802.11 PCF metropolitan networks. They also propose an adaptive polling model that not only addresses the hidden terminal problem but also optimizes access to the data transmission channel. It is worth noting that the adaptive polling scheme and its first protocol version were initially described in [27–29].

In the adaptive polling scheme, the order of queue visits remains cyclic, but the server does not visit (does not poll) queues that were empty at the time of polling in the previous cycle. For asymmetric systems with non-zero switchover times, adaptive polling reduces the mean polling cycle time and, consequently, enhances wireless network performance.

The formula for the mean cycle time from Section 2 in this case takes the form

$$C = \frac{\sum_{i=1}^{N} s_i u_i + \beta \prod_{i=1}^{N}(1-u_i)}{1-\rho}$$

where $u_i = \frac{1}{1+e^{-\lambda_i C}}$ is the probability that queue Q_i is visited by the server in an arbitrary cycle, $i=\overline{1,N}$, and β is the mean idle (waiting) time of the server when all queues in sequence are found empty during polling and must be skipped. The server waiting time is introduced here to eliminate empty polling cycles.

In the case of the gated service discipline, the system (1) for the PGFs $G_i(\mathbf{z})$, $i=\overline{1,N}$ takes the form:

$$\begin{aligned} G_i(\mathbf{z}) &= (1-u_i)u_{i-1}\mathcal{M}_{i+1}^{(1)}(\mathbf{z}) + \ldots + (1-u_1)\cdots(1-u_{N-1})u_N\mathcal{M}_{i+1}^{(N-1)}(\mathbf{z}) \\ &\quad + (1-u_1)\cdots(1-u_N)\mathcal{M}_{i+1}^{(N)}(\mathbf{z}) + u_i\mathcal{M}_{i+1}^{(0)}(\mathbf{z}), \end{aligned} \tag{4}$$

where

$$\mathcal{M}_{i+1}^{(l)}(\mathbf{z}) = G_{i-l}\left(z_1,\ldots,z_{i-l},\widetilde{B}_{i-l}\left(\sum_{j=1}^{N}\lambda_j(1-z_j)\right),z_{i-l+2},\ldots,z_N\right)\widetilde{S}_{i-l+1}\left[\sum_{j=1}^{N}\lambda_j(1-z_j)\right],$$

$$\begin{aligned} \mathcal{M}_{i+1}^{(N)}(\mathbf{z}) &= G_{i-N}\left(z_1,\ldots,z_{i-N},\widetilde{B}_{i-N}\left(\sum_{j=1}^{N}\lambda_j(1-z_j)\right),z_{i-N+2},\ldots,z_N\right) \\ &\quad \times \widetilde{S}_{i-N+1}\left[\sum_{j=1}^{N}\lambda_j(1-z_j)\right]\tilde{H}\left(\sum_{j=1}^{N}\lambda_j(1-z_j)\right), \quad l=\overline{0,N-1} \end{aligned}$$

and system (2) transforms to:

$$\begin{aligned} f_{i+1}(j) &= u_i\left[I_{\{i=j\}}f_i(j) + \lambda_j b_i f_i(i) + \lambda_j s_{i+1}\right] \\ &\quad + (1-u_i)u_{i-1}\left[\lambda_j(s_{i+1}+\beta) + I_{\{i-1=j\}}f_{i-1}(j) + \lambda_j b_{i-1}f_{i-1}(i-1) + \lambda_j s_{i+1}\right] + \ldots \\ &\quad + (1-u_1)\cdots(1-u_N)\left[I_{\{i-N=j\}}f_{i-N}(j) + \lambda_j b_{i-N}f_{i-N}(i-N)\right], \quad i,j=\overline{1,N}. \end{aligned}$$

As in the case of ordinary cyclic polling, the mean waiting times for the system with adaptive polling are determined as $W_i = \frac{(1+\rho_i)f_i(i,i)}{2\lambda_i^2 C}$, $i=\overline{1,N}$, where the second moments $f_i(j,k)$, $i,j,k=\overline{1,N}$ are computed as a solution to a system of linear equations obtained by differentiating the PGFs $G_i(\mathbf{z})$, $i=\overline{1,N}$. The case of exhaustive queue service is considered similarly [25].

In [30], a polling model with time-limited queue service, describing data transmission in IEEE 802.11e HCCA networks, is investigated. It is shown that the PGF $\omega_i(\mathbf{z})$ for the number of customers in the system at the moments of service completion in queue Q_i is expressed through the PGF $\bar{\omega}_i(\mathbf{z})$ for the number of customers in the system at the moments of polling queue Q_i as follows:

$$\omega_i(\mathbf{z}) = \frac{\gamma\tilde{B}_i\left(\sum_{k=1}^{N}\lambda_k(1-z_k)\right)}{z_i - \tilde{B}_i\left(\sum_{k=1}^{N}\lambda_k(1-z_k)\right)}\left[\bar{\omega}_{i-1}(\mathbf{z})\tilde{S}_i\left(\sum_{k=1}^{N}\lambda_k(1-z_k)\right) - \bar{\omega}_i(\mathbf{z})\right], \quad i=\overline{1,N}$$

where γ is a known constant determined by the mean cycle time C and the total arrival intensity into the system. The PGFs $\bar{\omega}_i(\mathbf{z})$, $i=\overline{1,N}$ are found from functional equations numerically using the Fourier transform. The LST $W_i(s)$ of the waiting time in queue Q_i satisfies the relation

$$W_i(s) = N_i(1-s/\lambda_i)/\tilde{B}_i(s), \quad i=\overline{1,N},$$

where $N_i(s)$ is the PGF for the number of customers in queue Q_i at service completion moments, computed as $N_i(z) = \omega_i(1, \ldots, 1, z_i, 1, \ldots, 1)/\omega_i(\mathbf{1})$.

Reference [31] also considers the application of a large-scale polling system with branching-type service disciplines to modeling the BACnet (Building Automation and Control Networks) protocol for building automation and control networks. Branching-type disciplines are defined as follows. If at some embedded moment, e.g. an arbitrary queue Q_i polling moment, its length is k_i then each of these k_i customers during the Q_i service time is replaced by some random number of customers having PGF $h_i(\mathbf{z})$. As shown in [32], polling systems with branching-type disciplines allow for exact analysis of performance characteristics whereas service disciplines not possessing this property typically necessitate approximate analytical methods.

Branching-type disciplines include, for example, the gated discipline with

$$h_i(z_1, \ldots, z_N) = \tilde{B}_i\left(\sum_{k=1}^{N} \lambda_k(1 - z_k)\right), \quad i = \overline{1, N}$$

and the exhaustive discipline with $h_i(z_1, \ldots, z_N) = \tilde{\theta}_i\left(\sum_{k=1}^{N} \lambda_k(1 - z_k)\right)$, $i = \overline{1, N}$, where the function $\tilde{\theta}(s)$ is defined by the functional relation (3). Limited disciplines, however, do not belong to this class, and systems with such disciplines require the development of approximate methods.

For a polling system with a branching-type discipline, the functional relation (1) for the PGFs $G_i(\mathbf{z})$, $i = \overline{1, N}$ transforms to

$$G_{i+1}(\mathbf{z}) = G_i\left(z_1, z_2, \ldots, z_{i-1}, h_i(\mathbf{z}), z_{i+1}, \ldots, z_N\right) \tilde{S}_{i+1}\left[\sum_{j=1}^{N} \lambda_j(1 - z_j)\right], \quad i = \overline{1, N}$$

which enables the derivation of performance characteristics based on the mean queue lengths at polling moments, obtained by differentiating the above relations.

The LST of the cycle time C is given by

$$\mathbf{E}\left[e^{-uC}\right] = G_1(\tilde{\theta}_1(\gamma_1(\mathbf{u})), \ldots, \tilde{\theta}(\gamma_N(\mathbf{u}))) \prod_{i=1}^{N} \tilde{S}_i(\mathbf{u}),$$

where $\mathbf{u} = (u, \ldots, u)$, $\gamma_N(\mathbf{z}) = z_N$, $\gamma_i(\mathbf{z}) = z_i + \sum_{j=i+1}^{N} \lambda_i(1 - \tilde{\theta}_j(\gamma_j(\mathbf{z})))$, $i = \overline{1, N-1}$.

For service disciplines not belonging to the branching class, such as k-limited service, [31] also proposes a so-called *flexible k-limited* discipline. It is similar to the ordinary k-limited discipline except when the server serves fewer than k customers in a queue. In this case, the server utilizes the unused service capacity to serve the next queue in order. This allows, for example, in the case of a highly asymmetric polling system, to transfer part of the service resource from less loaded queues to more loaded ones, thereby reducing the mean system waiting time.

References [33, 34] explore the possibility of organizing a metropolitan broadband wireless network using a tethered high-altitude long-endurance unmanned aerial platform [35–37]. An IEEE 802.11 PCF base station installed on an Unmanned Aerial Vehicle (UAV) at an altitude of 100–150 m can significantly increase the line-of-sight zone between subscribers and the base station antenna thereby expanding the network's coverage area. References [38, 39] investigate a polling model with retrials, describing the data transmission mechanism in FANET (Flying Ad Hoc Network) — wireless self-organizing networks based on UAVs.

In [40], the case of Markovian polling order is investigated, modeling Carrier-Sense Multiple-Access Collision-Avoidance (CSMA-CA) in IEEE 802.11 networks operating in random access mode. The server visits queues according to a random Markovian polling scheme, described by a stochastic matrix $\mathbf{P} = (p_{i,j})_{i,j=\overline{1,N}}$ and a probability vector $\mathbf{q} = (q_1, \dots, q_N)$. The vector $\mathbf{q}$ is calculated as the solution of the system $\mathbf{qP} = \mathbf{q}$, $\sum_{i=1}^N q_i = 1$.

The queue service discipline can be any branching type as described above. The mean cycle time for queue Q_i is calculated as $C_i = \frac{\sigma}{q_i(1-\rho)}$, where $\sigma = \sum_{j=1}^N q_j \sum_{k=1}^N p_{j,k}\mathbf{E}[S_{j,k}]$; $S_{j,k}$ is the random variable describing the switchover time between queues Q_j and Q_k, $j,k = \overline{1,N}$.

This case allows application of the PGF method yielding the functional relations similar to (1)

$$q_j G_j(\mathbf{z}) = \sum_{i=1}^N p_{i,j} q_i G_i\left(z_1, z_2, \dots, z_{i-1}, h_i(\mathbf{z}), z_{i+1}, \dots, z_N\right) \tilde{S}_{i+1}\left[\sum_{j=1}^N \lambda_j(1-z_j)\right], \quad j = \overline{1,N}.$$

The first moments $f_i(j)$, $i,j = \overline{1,N}$ of the queue lengths at the start of polling can be obtained explicitly facilitating the computation of key system performance characteristics.

The mean unconditional queue length L_j for Q_j is computed as:

$$L_j = \sum_{i=1}^N \left(\rho_i x_i(j) + \frac{(1-\rho)q_i}{\sigma} \sum_{k=1}^N p_{i,k}\mathbf{E}[S_{i,k}] y_{i,k}(j)\right), \quad j = \overline{1,N}$$

where $x_i(j)$ and $y_{i,k}(j)$ are known constants. The article [40] also generalizes the pseudo-conservation law for an arbitrary branching-type service discipline which can be used for system performance optimization

$$\begin{aligned}\sum_{i=1}^N \rho_i W_i = \rho\frac{\sum_{i=1}^N \lambda_i b_i^{(2)}}{2(1-\rho)} + \frac{1}{\sigma}\sum_{i=1}^N q_i \sum_{k=1}^N p_{i,k}\mathbf{E}[S_{i,k}] \sum_{j=1, j\neq i}^N f_i(j) b_j \\ + \frac{1}{1-\rho}\sum_{i=1}^N \lambda_i \frac{1-\rho_i}{1-h_i(i)} \sum_{k=1}^N p_{i,k}\mathbf{E}[S_{i,k}] \sum_{j=1}^N b_j h_i(j) + \frac{\rho}{2\sigma}\sum_{i=1}^N q_i \sum_{k=1}^N p_{i,k}\mathbf{E}[S_{i,k}^2].\end{aligned}$$

A model of a priority polling system with three queues and threshold-based service is used in [41] to analyze switching systems operating in Asynchronous Transfer Mode (ATM) for large-scale metropolitan networks, such as the DQDB (Distributed Queue Dual Bus) standard.

A similar application of a polling system with random polling order to describe a multiple access scheme in IEEE 802.16 WiMAX metropolitan networks operating in contention-based access mode is considered in [42]. Reference [43] proposes a multi-level MAC protocol with priority polling for military applications. For a system with three queues and limited service, describing a specific case of a polling system with an adaptive polling mechanism, [44] presents the stability condition.

In conclusion of this section, we note that subsequent versions of the IEEE 802.11 standard have introduced new access mechanisms (e.g., Triggered Uplink Access, TUA) [45, 46], which can also be effectively investigated using stochastic polling models.

4. SYSTEMATIZATION OF WORKS ON THE APPLICATION OF POLLING MODELS IN INTERNET OF THINGS NETWORKS AND 5G/6G CELLULAR NETWORKS

Internet of Things (IoT) networks have seen intensive development in recent years [47–49]. Stochastic polling models can be effectively used to evaluate characteristics and perform comparative analysis of protocols for such networks. Among the works devoted to this type of analysis, the following can be noted.

In [50], a Medium Access Control (MAC) protocol for IoT networks is analyzed using a star-type priority polling model. Such polling allows for separating users with different priorities to reduce delay and ensure fair access service. The model is represented by a system of $N+1$ $M/G/1$-type queues. Queue Q_h has priority over the other N queues and is served by the server according to a gated discipline each time after the service of any other queue is completed. The non-priority queues receive exhaustive service.

The vector describing the number of customers in the queues at the moment of polling Q_i has dimension $N+1$; therefore, the corresponding PGF of stationary probabilities has an additional argument z_h. The functional relations (1) in this case connect the PGFs $G_i(\mathbf{z}, z_h)$ and $G_{ih}(\mathbf{z}, z_h)$ as follows:

$$G_{i+1}(\mathbf{z}, z_h) = G_{ih}\left(\mathbf{z}, \tilde{B}_h\left(\sum_{j=1}^{N}\lambda_j(1-z_j) + \lambda_h(1-z_h)\right)\right), \tag{5}$$

$$G_{ih}(\mathbf{z}, z_h) = G_i\left(z_1, \ldots, z_{i-1}, \tilde{\theta}_i\left(\sum_{j=1}^{N}\lambda_j(1-z_j) + \lambda_h(1-z_h)\right), z_{i+1}, \ldots, z_N, z_h\right)$$
$$\times \tilde{S}_i\left[\sum_{j=1}^{N}\lambda_j(1-z_j) + \lambda_h(1-z_h)\right], \quad i = \overline{1, N}.$$

As shown in Sections 2 and 3, these functional equations allow for the computation of key performance characteristics. Note that here $\tilde{S}_i(s)$ is the LST of the server disconnection time from queue Q_i.

In [51], a neural network algorithm is used to predict and analyze network performance in Wireless Sensor Networks (WSN). The model enables solving priority tasks in the IoT domain. A neural network is constructed to predict system performance. It is demonstrated how the proposed priority scheme helps improve system performance compared to a similar polling system with a single-level priority.

In [52], a new fast polling algorithm for wireless mesh networks in narrowband IoT systems is presented. It is shown that the algorithm significantly minimizes the polling time required by the data collection and management center to receive responses from all remote telemetry devices. In [53], a polling model with two queues and a correlated Batch Markovian Arrival Process (BMAP) input stream is investigated for the design and optimization of high-definition video streaming in IoT networks.

Stochastic polling models can be effectively used for performance evaluation and design of protocols for next-generation 5G/6G cellular networks. The deployment of 5G/6G networks provides a significant expansion of the service range for users, increased network capacity, and higher data transfer speeds compared to previous generations [54]. To achieve high and ultra-high data rates in access and transport segments of next-generation networks, not only traditional centimeter-wave radio channels (up to 6 GHz) but also millimeter-wave channels (up to 100 GHz) with wide bandwidth are utilized [55, 56].

However, considering the propagation characteristics of millimeter waves (short coverage zones), the 3GPP consortium proposed the Integrated Access and Backhaul (IAB) technology [57] which integrates access and backhaul functions within a single network. This technology allows for the creation of a cost-effective network by using inexpensive relay nodes (IAB nodes) instead of fully equipped base stations, which is particularly important for deploying 5G networks along transportation corridors.

In [58], a polling model with two queues is used to analyze the performance of a terminal node in an IAB network. One queue stores data transmitted via the downlink from the parent IAB node, the second stores data transmitted via the uplink from child nodes and user equipment associated with the terminal node. In [59–61], a boundary node is modeled as an $M_K/G_K/1$-type polling system with cyclic polling and instantaneous server switchovers between queues ($s_i = 0$, $i = \overline{1, K}$) except for a period s_0 at the beginning of the cycle which can be interpreted as the connection time to queue Q_1. It is assumed that customers can only arrive at the queues during time s_0 representing a combination of globally-gated and exhaustive service. In this case, the PGF (1) for the number of customers at queue polling moments takes the form

$$G_i(\mathbf{z}) = \tilde{S}\left(\sum_{j=i}^{K}(\lambda_j(1 - z_j))\right), \quad i = \overline{1, K},$$

where $\tilde{S}(w)$ is the LST of the distribution function of the random variable s_0. Subsequently, moments of arbitrary order for the random variables X_i^j, $i, j = \overline{1, K}$, describing the number of customers in queues at polling moments, are obtained explicitly. The LST of the delay time distribution in queue Q_i with an arbitrary service time distribution function has the form

$$\tilde{\Delta}_i(w) = s^i \prod_{j=1}^{i} \frac{1}{s + \lambda_j(1 - \tilde{B}_j(w))}$$

with mean $\Delta_i = \frac{1}{s}\sum_{j=1}^{i} \lambda_j b_j$, $i = \overline{1, K}$.

Polling models have found application in modeling congestion control for SIP servers [62]. SIP (Session Initiation Protocol) is a data transmission protocol that describes a method for establishing and terminating a user communication session in 5G networks. In [62], a scheme for threshold-based control of load generated by non-priority clients is proposed. In [63], the analysis of the SIP server congestion control mechanism is continued by testing various service disciplines. The considered polling model assumes a threshold-based discipline for switching to a priority queue and two service disciplines: exhaustive and gated. It is shown that in some cases, using gated service for the priority queue is preferable.

In conclusion of this section, we also note the article [64] which considers a model with a controlled (dynamic) polling order to achieve the required level of fairness — the system's ability to provide the same quality of service to users in generalized scheduling algorithms in modern wireless networks, and the paper [65] which investigates various polling policies (rules for polling and service) to solve scalability problems of the standard Transmission Control Protocol (TCP).

5. APPLICATION OF STOCHASTIC POLLING MODELS IN TRANSPORTATION AND VEHICULAR TRAFFIC MANAGEMENT AT INTERSECTIONS

The application of polling models in traffic control has an extensive history, detailed in the review [9] published in 2011. The further development of this application area is reflected in a number of later articles. In particular, [66] considers a polling model with random duration of green traffic signals for each direction depending on its traffic load, aiming to minimize waiting time at an intersection. To model automated traffic light control regulating vehicle movement at an intersection, [67, 68] investigates a two-queue polling system with a threshold-type priority for switching the server between queues. The traffic light determines the crossing priority for two directions in a threshold-based manner: when the number of vehicles in any direction reaches a threshold value, the green light turns on for that direction. In [68], a matrix-analytic approach is applied to obtain the key performance characteristics of such a model. A variation of threshold-based strategies for switching the server between queues to regulate intersection traffic is also considered in [7].

In [69], an exact solution for a two-queue polling system with threshold priority is proposed. Simulation results of a polling mechanism at fully autonomous intersections are presented in [70].

Autonomous vehicles approaching an intersection adjust their speed to pass through the intersection while avoiding collisions.

For traffic flows, the assumption of independence of flows moving through an intersection is not realistic; therefore, accounting for the correlated nature of real traffic flows is a highly relevant task. In [71, 72], for a two-queue polling model, the correlated and non-stationary nature of input flows is accounted for using a Markovian Arrival Process (MAP) and a Marked Markovian Arrival Process (MMAP). In [71], two queues of the $MAP/PH/1/N$ type with alternating server service and phase-type service time distribution are considered. The authors provide a detailed analysis of waiting time in the queue, which can serve as a guide for analyzing waiting times in other polling models, not only through their mean values but also in terms of distributions. For example, the LST of the waiting time distribution for customers in the first queue has the form:

$$
\begin{aligned}
v_1(s) = \frac{1}{\lambda_1} \sum_{i_1=0}^{N_1-1} \sum_{i_2=0}^{N_2} \Bigg(& \boldsymbol{\pi}^{(-1)}(i_1, i_2) \left(I_{M_{-1}} \otimes \hat{D}_1^{(1)} \right) \mathbf{e} v_1^{(-1)}(s, i_1) \\
& + \boldsymbol{\pi}^{(-2)}(i_1, i_2) \left(I_{M_{-2}} \otimes \hat{D}_1^{(1)} \right) \mathbf{e} v_1^{(-2)}(s, i_1, i_2) \\
& + \sum_{j=1}^{N_1} \boldsymbol{\pi}^{(1)}(j, i_1, i_2) \left(I_{M_1} \otimes \hat{D}_1^{(1)} \right) \mathbf{e} v_1^{(1)}(s, j, i_1, i_2) \\
& + \sum_{j=1}^{N_2} \boldsymbol{\pi}^{(2)}(j, i_1, i_2) \left(I_{M_2} \otimes \hat{D}_1^{(1)} \right) \mathbf{e} v_1^{(2)}(s, j, i_1, i_2) \Bigg) + P_1^{(loss)},
\end{aligned}
$$

where $P_1^{(loss)}$ is the loss probability in queue Q_1, $\boldsymbol{\pi}^{(k)}(*)$ are stationary probabilities corresponding to state $(*)$ when the server visits queue Q_k, $v_1^{(k)}(s, *)$ is the LST of the conditional waiting time distribution in queue Q_1 given that the server is in state k and the system is in state $*$, $D_1^{(1)}$ is one of the matrices describing arrivals in the MAP flow.

In conclusion of this section, we also note other applications of polling models in transportation systems. In [73], the application of Radio Frequency Identification (RFID) technology [74] for monitoring the number of vehicles at an intersection to adapt traffic signals to the current traffic situation is investigated. For transportation systems, [75] analyzes the problem of improving priority order delivery without reducing the efficiency of regular order delivery. For this purpose, it is proposed to designate some order-picking machines as priority with exhaustive service discipline, while the others are non-priority with a service limit of one order (1-limited service discipline). In [76], a modification of service disciplines in a polling system describing an order delivery system is proposed to reduce the assembly time for multiple orders for a single customer.

6. APPLICATION OF STOCHASTIC POLLING MODELS IN HEALTHCARE SYSTEMS

Currently, polling system models are actively applied in the field of medical monitoring. For example in [77], a polling system is used to evaluate the performance of wireless body sensor networks which are modern monitoring systems using wireless technologies for disease prediction and diagnosis. The article [78] analyzes a sleep mode scheme in Wireless Body Area Networks (WBAN) based on the IEEE 802.15 standard via a discrete-time polling model with adaptive polling order, assuming the server skips empty queues if they were empty at the previous polling moment.

The model considered in [19] describes the initial patient intake process at a medical clinic. Patients are admitted by several nurses, each organizing their own queue of patients, and an on-duty doctor. A patient is directed to the nurse with the shortest queue for examination. Afterward, the doctor admits patients, each time choosing a patient from the longest queue. This scheme involves a combination of two strategies for customer arrival and service: joining an incoming customer to the queue with the shorter length and serving the queue with the larger number of customers by the server. This allows describing the behavior of the corresponding two-queue polling model with a two-dimensional Markov chain where the first component is the length of Q_1, and the second one is the current difference in queue lengths. The Markov chain in this case belongs to the class of quasi birth-and-death processes and has an infinitesimal generator of the form

$$Q = \begin{bmatrix} B_1 & B_0 & 0 & \dots & \dots & \dots \\ B_2 & A_1 & A_0 & 0 & \dots & \dots \\ 0 & A_2 & A_1 & A_0 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix},$$

where the elements of matrices A_i and B_i are input parameters of the system.

The vector $\mathbf{p}_n$ of stationary state probabilities of the system, corresponding to the number n of customers in queue Q_1 calculated as $\mathbf{p}_n = \mathbf{p}_1 R^{n-1}$, $n \geqslant 1$ where matrix R is the minimal non-negative solution to the matrix equation $A_0 + RA_1 + R^2 A_2 = O$. The vectors $\mathbf{p}_0$ and $\mathbf{p}_1$ are found from the system

$$\mathbf{p}_0 B_1 + \mathbf{p}_1 B_2 = \mathbf{0}, \quad \mathbf{p}_0 B_0 + \mathbf{p}_1 (A_1 + RA_2) = \mathbf{0}, \quad \mathbf{p}_0 \mathbf{e} + \mathbf{p}_1 [I - R]^{-1} \mathbf{e} = 1,$$

where $\mathbf{e} = (1, \dots, 1)$, I is the identity matrix. The mean length of queue Q_1 is calculated as $L_1 = \mathbf{p}_1 [I - R]^{-2} \mathbf{e}$.

In [79], it is proposed to use a polling scheme for scheduling edge computing to adapt scheduling methods for multiple services. An edge server provides services to a hospital information center. The authors consider the problem of maximizing the efficiency of edge server utilization using a star-type priority polling scheme. As in [50], the model has N ordinary queues and one priority queue. All queues receive exhaustive service except Q_h visited by the server each time the server leaves a non-priority queue. The PGFs of the number of customers in queues at polling moments satisfy functional relations similar to (5) where the LST $\tilde{B}_h(s)$ is replaced by $\tilde{\theta}_h(s)$ defined by formula (3).

7. OTHER APPLICATIONS OF POLLING SYSTEMS

Polling systems are also used in other areas not covered in previous sections, such as passive optical networks, robotics, satellite systems, and production processes.

Passive Optical Networks (PON) are a telecommunications technology that uses fiber-optic communication channels to provide users with high-speed Internet access, video, and voice services. The term *passive* stems from the use of passive optical splitters which do not require active components to distribute the optical signal to multiple endpoints. In [80], the issue of the existence of a stationary regime in a multi-server polling system with limited service, modeling the data transmission process in networks based on PON technology, is investigated.

In [81], a scheme for accumulating and servicing customers in polling systems modeling data transmission based on PON technology is proposed. This is the so-called *multi-phase gated* service discipline. A customer arriving at queue Q_i must wait K_i cycles before it is served, $i = \overline{1, N}$. As the authors note, such a multi-phase discipline helps solve the problem of server monopolization by more loaded queues by selecting appropriate values $(K_1, \dots, K_N)$. Despite its complexity, this discipline allows for an exact analysis of the system and explicit derivation of performance characteristics.

The system state at the moment of polling queue Q_1 (start of a cycle) is described by the set of random variables $X_{i,n}^{(k)}$, $k = \overline{1, K_i}$, $i = \overline{1, N}$, $n \geqslant 0$, where $X_{i,n}^{(k)}$ is the number of customers in queue Q_i at phase k at the nth moment of polling queue Q_1. The process $X^{(n)} = \left(X_{1,n}^{(1)}, \ldots, X_{1,n}^{(K_1)}, \ldots, X_{N,n}^{(1)}, \ldots, X_{N,n}^{(K_N)}\right)$, $n \geqslant 0$ is a K-dimensional multi-type branching process with immigration [32], with the PGF for the number of offspring:

$$f(\mathbf{s}) = \left(f^{(1,1)}(\mathbf{s}), \ldots, f^{(1,K_1)}(\mathbf{s}), \ldots, f^{(N,1)}(\mathbf{s}), \ldots, f^{(N,K_N)}(\mathbf{s})\right),$$

where $\mathbf{s} = \left(s_1^{(1)}, \ldots, s_1^{(K_1)}, \ldots, s_N^{(1)}, \ldots, s_N^{(K_N)}\right)$, and for $i = \overline{1,N}$

$$f^{(i,k)}(\mathbf{s}) = s_i^{(k+1)}, \quad k = \overline{1, K_i - 1},$$
$$f^{(i,K_i)}(\mathbf{s}) = \tilde{B}_i \left(\sum_{j=1}^{i} \lambda_j (1 - s_j^{(1)}) + \sum_{j=i+1}^{N} \lambda_j (1 - f^{(j,1)}(\mathbf{s}))\right),$$

and the PGF of immigration is

$$g(\mathbf{s}) = \prod_{i=1}^{N} \tilde{S}_i \left(\sum_{j=1}^{i} \lambda_j (1 - s_j^{(1)}) + \sum_{j=i+1}^{N} \lambda_j (1 - f^{(j,1)}(\mathbf{s}))\right).$$

The pseudo-conservation law (expression for the weighted sum of mean waiting times in the system) for the multi-phase gated discipline is

$$\sum_{i=1}^{N} \rho_i \mathbf{E}[W_i] = \rho \frac{\rho}{1-\rho} \frac{b^{(2)}}{2b} + \rho \frac{s^{(2)}}{2s} + \frac{s}{2(1-\rho)} \left[\rho^2 - \sum_{i=1}^{N} \rho_i^2\right] + \sum_{i=1}^{N} \mathbf{E}[M_i]$$

where M_i is the mean amount of work remaining in queue Q_i when the server leaves queue Q_i, $b^{(2)} = \sum_{i=1}^{N} \frac{\lambda_i}{\Lambda} b_i^{(2)}$, $b = \sum_{i=1}^{N} \frac{\lambda_i}{\Lambda} b_i$, $\Lambda = \sum_{i=1}^{N} \lambda_i$. For the multi-phase gated discipline, $\mathbf{E}[M_i] = \rho_i((K_i - 1) + \rho_i)\frac{s}{1-\rho}$, $i = \overline{1,N}$.

In [82], another model for data reservation and transmission in PON networks is presented: a polling system with retrials and so-called *glue periods*. Customers arriving at queue Q_i can accumulate in the queue and then receive service only during a so-called glue period, which has a fixed length G_i. Customers arriving outside this period go into orbit and make repeated attempts to join the queue during the glue period. The repeated attempts from the orbit of queue Q_i form a Poisson process with parameter ν_i, $i = \overline{1,N}$.

The PGF for the number of customers in queues at the start of the glue period for queue Q_1 has the explicit form

$$X(\mathbf{z}) = \prod_{m=0}^{\infty} K\left(h_1^{(m)}(\mathbf{z}), h_2^{(m)}(\mathbf{z}), \ldots, h_N^{(m)}(\mathbf{z})\right), \quad i = \overline{1,N},$$

where $h_i^{(0)}(\mathbf{z}) = z_i$, $h_i^{(n)}(\mathbf{z}) = h_i\left(h_i^{(n-1)}(\mathbf{z}), h_2^{(n-1)}(\mathbf{z}), \ldots, h_N^{(i-1)}(\mathbf{z})\right)$, the functions $h_i(\mathbf{z})$, $i = \overline{1,N}$ are defined as

$$h_i(\mathbf{z}) = f_i\left(z_1, \ldots, z_i, h_{i+1}(\mathbf{z}), \ldots, h_N(\mathbf{z})\right), \; i = \overline{1, N-1}, \quad h_N(\mathbf{z}) = f_N(\mathbf{z}),$$

where $f_i(\mathbf{z}) = (1 - e^{-\nu_i G_i})\beta_i(\mathbf{z}) + e^{-\nu_i G_i} z_i$, $\beta_i(\mathbf{z}) = \tilde{B}_i\left(\sum_{j=1}^{N} \lambda_j (1 - z_j)\right)$, $i = \overline{1,N}$.

The function $K(\mathbf{z})$ is defined as follows:

$$K(\mathbf{z}) = \prod_{i=1}^{N} \sigma_i\left(z_i, \ldots, z_i, h_{i+1}(\mathbf{z}), \ldots, h_N(\mathbf{z})\right) \prod_{i=1}^{N} e^{-G_i D_i(\mathbf{z})}$$

where

$$D_i(\mathbf{z}) = \sum_{j=1}^{i-1} \lambda_j(1 - z_j) + \lambda_i \left(1 - \beta_i \left(z_i, \dots, z_i, h_{i+1}(\mathbf{z}), \dots, h_N(\mathbf{z})\right)\right) + \sum_{j=i+1}^{N} \lambda_j \left(1 - h_j(\mathbf{z})\right),$$

$$\sigma_i(\mathbf{z}) = \tilde{S}_i \left(\sum_{j=1}^{N} \lambda_j(1 - z_j) \right).$$

Here, this refers to the LST of the disconnection time from queue Q_i, unlike the main model in Section 2, $i = \overline{1, N}$. In [82], the issue of optimizing the duration of glue periods is also considered.

Wireless Local Area Networks. The authors in [83] analyze a continuous polling system without separate queues, represented by a common two-dimensional waiting area for all customers, where arrivals occur randomly. The system models a Ferry-based Wireless Local Area Network (FWLAN). FWLAN is a concept for a wireless local area network where communication between nodes (users) and the external world or between nodes is possible via a mobile relay — a *ferry*. In such a network, isolated nodes are scattered over a certain area, and direct global communication between them is impossible. The moving relay, traveling along a predetermined cyclic route, communicates with static network nodes (or users) via a wireless communication channel. Mobile base stations are considered in the context of Mobile Ad-Hoc Networks (MANETs), Vehicular Ad-hoc Networks (VANETs), and wireless (static) sensor networks.

Networks-on-Chip. In [84, 85], the case of three-phase gated service is considered as a scheduling algorithm for Medium Access Control (MAC) in a Network-on-Chip (NoC) communication subsystem taking into account the specific features of the two-dimensional interconnection structure.

Energy Saving. In [86], an analysis of a polling system with two queues and time-limited queue service is presented from the perspective of energy consumption. The study primarily focuses on the balance between processor energy consumption in personal computers and their performance for users. It is assumed that in the polling model, the service rate of customers in a queue is not constant but depends on the queue load. This approach allows personal computers to adjust data processing speed and optimize energy consumption depending on the volume of tasks being performed.

Memory Usage Control. The model proposed in [67, 68] has another application — controlling the occupied memory volume in data centers when the data volume on individual disks exceeds a certain limit (threshold), leading to inefficient operation and necessitating memory cleanup.

Robotics. In [87], the use of a polling mechanism for processing environmental and social information received from a robot swarm is described. It is shown that the swarm's ability to adapt to load improves as the number of communication channels between robots decreases, which requires solving the corresponding optimization problem.

Satellite Systems. In [88], a polling model considering time-division duplex communication is investigated for developing a unified control system in global navigation satellite systems to improve data interaction.

Production Processes. In [8], a tandem of two $M/M/1$-type polling systems each having two queues is analyzed. This model represents the aluminum rolling production process. Each production stage is modeled by a polling system which describes the process of sequentially processing different types of alloys.

8. CONCLUSION

This review systematizes new directions of practical application for stochastic polling models based on an analysis of results from numerous articles by domestic and international authors published in recent years. The application of such models for evaluating performance characteristics and designing broadband wireless metropolitan networks, 5G/6G cellular networks, Internet of Things networks, transportation systems, and healthcare systems has been considered. The effectiveness of applying polling models in other areas, including robotics, satellite systems, passive optical networks, and production processes, has been briefly noted. For individual works of theoretical and/or practical interest, the review provides descriptions of original mathematical models of stochastic polling (differing from the main model discussed in Section 2 of the article) and the results of their investigation.

Directions for further development of applied work in the field of stochastic polling have been discussed. These include research on polling models with correlated input flows and generalized Phase-Type (PH) service time distributions, which adequately describe the operation of modern computing systems and networks; the application and development of machine learning methods and the matrix-geometric approach for predicting and evaluating the characteristics of systems with centralized control mechanisms. Additionally, a promising direction for further applied research is modeling new protocols for using multiple frequency channels and serving multiple users, as well as new access mechanisms such as Triggered Uplink Access (TUA).

The presented review summarizes applied research on stochastic polling systems, enables the formulation of new problems, and defines directions for future work in this promising area of queueing theory.

REFERENCES

1. Takagi, H., *Analysis of Polling Systems*, MIT Press, 1986.
2. Borst, S.C., *Polling Systems*, Stichting Mathematisch Centrum, 1996.
3. Borst, S.C. and Boxma, O.J., Polling: past, present, and perspective, *TOP*, 2018, vol. 26, pp. 335–369.
4. Vishnevskii, V.M. and Semenova, O.V., Mathematical methods to study the polling systems, *Autom. Remote Control*, 2006, vol. 67, pp. 173–220.
5. Vishnevsky, V. and Semenova, O., *Polling Systems: Theory and Applications for Broadband Wireless Networks*, LAMBERT Academic Publishing, 2012.
6. Vishnevsky, V. and Semenova, O., Polling systems and their application to telecommunication networks, *Mathematics*, 2021, vol. 9, no. 2, Art. no. 117.
7. Perel, E. and Yechiali, U., Two-queue polling systems with switching policy based on the queue that is not being served, *Stoch. Models*, 2017, vol. 33, pp. 430–450.
8. Suman, R. and Krishnamurthy, A., Analysis of tandem polling queues with finite buffers, *Ann. Oper. Res.*, 2019, vol. 293, pp. 343–369.
9. Boon, M.A.A., van der Mei, R.D., and Winands, E.M.M., Applications of polling systems, *SURV. Oper. Res. Manag. Sci.*, 2011, vol. 16, no. 2, pp. 67–82.
10. Dudin, A.N., Klimenok, V.I., and Vishnevsky, V.M., *The Theory of Queuing Systems with Correlated Flows*, Springer, 2020.
11. Chakravarthy, S.R., Queueing models in services — analytical and simulation approach, in *Queueing Theory 2: Advanced Trends*, Wiley, 2021, pp. 33–80.
12. Vishnevsky, V.M., Efrosinin, D.V., and Stepanova, N.V., *Queueing Theory and Machine Learning*, Springer, 2026.
13. Efrosinin, D., Vishnevsky, V., and Stepanova, N., Optimal scheduling in general multi-queue system by combining simulation and neural network techniques, *Sensors*, 2023, vol. 23, no. 12, Art. no. 5479.
14. Winands, E.M.M., Adan, I.J.B.F., and van Houtum, G.J., Mean value analysis for polling systems, *Queueing Syst.*, 2006, vol. 54, pp. 35–44.

15. Konheim, A.G., Levy, H., and Srinivasan, M.M., Descendant set: an efficient approach for the analysis of polling systems, *IEEE Trans. Commun.*, 1994, vol. 42, no. 2–4, pp. 1245–1253.

16. Rykov, V.V., On analysis of periodic polling systems, *Autom. Remote Control*, 2009, vol. 70, pp. 997–1018.

17. Yechiali, U., Analysis and control of polling systems, in *Performance Evaluations of Computer and Communication Systems*, 1993, pp. 630–650.

18. Hirayama, T., Analysis of multiclass Markovian polling systems with feedback and composite scheduling algorithms, *Ann. Oper. Res.*, 2012, vol. 198, no. 1, pp. 83–123.

19. Perel, E., Perel, N., and Yechiali, U., A polling system with "Join the shortest — serve the longest" policy, *Comput. Oper. Res.*, 2019, vol. 114, Art. no. 104809.

20. Boon, M.A.A. and Winands, E.M.M., Heavy-traffic analysis of k-limited polling systems, *Probab. Eng. Inf. Sci.*, 2014, vol. 28, pp. 451–471.

21. Bekker, R., Vis, P., Dorsman, J.L., van der Mei, R.D., and Winands, E.M.M., The impact of scheduling policies on the waiting-time distributions in polling systems, *Queueing Syst.*, 2015, vol. 79, pp. 145–172.

22. Boxma, O.J. and Groenendijk, W.P., Pseudo conservation laws in cyclic-service systems, *J. Appl. Probab.*, 1987, vol. 24, pp. 949–964.

23. IEEE Standard for Information Technology — Telecommunications and Information Exchange between Systems — Local and Metropolitan Area Networks — Specific Requirements — Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: 320MHz Positioning, 2025.

24. Vishnevsky, V.M., Dudin, A.N., Semenova, O.V., and Klimenok, V.I., Performance analysis of the $BMAP/G/1$ queue with gated servicing and adaptive vacations, *Perform. Eval.*, 2011, vol. 68, no. 5, pp. 446–462.

25. Vishnevsky, V.M., Semenova, O.V., and Bui, D.T., Investigation of the stochastic polling system and its applications to broadband wireless networks, *Autom. Remote Control*, 2021, vol. 82, no. 9, pp. 1607–1613.

26. Vishnevsky, V.M., Semenova, O.V., and Bui, D.T., Using a machine learning approach for analysis of polling systems with correlated arrivals, *Lect. Notes Comput. Sci.*, 2021, vol. 13144, pp. 336–345.

27. Vishnevsky, V.M., Lyakhov, A.I., and Guzakov, N.N., An adaptive polling strategy for IEEE 802.11 PCF, *Proc. 7th Int. Symp. Wirel. Pers. Multimed. Commun.*, 2004, vol. 1, pp. 87–91.

28. Vishnevsky, V. and Lyakhov, A., Analytical study of IEEE 802.11 PCF for regional and metropolitan area networks, *Cybern. Inf. Technol.*, 2005, vol. 5, no. 2, pp. 117–136.

29. Vishnevsky, V., Lyakhov, A., Portnoy, S., and Shakhnovich, I., *Broadband Wireless Information Transmission Networks*, Tekhnosfera, 2005.

30. Leonovich, A. and Ferng, H.-W., Modeling the IEEE 802.11e HCCA mode, *Wirel. Netw.*, 2013, vol. 19, no. 5, pp. 771–783.

31. Meyfroyt, T.M.M., Boon, M.A.A., Borst, S.C., and Boxma, O.J., Performance of large-scale polling systems with branching-type and limited service, *Perform. Eval.*, 2019, vol. 133, pp. 1–24.

32. Resing, J.A.C., Polling systems and multitype branching processes, *Queueing Syst.*, 1993, vol. 13, pp. 413–426.

33. Vishnevsky, V.M., Semenova, O.V., Nguyen, V.H., and Dang, M.C., Batch service polling system: mathematical analysis and simulation modeling, *Lect. Notes Comput. Sci.*, 2024, vol. 14123, pp. 142–155.

34. Vishnevsky, V.M., Avramenko, Y.A., Kalmykov, N.S., and Nguyen, V.H., Broadband Wireless networks based on tethered high-altitude unmanned platforms, *Commun. Comput. Inf. Sci.*, 2024, vol. 2129, pp. 101–112.

35. Belmekki, B.E.Y. and Alouini, M.-S., Unleashing the potential of networked tethered flying platforms: prospects, challenges, and applications, *IEEE Open J. Veh. Technol.*, 2022, vol. 3, pp. 278–320.

36. Marques, M.N., Magalhães, S.A., Dos Santos, F.N., and Mendonca, H.S., Tethered unmanned aerial vehicles — a systematic review, *Robotics*, 2023, vol. 12, no. 4, p. 117.

37. Zhao, W., Zhang, J., and Li, D., Clustering and beamwidth optimization for UAV-assisted wireless communication, *Sensors*, 2023, vol. 23, no. 23, Art. no. 9614.

38. Nazarov, A.A., Shulgina, K.S., Salimzyanov, R.R., et al., A Study of a Mathematical Model of a Wireless Data Transmission Network as a Cyclic Random Multiple Access System, *Tekhnologii Bezopasnosti Zhiznedeyatel'nosti*, 2023, no. 2, pp. 58–65.

39. Nazarov, A., Phung-Duc, T., Shulgina, K., et al., Cyclic retrial queue for building data transmission networks, *Lect. Notes Inst. Comput. Sci. Soc.-Inform. Telecommun. Eng.*, 2023, vol. 482, pp. 167–179.

40. Dorsman, J.-P.L., Borst, S.C., Boxma, O.J., et al., Markovian polling systems with an application to wireless random-access networks, *Perform. Eval.*, 2015, vol. 85–86, pp. 33–51.

41. Liu, Z., Chu, Y., and Wu, J., On the three-queue priority polling system with threshold service policy, *J. Appl. Math. Comput.*, 2017, vol. 53, pp. 445–470.

42. Fiems, D. and Altman, E., Gated polling with stationary ergodic walking times, Markovian routing and random feedback, *Ann. Oper. Res.*, 2012, vol. 198, no. 1, pp. 145–164.

43. Ding, H., Li, C., Bao, L., et al., Research on multi-level priority polling MAC protocol in FPGA tactical data chain, *IEEE Access*, 2019, vol. 7, pp. 33506–33516.

44. Chernova, N., Foss, S., and Kim, B., On the stability of a polling system with an adaptive service mechanism, *Ann. Oper. Res.*, 2012, vol. 198, pp. 125–144.

45. Morais, D.H., Wi-Fi 6/6E and Wi-Fi 7 Overview, in *5G/5G-Advanced, Wi-Fi 6/7, and Bluetooth 5/6*, Springer, 2025.

46. Karamyshev, A., Levitsky, I., Bankov, D., et al., A Tutorial on Wi-Fi 8: the journey to ultra high reliability, *Probl. Inf. Transm.*, 2025, vol. 61, pp. 164–210.

47. Rahaman, M., A review on Internet of Things-IoT- architecture, technologies, future applications & challenges, *Int. J. Sci. Bus.*, 2022, vol. 14, no. 1, pp. 80–92.

48. Nguyen, D.C., Ding, M., Pathirana, P.N., et al., 6G Internet of Things: a comprehensive survey, *IEEE Internet Things J.*, 2022, pp. 359–383.

49. Wei, Z., Jia, J., Niu, Y., Wang, L., et al., Integrated sensing and communication channel modeling: a survey, *IEEE Internet Things J.*, 2025, pp. 18850–18864.

50. Yang, Z., Mao, L., Ding, H., et al., Research on two-level priority polling access control protocol based on continuous time, *Proc. 2020 IEEE 6th Int. Conf. Comput. Commun. (ICCC)*, 2020, pp. 136–140.

51. Yang, Z., Mao, L., Yan, B., Wang, J., and Gao, W., Performance analysis and prediction of asymmetric two-level priority polling system based on BP neural network, *Appl. Soft Comput.*, 2021, vol. 99, Art. no. 106880.

52. Ar-Reyouchi, E.M., Maslouhi, I., and Ghoumid, K., A new fast polling algorithm in wireless mesh network for narrowband Internet of Things, *Telecommun. Syst.*, 2020, vol. 74, no. 3, pp. 405–410.

53. Cao, J., Feng, W., Chen, Y., et al., Performance analysis of a polling model with BMAP and across-queue state-dependent service discipline, *IEEE Access*, 2019, vol. 7, pp. 127230–127253.

54. Molchanov, D.A., Begishev, V.O., Samuilov, K.E., et al., *Seti 5G/6G: arkhitektura, tekhnologii, metody analiza i rascheta*, RUDN, 2022.

55. Cudak, M., Ghosh, A., Ghosh, A., et al., Integrated access and backhaul: A key enabler for 5G millimeter-wave deployments, *IEEE Commun. Mag.*, 2021, vol. 59, no. 4, pp. 88–94.

56. Gupta, M., Rao, A., Visotsky, E., et al., Learning link schedules in self-backhauled millimeter wave cellular networks, *IEEE Trans. Wirel. Commun.*, 2020, vol. 19, no. 12, pp. 8024–8038.

57. 3GPP TR 38.901 V14.1.1, Study on channel model for frequencies from 0.5 to 100 GHz (Release 17).

58. Nikolaev, D. and Gaidamaka, Y., Leaf node polling model analysis in an integrated access and backhaul network, *Commun. Comput. Inf. Sci.*, 2023, vol. 2163.

59. Beschastny, V.A., Machnev, E.A., Ostrikova, D.Yu., et al., Probabilistic Characteristics of Packet Service Process in an IAB Network Node, *Elektrosvyaz*, 2024, no. 1, pp. 43–46.

60. Nikolaev, D., Gorshenin, A., and Gaidamaka, Y., Polling model for analysis of round-trip time in the IAB network, *Lect. Notes Comput. Sci.*, 2025, vol. 15460, pp. 219–241.

61. Nikolaev, D.I., Beschastny, V.A., and Gaidamaka, Y.V., Two-queue polling system as a model of an integrated access and backhaul network node in half-duplex mode, *Discrete Contin. Models Appl. Comput. Sci.*, 2024, vol. 32, no. 4, pp. 362–369.

62. Gaidamaka, Yu.V., Model with threshold control for analysing a server with SIP protocol in the overload mode, *Autom. Control Comput. Sci.*, 2013, vol. 47, pp. 211–218.

63. Shorgin, S., Samouylov, K., Gaidamaka, Y., et al., Polling system with threshold control for modeling of SIP server under overload, *Adv. Intell. Syst. Comput.*, 2014, vol. 240, pp. 97–107.

64. Singh, V. and Kavitha, V., Fair opportunistic schedulers for Lossy Polling systems, *arXiv preprint arXiv:2208.05846*, 2022.

65. Czerniak, O., Altman, E., and Yechiali, U., Orchestrating parallel TCP connections: cyclic and probabilistic polling policies, *Perform. Eval.*, 2012, vol. 69, no. 3–4, pp. 150–163.

66. Nimisha, M., Manoharan, M., and Krishnamoorthy, A., Polling models: a short survey and some new results, *Queueing Models Serv. Manag.*, 2024, vol. 7, no. 1, pp. 25–42.

67. Avrachenkov, K., Perel, E., and Yechiali, U., Finite-buffer polling system with threshold-based switching policy, *TOP*, 2016, vol. 24, pp. 541–571.

68. Jolles, A., Perel, E., and Yechiali, U., Alternating server with non-zero switchover times and opposite-queue threshold-based switching policy, *Perform. Eval.*, 2018, vol. 126, pp. 22–38.

69. Choi, D.I. and Lim, D.-E., An analytical analysis of Markovian queueing model with an alternating server and state-dependent alternating priority policy, *ICIC Express Lett.*, 2024, vol. 18, no. 4, pp. 385–391.

70. Miculescu, D. and Karaman, S., Polling-systems-based autonomous vehicle coordination in traffic intersections with no traffic signals, *IEEE Trans. Autom. Control*, 2020, vol. 65, pp. 680–694.

71. Dudin, A. and Sinyugina, Y., Analysis of the polling system with two Markovian arrival flows, finite buffers, gated service and phase-type distribution of service and switching times, *Commun. Comput. Inf. Sci.*, 2021, vol. 1605, pp. 1–15.

72. Dudin, A. and Dudina, O., Analysis of polling queueing system with two buffers and varying service rate, *Lect. Notes Comput. Sci.*, 2025, vol. 15460, pp. 129–143.

73. Semenova, O.V. and Lykov, S.S., Application of RFID Technology for Controlling Regulated Intersections, in *Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN-2013)*, Technosphere, 2013, pp. 318–324.

74. Abramian, V.L., Vishnevsky, V.M., and Larionov, A.A., Radio frequency identification in transport applications, *Control Sci.*, 2024, no. 1, pp. 2–12.

75. Ran, W., Liu, S., and Zhang, Z., A polling-based dynamic order-picking system considering priority orders, *Complexity*, 2020, vol. 2020, Art. ID 4595316.

76. Claeys, D., Adan, I., and Boxma, O., Stochastic bounds for order flow times in parts-to-picker warehouses with remotely located order-picking workstations, *Eur. J. Oper. Res.*, 2016, vol. 254, no. 3, pp. 895–906.

77. Kaur, M., Bajaj, R., and Kaur, N., Wireless Body Area Network: performance analysis of polling access MAC protocol, *Proc. 2021 2nd Glob. Conf. Adv. Technol. (GCAT)*, 2021, pp. 1–7.

78. He, M., Guan, Z., Wu, Z., et al., A polling access control with exhaustive service in wireless body area networks for mobile healthcare using the sleeping schema, *J. Ambient Intell. Humaniz. Comput.*, 2019, vol. 10, no. 10, pp. 3761–3774.

79. Wang, W., Yang, Z., and Ding, H., Application of polling scheduling in mobile edge computing, *Axioms*, 2023, vol. 17, no. 7, Art. no. 709.

80. Antunes, N., Fricker, C., and Roberts, J., Stability of multi-server polling system with server limits, *Queueing Syst.*, 2011, vol. 68, pp. 229–235.

81. van der Mei, R.D. and Roubos, A., Polling models with multi-phase gated service, *Ann. Oper. Res.*, 2012, vol. 198, no. 1, pp. 25–56.

82. Abidini, M.A., Boxma, O., and Resing, J., Analysis and optimization of vacation and polling models with retrials, *Perform. Eval.*, 2016, vol. 98, pp. 52–69.

83. Kavitha, V. and Altman, E., Continuous polling models and application to ferry assisted WLAN, *Ann. Oper. Res.*, 2012, vol. 198, pp. 185–218.

84. Bao, L., Zhao, D., and Zhao, Y., A priority-based polling scheduling algorithm for arbitration policy in Network on Chip, *J. Electron. (China)*, 2012, vol. 29, no. 1–2, pp. 120–127.

85. Ling, Y., Liu, C., and Li, Y., Study on queue strategy of gated polling multi-access communication system, *Lect. Notes Electr. Eng.*, 2012, vol. 124, pp. 99–105.

86. Boxma, O.J., Saxena, M., and Janssen, A., Two queues with time-limited polling and workload-dependent service speeds, *Stoch. Models*, 2021, vol. 37, no. 2, pp. 265–299.

87. Talamali, M.S., Saha, A., Marshall, J.A., and Reina, A., When less is more: Robot swarms adapt better to changes with constrained communication, *Sci. Robot.*, 2021, vol. 6, p. eabf1416.

88. Yan, Z., Zhao, K., Li, W., et al., Topology design for GNSSs under polling mechanism considering both inter-satellite links and ground-satellite links, *IEEE Trans. Veh. Technol.*, 2022, vol. 71, no. 2, pp. 2084–2097.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 1, pp. 18–28.
© The Author(s), 2026 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2026.
Russian Text © The Author(s), 2026, published in Avtomatika i Telemekhanika, 2026, No. 1, pp. 28–41.

On an Algebro-Geometric Approach to the Study of Stability with Respect to a Part of the Variables of Linear Systems with Constant Coefficients

V. I. Nikonov

National Research Mordovia State University, Saransk, Russia
e-mail: nik_vl_@mail.ru

Received March 11, 2025
Revised July 8, 2025
Accepted July 23, 2025

Abstract—An algebraic-geometric approach to the study of the stability of linear systems with constant coefficients with respect to some of the variables is proposed. The proposed approach establishes a relationship between the conditions of partial stability of a linear system and the existence of invariant subspaces in the original and dual spaces. An example is given to illustrate the results obtained.

Keywords: partial stability, minimal annihilating polynomial, invariant subspace, annihilator

DOI: 10.7868/S1608303226010023

1. INTRODUCTION

As noted in [1], interest in the problem of stability of linear systems with respect to some variables arose at the beginning of the 20th century. By now, the theory of stability of linear systems with constant coefficients is almost complete. The main stages of development of this theory are reflected in [1–8]. In the series of works [9–11] on achieving coordinate consensus in multi-agent systems, it was established that the problem being solved, with a special choice of control action and a linear transformation of the phase space, is reduced to a problem of partial stability of a linear system with constant coefficients. In [12] the problem of partial construction of solutions was considered, which, as noted by the authors, is “consonant” with the stability problems [1, 2]. The results of these studies allow us to obtain necessary and sufficient stability conditions for selected components of the phase vector of an arbitrary system of linear differential equations with constant coefficients. This algebraic approach was further developed in the works [13, 14].

Analyzing the obtained results, it can be noted that the main approaches to the study of partial stability of linear systems with constant coefficients are the Lyapunov function method, the auxiliary μ-systems method and methods based on differential and computer algebra methods. A stability criterion for linear systems was obtained in [2] and formulated in terms of μ-systems. Another stability criterion in [15] relies on the Jordan normal form of the linear operator defined by the system matrix in the original basis space. In [16], the geometric nature of partial stability is noted. A geometric approach to the study of partial stability is proposed in [17, 18].

This article is devoted to the study of stability of linear systems with constant coefficients. Using the algebraic-geometric properties of the original system, necessary and sufficient conditions for the stability of linear systems are obtained, expressed in terms of dual linear spaces on which dual linear operators act. It turns out that if, under certain conditions, some parameters of the system under study lie in the invariant subspace of one of the linear operators, then the problem of partial stability is broken down into two simpler problems. The obtained results establish a connection between the stability of the system and the internal geometry in the space of parameters that determine the dynamics of the system under study. The geometric properties of the system make it possible to obtain conditions under which the property of invariance with respect to the property of partial stability arises.

2. PROBLEM STATEMENT

The stability with respect to a given part of the components of the phase vector of the system is investigated

$$\frac{dx(t)}{dt} = A_* x(t), \tag{1}$$

with the following notations: x is the n-dimensional phase vector, A_* is a constant matrix with elements from the numerical field $\mathbb{F}$, $t \in \mathbb{R}$, $t \geq 0$.

Let's assume that we are investigating the stability of the first m components of the phase vector x of the system (1). We denote the first group of coordinates of the phase vector by the vector y, and the remaining components will form the vector z. In this regard, we represent the system (1) in the form

$$\begin{aligned} \frac{dy(t)}{dt} &= Ay(t) + Bz(t), \\ \frac{dz(t)}{dt} &= Cy(t) + Dz(t), \end{aligned} \tag{2}$$

where $y \in \mathbb{F}^m,\ z \in \mathbb{F}^p,\ A \in \mathbb{F}^{m\times m},\ B \in \mathbb{F}^{m\times p},\ C \in \mathbb{F}^{p\times m},\ D \in \mathbb{F}^{p\times p},\ n = m + p,\ p > 0.$

It is required to determine the conditions on the coefficients of the system under study so that the solution $x(t) \equiv 0$ of the system (1), (2) is stable with respect to the components included in the vector y, i.e. y-stable.

3. SOME INFORMATION FROM LINEAR ALGEBRA

3.1. Conjugate Space. Dual Spaces. Annihilators

Let $\mathcal{V}$ be an arbitrary linear space over a field $\mathbb{F}$.

Definition 1 [19, p. 33]. A function $\xi : \mathcal{V} \to \mathbb{F}$ is called a linear functional if it is a homomorphism of linear spaces, i.e., if

$$\xi(\nu_1 + \nu_2) = \xi(\nu_1) + \xi(\nu_2)$$

and

$$\xi(k\nu) = k\xi(\nu)$$

$\forall \nu, \nu_1, \nu_2 \in \mathcal{V}$, $\forall k \in \mathbb{F}$. Linear functionals are also called covectors of the space $\mathcal{V}$.

The set of all linear functionals is a subspace of the space of all functions on $\mathcal{V}$ and is denoted by $\mathcal{V}'$.

Definition 2 [19, p. 33]. The linear space $\mathcal{V}'$ is called the dual space of $\mathcal{V}$.

A basis $e^1, \ldots, e^n$ of $\mathcal{V}'$ that satisfies the relation

$$e^j(e_i) = \delta_i^j, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n,$$

is called the adjoint or dual basis of $e_1, \ldots, e_n$ of $\mathcal{V}$, where δ_i^j is the Kronecker delta.

Let $\mathcal{V}$ and $\mathcal{W}$ be two linear spaces over a field $\mathbb{F}$. Suppose that any two vectors $\nu \in \mathcal{V}$, $w \in \mathcal{W}$ are assigned a number $\nu \mapsto \langle \nu, w \rangle \in \mathbb{F}$ such that the following conditions are satisfied:

(1) For each fixed $w \in \mathcal{W}$, the function $\nu \mapsto \langle \nu, w \rangle$ is a linear functional on $\mathcal{V}$;
(2) For each fixed $\nu \in \mathcal{V}$, the function $w \mapsto \langle \nu, w \rangle$ is a linear functional on $\mathcal{W}$;
(3) For every vector $\nu \in \mathcal{V}$, there exists a vector $w \in \mathcal{W}$ such that $\langle \nu, w \rangle \neq 0$, and conversely, for every vector $w \in \mathcal{W}$, there exists a vector $\nu \in \mathcal{V}$ such that $\langle \nu, w \rangle \neq 0$.

Definition 3 [19, p. 36]. A function $\nu, w \mapsto \langle \nu, w \rangle$ satisfying conditions (a), (b), and (c) is called a pairing between the spaces $\mathcal{V}$ and $\mathcal{W}$. Spaces $\mathcal{V}$ and $\mathcal{W}$ for which at least one pairing exists are called dual. Notation: $\mathcal{V}|\mathcal{W}$.

Proposition 1 [19, p. 36]. *The linear space $\mathcal{V}$ is dual to the dual space $\mathcal{V}'$:*

$$\mathcal{V}|\mathcal{V}'.$$

Setting for any $\nu \in \mathcal{V}$ and $\varphi \in \mathcal{V}'$

$$\langle \nu, \varphi \rangle := \varphi(\nu),$$

we see that the indicated relation defines a pairing of the spaces $\mathcal{V}$ and $\mathcal{V}'$.

Let $S \subset \mathcal{V}$ be an arbitrary subset of the linear space $\mathcal{V}$.

Definition 4 [19, p. 40]. The set of all linear functionals $\xi \in \mathcal{V}'$ that vanish on any vector $\nu \in S$ is called the annihilator of the set S and is denoted by the symbol $\operatorname{Ann}(S)$.

Theorem 1 [20, p. 41]. *If $\mathcal{M}$ is an m-dimensional subspace of an n-dimensional vector space $\mathcal{V}$, then $\operatorname{Ann}(\mathcal{M})$ is an $(n-m)$-dimensional subspace of $\mathcal{V}'$.*

Theorem 2 [20, p. 42]. *If $\mathcal{M}$ is a subspace of a finite-dimensional vector space $\mathcal{V}$, then*

$$\operatorname{Ann}(\operatorname{Ann}(\mathcal{M})) = \mathcal{M}.$$

Let us list the main properties of annihilators:

1) If $\mathcal{I}$ and $\mathcal{J}$ are subsets of a vector space and $\mathcal{I} \subset \mathcal{J}$, then $\operatorname{Ann}(\mathcal{J}) \subset \operatorname{Ann}(\mathcal{I})$.

2) If $\mathcal{M}$ and $\mathcal{N}-$ are subspaces of a finite-dimensional vector space, then

$$\operatorname{Ann}(\mathcal{M} \cap \mathcal{N}) = \operatorname{Ann}(\mathcal{M}) + \operatorname{Ann}(\mathcal{N}), \ \operatorname{Ann}(\mathcal{M} + \mathcal{N}) = \operatorname{Ann}(\mathcal{M}) \cap \operatorname{Ann}(\mathcal{N}).$$

3) $\dim \mathcal{M} + \dim \operatorname{Ann}(\mathcal{M}) = \dim \mathcal{V}$.

3.2. Cyclic Invariant Subspaces

Consider an n-dimensional vector space $\mathcal{V}$ over some field $\mathbb{F}$ and a linear operator $\mathcal{A}$ on this space.

Definition 5 [21, p. 68]. A polynomial $f(t)$ is said to annihilate a linear operator $\mathcal{A}$ if $f(\mathcal{A}) = \mathcal{O}$. A normalized polynomial of minimal degree that annihilates $\mathcal{A}$ is called a minimal polynomial of $\mathcal{A}$.

Let us denote the minimal polynomial of the operator $\mathcal{A}$ as

$$\mu_{\mathcal{A}}(t) = t^m + \mu_m t^{m-1} + \cdots + \mu_2 t + \mu_1.$$

Let $x \in \mathcal{V}$ be a vector. Let $f(t)$ be a polynomial satisfying the relation

$$f(\mathcal{A})x = 0,$$

is called the annihilating polynomial of vector x, and the polynomial of the least degree that annihilates vector x is called the minimal annihilating polynomial of vector x.

If $f(t) = t^k + \alpha_k t^{k-1} + \cdots + \alpha_2 t + \alpha_1 -$ is the minimal annihilating polynomial of the vector x, then the linear subspace with basis

$$x, \mathcal{A}x, \ldots, \mathcal{A}^{k-1}x$$

is called a cyclic subspace with respect to the linear operator $\mathcal{A}$.

4. MAIN RESULTS

Let us fix some numerical field $\mathbb{F}$. We will investigate the partial stability of the system (1), (2) over the chosen field.

It is obvious that the y-stability property of the system (2) is invariant with respect to z-transformations of the form $z = S\overline{z}$, where $S \in \mathbb{F}^{p\times p}$ is a constant non-singular matrix, i.e. this property does not depend on the choice of basis in the space $\mathbb{F}^p$. Using the algebraic-geometric properties of the system under study, we will try to choose a basis in the phase subspace of variables $z_1, \ldots, z_p$ so as to obtain the desired stability conditions.

By carrying out a non-degenerate replacement of z-variables, we obtain a system of differential equations that is equivalent with respect to the property of partial stability:

$$\begin{aligned} \frac{dy(t)}{dt} &= Ay(t) + BS\overline{z}(t), \\ \frac{d\overline{z}(t)}{dt} &= S^{-1}Cy(t) + S^{-1}DS\overline{z}(t). \end{aligned} \tag{3}$$

In connection with the above, we consider the linear operator $\mathcal{D} : \mathbb{F}^p \to \mathbb{F}^p$ and the corresponding adjoint operator $\mathcal{D}^*$, acting in the adjoint space $\mathbb{F}^{p*}$ of linear functionals on the space $\mathbb{F}^p$. In this case, the following relation holds:

$$\langle \mathcal{D}s, \psi\rangle = \langle s, \mathcal{D}^*\psi\rangle = \psi\left(\mathcal{D}s\right) \ \forall s \in \mathbb{F}^p, \forall \psi \in \mathbb{F}^{p*}.$$

Let us introduce into consideration the linear functionals $b_i^* = b_i(s) \in \mathbb{F}^{p*}$, $i = 1, \ldots, m$, and construct invariant cyclic subspaces with respect to the linear operator $\mathcal{D}^*$ with generators b_i^*, $(i = 1, \ldots, m)$:

$$\begin{gathered} U_1^* = \operatorname{span}\left(b_1^*, \mathcal{D}^*b_1^*, \ldots, \mathcal{D}^{*k_1-1}b_1^*\right), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ U_i^* = \operatorname{span}\left(b_i^*, \mathcal{D}^*b_i^*, \ldots, \mathcal{D}^{*k_i-1}b_i^*\right), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ U_m^* = \operatorname{span}\left(b_m^*, \mathcal{D}^*b_m^*, \ldots, \mathcal{D}^{*k_m-1}b_m^*\right). \end{gathered}$$

Thus, the following relationships hold:

$$\mathcal{D}^*U_i^* \subseteq U_i^*, \quad i = 1, \ldots, m.$$

Proposition 2. *The linear subspaces $W_i^* = U_1^* + \cdots + U_i^*$, $i = 1, \ldots, m$, are invariant subspaces of the linear operator $\mathcal{D}^*$.*

Thus,

$$\mathcal{D}^*\left(U_1^* + \cdots + U_i^*\right) \subseteq U_1^* + \cdots + U_i^*, \ \forall i \in \{1, \ldots, m\}.$$

Obviously, there is a sequence of inclusions of subspaces:

$$\{0\} \subseteq U_1^* \subseteq U_1^* + U_2^* \subseteq \cdots \subseteq U_1^* + U_2^* + \cdots + U_m^* \subseteq \mathbb{F}^{p*}.$$

But taking into account invariance, the following chain of inclusions will also be valid:

$$\mathcal{D}^*\{0\} \subseteq \mathcal{D}^* U_1^* \subseteq \mathcal{D}^* \left(U_1^* + U_2^*\right) \subseteq \dots \subseteq \mathcal{D}^* \left(U_1^* + U_2^* + \dots + U_m^*\right) \subseteq \mathcal{D}^* \mathbb{F}^{p*}.$$

Using the properties of annihilators, we have:

$$\operatorname{Ann}(\{0\}) \supseteq \operatorname{Ann}(U_1^*) \supseteq \operatorname{Ann}(U_1^* + U_2^*) \supseteq \dots \supseteq \operatorname{Ann}(U_1^* + U_2^* + \dots + U_m^*) \supseteq \operatorname{Ann}(\mathbb{F}^{p*}).$$

The last chain of inclusions can be represented as

$$\mathbb{F}^p \supseteq \operatorname{Ann}(U_1^*) \supseteq \cap_{i=1}^2 \operatorname{Ann}(U_i^*) \supseteq \dots \supseteq \cap_{i=1}^m \operatorname{Ann}(U_i^*) \supseteq \{0\}.$$

Thus, the following inclusions are valid:

$$\mathcal{D}^* \left(U_1^* + \dots + U_i^*\right) \subseteq U_1^* + \dots + U_i^*, \ \forall i \in \{1, \dots, m\}.$$

Theorem 3. *The subspace* $\sum\limits_{i=1}^m U_i^*$ *is an invariant subspace of the operator* $\mathcal{D}^*$*,* $\dim\left(\sum\limits_{i=1}^m U_i^*\right) = k \Leftrightarrow \cap_{i=1}^m \operatorname{Ann}(U_i^*)$ *is an invariant subspace of the operator* $\mathcal{D}$*,* $\dim\left(\cap_{i=1}^m \operatorname{Ann}(U_i^*)\right) = p - k$.

Corollary 1. $\sum_{i=1}^m U_i^* = \mathbb{F}^{p*} \Leftrightarrow \cap_{i=1}^m \operatorname{Ann}(U_i^*) = \{0\}$.

Proposition 3. $\dim\left(\sum\limits_{i=1}^m U_i^*\right) = \operatorname{rang}(B, BD, \dots, BD^{p-1})$.

Theorem 4 (reducibility theorem). *If* $\dim\left(\cap_{i=1}^m \operatorname{Ann}(U_i^*)\right) = k < p$*, then there exists a nondegenerate transformation* $z = S\overline{z}$ *that reduces the system* (2) *to the form*

$$\begin{aligned}
\frac{dy(t)}{dt} &= Ay(t) + B_{11}\overline{z}_1(t), \\
\frac{d\overline{z}_1(t)}{dt} &= \overline{C}_1 y(t) + \overline{D}_{11}\overline{z}_1(t), \\
\frac{d\overline{z}_2(t)}{dt} &= \overline{C}_2 y(t) + \overline{D}_{21}\overline{z}_1(t) + \overline{D}_{22}\overline{z}_2(t),
\end{aligned} \tag{4}$$

where $\overline{z}_1 \in \mathbb{F}^k$*,* $\overline{z}_2 \in \mathbb{F}^{p-k}$.

Theorem 5 (dual reducibility theorem). *If* $\dim\left(\sum\limits_{i=1}^m U_i^*\right) = p - k > 0$*, then there exists a nondegenerate transformation* $\overline{z} = Wz$ *that reduces the system* (2) *to the form* (4).

Theorem 6 (on partial stability). *Let the conditions of Theorem 4 be satisfied. For system* (1) *to be stable with respect to the first* m *components of the phase vector* x*, it is necessary and sufficient that the system*

$$\begin{aligned}
\frac{dy(t)}{dt} &= Ay(t) + \overline{B}_{11}\overline{z}_1(t), \\
\frac{d\overline{z}_1(t)}{dt} &= \overline{C}_1 y(t) + \overline{D}_{11}\overline{z}_1(t)
\end{aligned} \tag{5}$$

be stable with respect to all variables.

In some special cases, the study of partial stability of the system (1) can be simplified.

Proposition 4. *If, in addition to the conditions of Theorem 5, the condition*

$$\operatorname{im} C \subseteq \operatorname{Ann}\left(\sum_{i=1}^m U_i^*\right),$$

then the system (2) *is reducible to the form*

$$\begin{aligned} \frac{dy(t)}{dt} &= Ay(t) + \overline{B}_{11}\overline{z}_1(t), \\ \frac{d\overline{z}_1(t)}{dt} &= \overline{D}_{11}\overline{z}_1(t), \\ \frac{d\overline{z}_2(t)}{dt} &= \overline{C}_2 y(t) + \overline{D}_{21}\overline{z}_1(t) + \overline{D}_{22}\overline{z}_2(t), \end{aligned} \tag{6}$$

and the question of y-stability is reduced to the study of the stability of the system

$$\begin{aligned} \frac{dy(t)}{dt} &= Ay(t) + \overline{B}_{11}\overline{z}_1(t), \\ \frac{d\overline{z}_1(t)}{dt} &= \overline{D}_{11}\overline{z}_1(t). \end{aligned} \tag{7}$$

Proofs of the theorems and propositions are given in the Appendix.

Remark 1. If the conditions of proposition 4 are met, the problem of studying partial stability is reduced to studying the stability of two lower-order systems. The subspace $\mathrm{Ann}(U^*)$ is an invariant subspace with respect to y-stability in the sense that the y-stability property is preserved for any choice of the matrix C, whose columns are arbitrary vectors from the subspace $\mathrm{Ann}\left(\sum\limits_{i=1}^{m} U_i^*\right)$. Generally speaking, small variations in the parameters of the system (2) can lead to a violation of the partial stability property.

Remark 2. Note that the proposed approach allows us to investigate the stability of any component of the phase vector of the system under study. It is sufficient to designate the component under study in the system as y, and the remaining components will form the vector z. This implements the algorithm for solving the variable elimination problem proposed in [22].

5. EXAMPLE

Let the system under study have the following initial data:

$$y \in \mathbb{R}^2, \quad z \in \mathbb{R}^4, \quad A \in \mathbb{R}^{2\times 2}, \quad B \in \mathbb{R}^{2\times 4}, \quad C \in \mathbb{R}^{4\times 2}, \quad D \in \mathbb{R}^{4\times 4}, \quad n = 6, \quad p = 4,$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 1 & 2 & -2 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 0 \\ 1 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 9 & -23 & -46 & 8 \\ -3 & 5 & 22 & 16 \\ -6 & -2 & -10 & -4 \\ -15 & 19 & 38 & -10 \end{pmatrix}.$$

Based on the data of the problem, we consider the linear space $\mathbb{R}^4$ and the corresponding dual linear space $\mathbb{R}^{4*}$, in which the linear operators $\mathcal{D}$ and $\mathcal{D}^*$ act, respectively.

Covectors $b_1^* = -\xi_1 + \xi_2 + 2\xi_3 - 2\xi_4$ and $b_2^* = \xi_1 + \xi_2 + 2\xi_3$ are generating vectors of invariant cyclic subspaces with respect to the linear operator $\mathcal{D}^*$.

The corresponding subspaces — the linear spans of their bases — have the form:

$$\begin{aligned} U_1^* &= \mathrm{span}(-\xi_1 + \xi_2 + 2\xi_3 - 2\xi_4, 6\xi_1 - 14\xi_2 - 28\xi_3 + 20\xi_4), \\ U_2^* &= \mathrm{span}(\xi_1 + \xi_2 + \xi_3, -20\xi_2 - 34\xi_3 + 20\xi_4, -36\xi_1 + 348\xi_2 + 660\xi_3 - 384\xi_4). \end{aligned}$$

The minimal annihilating polynomials of the covectors b_1^* and b_2^* have the form

$$\mu_1(t) = t^2 + 24t + 108, \quad \mu_2(t) = t^3 + 30t^2 + 252t + 648.$$

At the same time $\dim U_1^* = 2$, $\dim U_2^* = 3$, $\dim(U_1^* + U_2^*) = 3 < 4 = \dim \mathbb{R}^{4*}$.

$$\operatorname{Ann}(U_1^* + U_2^*) : \begin{cases} -\xi_1 + \xi_2 + 2\xi_3 - 2\xi_4 = 0, \\ 6\xi_1 - 14\xi_2 - 28\xi_3 + 20\xi_4 = 0, \\ \xi_1 + \xi_2 + \xi_3 = 0, \\ -20\xi_2 - 34\xi_3 + 20\xi_4 = 0, \\ -36\xi_1 + \xi_2 + 660\xi_3 - 384\xi_4 = 0. \end{cases}$$

By solving a homogeneous system of linear equations, we find the basis of the solution space:

$$\operatorname{Ann}(U_1^* + U_2^*) = \operatorname{span}(-1, 1, 0, 1).$$

By choosing a basis of the subspace $U_1^* + U_2^*$, partially consistent with the bases of the subspaces U_1^* and U_2^*,

$$e^1 = -\xi_1 + \xi_2 + 2\xi_3 - 2\xi_4, \quad e^2 = 6\xi_1 - 14\xi_2 - 28\xi_3 + 20\xi_4, \quad e^3 = \xi_1 + \xi_2 + \xi_3$$

and complementing it to the basis of the entire space $\mathbb{R}^{4*}$ with the covector $e^4 = -\xi_1 + \xi_2 + \xi_4$, we obtain the corresponding dual basis of the space $\mathbb{R}^4$:

$$e_1 = (1/4, 9/4, -5/2, -2), \quad e_2 = (1/24, 5/24, -1/4, -1/6),$$
$$e_3 = (2/3, 4/3, -1, -2/3), \quad e_4 = (-1/3, 1/3, 0, 1/3).$$

Then the transition matrix S from the original basis to the new one has the form

$$S = \begin{pmatrix} 1/4 & 1/24 & 2/3 & -1/3 \\ 9/4 & 5/24 & 4/3 & 1/3 \\ -5/2 & -1/4 & -1 & 0 \\ -2 & -1/6 & -2/3 & 1/3 \end{pmatrix}.$$

The matrix of the operator $\mathcal{D}$ in the new basis takes the form

$$D_e = S^{-1} D S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -108 & -24 & 0 & 0 \\ 0 & 1 & -6 & 0 \\ -162 & -35/2 & -60 & 24 \end{pmatrix}.$$

At the same time

$$BS = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad S^{-1} C = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 3 & -3 \end{pmatrix}.$$

As a result of the transition to a new basis using the non-degenerate transformation $z = S\overline{z}$, we arrive at the system:

$$\frac{dy(t)}{dt} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} y(t) + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \overline{z}(t),$$

$$\frac{d\overline{z}(t)}{dt} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 3 & -3 \end{pmatrix} y(t) + \begin{pmatrix} 0 & 1 & 0 & 0 \\ -108 & -24 & 0 & 0 \\ 0 & 1 & -6 & 0 \\ -162 & -\frac{35}{2} & -60 & 24 \end{pmatrix} \overline{z}(t)$$

where $y \in \mathbb{R}^2$, $z \in \mathbb{R}^4$, $A \in \mathbb{R}^{2\times 2}$, $B \in \mathbb{R}^{2\times 4}$, $C \in \mathbb{R}^{4\times 2}$, $D \in \mathbb{R}^{4\times 4}$, $n = 6$, $p = 4$.

Thus, the original problem is divided into two simpler subtasks: studying the stability of two systems

$$\frac{dy(t)}{dt} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} y(t) \quad \text{and} \quad \frac{d\overline{z}_1(t)}{dt} = \begin{pmatrix} 0 & 1 & 0 \\ -108 & -24 & 0 \\ 0 & 1 & -6 \end{pmatrix} \overline{z}_1(t).$$

In this case, for the second system the characteristic polynomial has the form

$$\chi(t) = (-6 - t)\mu_1(t) = -\mu_2(t),$$

which is stable. Thus, for y-stability, the stability of the first system is necessary and sufficient.

6. CONCLUSION

An algebraic-geometric approach to the study of partial stability of linear systems with constant coefficients is proposed, which makes it possible to determine the stability conditions of the y-components of the phase vector by "rotating" the phase subspace of the z-components. By using the geometric properties of this subspace, conditions are obtained under which stability analysis is reduced to solving two problems of lower dimension. It is shown that the system exhibits some invariance properties with respect to partial stability. The proposed approach can be applied to studying the stability of linear discrete and delayed systems, as well as to solving polystability problems.

Further research directions can be carried out in the areas of partial controllability and stabilization of linear systems, as well as in the area of dissemination of this approach to the study of problems of partial stability of linear non-stationary systems and nonlinear systems of differential equations.

APPENDIX

Proof of Proposition 2. Let $\forall w_i^* = u_1^* + \cdots + u_i^* \in W_i^* = U_1^* + \cdots + U_i^*$. Since $\mathcal{D}^* u_i \in U_i$ $(\forall i = 1, \ldots, m)$ due to the invariance of subspaces U_i^*, then $\mathcal{D}^* w_i = \mathcal{D}^* (u_1 + \cdots + u_i) = \mathcal{D}^* u_1 + \cdots + \mathcal{D}^* u_i \in U_1^* + \cdots + U_i^* = W_i^*$ $(\forall i = 1, \ldots, m)$. Thus, W_i $(\forall i = 1, \ldots, m)$ are invariant subspaces of the operator $\mathcal{D}^*$. Proposition 2 has been proven.

Proof of Proposition 3. The proof follows from the isomorphism of the two dual linear spaces L and L^*, as well as from the properties of annihilators. Indeed, since the relation

$$\operatorname{Ann}\left(\sum_{i=1}^{m} U_i^*\right) = \cap_{i=1}^{m} \operatorname{Ann}\left(U_i^*\right),$$

then in the original dual bases we represent the annihilator in the form

$$\operatorname{Ann}\left(\sum_{i=1}^{m} U_i^*\right) : \begin{cases} B\xi &= 0, \\ BD\xi &= 0, \\ \cdots & \\ BD^{p-1}\xi &= 0. \end{cases}$$

But then

$$\dim \operatorname{Ann}\left(\sum_{i=1}^{m} U_i^*\right) = \dim \mathbb{F}^{p*} - \operatorname{rang}(B, BD, \ldots, BD^{p-1}).$$

On the other hand, the following relation is valid:

$$\dim \sum_{i=1}^{m} U_i^* + \dim \operatorname{Ann}\left(\sum_{i=1}^{m} U_i^*\right) = \dim \mathbb{F}^{p*},$$

Taking into account the last two relations, we have

$$\dim \sum_{i=1}^{m} U_i^* = \operatorname{rang}(B, BD, \ldots, BD^{p-1}).$$

Proposition 3 has been proven.

Proof of Theorem 3. Let us prove the necessity. From Proposition 2 it follows that $W^* = \sum\limits_{i=1}^{m} U_i^*$ is an invariant subspace with respect to the operator $\mathcal{D}^*$. Let $\forall w^* \in W^*,\ \forall w \in \operatorname{Ann}(W^*) \Leftrightarrow \langle w, w^* \rangle = 0 \Leftrightarrow \langle \mathcal{D}w, w^* \rangle = \langle w, \mathcal{D}^* w^* \rangle = 0$, since $\mathcal{D}^* w^* \in W^*$. But then $\mathcal{D}w \in \operatorname{Ann}(W^*)$. Since $\dim(W^*) + \dim(\operatorname{Ann}(W^*)) = k + \dim(\operatorname{Ann}(W^*)) = p$, then $\dim(\operatorname{Ann}(W^*)) = p - k$. Finally, since $\operatorname{Ann}(W^*) = \operatorname{Ann}\left(\sum_{i=1}^{m} U_i^*\right) = \cap_{i=1}^{m} \operatorname{Ann}\left(U_i^*\right)$, then necessity is proved.

The proof of sufficiency is obtained by verbatim repetition of necessity, provided that the proof proceeds in the opposite direction, i.e., from the end of the proof to its beginning. Theorem 3 is proven.

Proof of Theorem 4. Let $U^* = \sum\limits_{i=1}^{m} U_i^*$, $\dim U^* = p - k > 0$. Thus $U^* = \operatorname{span}\left(e^1, \ldots, e^{p-k}\right)$ and $e^1, \ldots, e^{p-k}$ is a basis of this subspace. We complement this system of vectors to a basis of the space $\mathbb{F}^{p*}$ using the vectors $e^{p-k+1}, \ldots, e^p$. This basis in the space $\mathbb{F}^p$ corresponds to the dual basis $e_1, \ldots, e_p$. Since by Proposition 2 it follows that $\sum\limits_{i=1}^{m} U_i^*$ is an invariant subspace with respect to the operator $\mathcal{D}^*$, we have $\forall u^* \in U^* \Rightarrow \mathcal{D}^* u^* \in U^* \Rightarrow \langle \mathcal{D}v, u^* \rangle = \langle v, \mathcal{D}^* u^* \rangle = u\left(\mathcal{D}v\right)$ $\forall u \in \mathbb{F}^p$, $\forall u^* \in \mathbb{F}^{p*}$. Then

$$\langle \mathcal{D}e_j, e^i \rangle = \langle e_j, \mathcal{D}^* e^i \rangle = e^i\left(\mathcal{D}e_j\right) = 0 \ \ \forall i = 1, \ldots, p-k,\ j = p-k+1, \ldots, p.$$

But this means that the linear operator $\mathcal{D}$ has an invariant subspace $\operatorname{Ann}(U^*) \subseteq \mathbb{F}^p$. Therefore, in the chosen basis, the matrix of the linear operator $\mathcal{D}$ takes the form

$$D_e = \begin{pmatrix} \overline{D}_{11} & 0 \\ \overline{D}_{21} & \overline{D}_{22} \end{pmatrix}.$$

It remains to construct the matrix of the non-degenerate transformation. Since the annihilators of the subspaces U_i^* are representable as a set of homogeneous linear systems

$$\operatorname{Ann}(U_i^*)\ (i = 1, \ldots, m) : \left\{ \begin{array}{rcl} b_i\xi & = & 0, \\ b_i D\xi & = & 0, \\ \cdots & & \\ b_i D^{k_i-1}\xi & = & 0, \end{array} \right.$$

then the subspace $\cap_{i=1}^{m} \operatorname{Ann}\left(U_i^*\right)$ can be represented as

$$\left\{ \begin{array}{rcl} B\xi & = & 0, \\ BD\xi & = & 0, \\ \cdots & & \\ BD^{p-1}\xi & = & 0. \end{array} \right. \tag{A.1}$$

Having found the fundamental system of solutions of the homogeneous system (A.1) $\xi_1, \ldots, \xi_{p-k}$ and complementing it to the basis of the space $\mathbb{F}^p$ with the system of vectors $\nu_1, \ldots, \nu_k$, we obtain the desired transformation matrix:

$$S = (\nu_1, \cdots, \nu_k, \xi_1, \cdots, \xi_{p-k}).$$

Let's make a change of variables $z = S\overline{z}$ in system (2). System (3) then takes the form (4). Theorem 4 is proven.

Proof of Theorem 5. Let $U^* = \sum\limits_{i=1}^{m} U_i^*$, $\dim U^* = p - k > 0$. Thus, $U^* = \operatorname{span}\left(e^1, \ldots, e^{p-k}\right)$ and $e^1, \ldots, e^{p-k}-$ are the basis of this subspace. Let us supplement this system of vectors to the basis of the space $\mathbb{F}^{p*}$ using the vectors $e^{p-k+1}, \ldots, e^p$. Then, according to Proposition 2, it follows that U^* is an invariant subspace with respect to the operator $\mathcal{D}^*$. Therefore, the matrix of the linear operator $\mathcal{D}^*$ in the chosen basis has the form

$$D_e^* = W^{-1} D^* W = \begin{pmatrix} \overline{D^*}_{11} & \overline{D^*}_{12} \\ 0 & \overline{D^*}_{22} \end{pmatrix}.$$

But then the adjoint operator $\mathcal{D} = (\mathcal{D}^*)^*$ in the dual basis takes the form

$$D = (D_e^*)^T = (W^{-1} D^* W)^T = W^T (D^*)^T W^{-T} = \begin{pmatrix} \overline{D}_{11} & 0 \\ \overline{D}_{21} & \overline{D}_{22} \end{pmatrix}, \ \ \overline{D}_{ij} = \overline{D^*}_{ij}^T, \ i, j \in \{1, 2\}.$$

From here $S = W^{-T} \Rightarrow W = S^{-T}$. The required non-degenerate transformation is equal to $\overline{z} = Wz$. By replacing $z = W^{-1}\overline{z}$ in system (2), we arrive at system (4). Theorem 5 is proven.

Proof of Theorem 6. According to Theorem 4, there exists a transformation that reduces system (2) to form (4). Moreover, the transformation found preserves the y-stability property. As a result, a subsystem of $(m + p - k)$th order is identified. According to Corollary 1, the resulting subsystem satisfies the relation $\sum\limits_{i=1}^{m} \overline{U}_i^* = \mathbb{F}^{p-k^*}$. Thus, $\operatorname{rang}(\overline{B}_{11}, \overline{B}_{11}\overline{D}_{11}, \ldots, \overline{B}_{11}\overline{D}_{11}^{p-k-1}) = p - k$. Since the order of the resulting subsystem 5 is $m + p - k$, by Corollary 1.1.2 [4, p. 38] we see that this condition is both necessary and sufficient for y-stability of the system (2). Theorem 6 is proved.

Proof of Proposition 4. Let the conditions of theorem 6 be satisfied and, in addition, $\operatorname{im} C \subseteq \operatorname{Ann}\left(\sum\limits_{i=1}^{m} U_i^*\right)$. Then

$$\begin{cases} BC &= 0, \\ BDC &= 0, \\ \cdots & \\ BD^{p-1}C &= 0. \end{cases} \tag{A.2}$$

As a result, in system (5), the matrix $\overline{C}_1 = 0$. Proposition 4 is proven.

REFERENCES

1. Vorotnikov, V.I. and Rumyantsev, V.V., *Ustoichivost' i upravlenie po chasti koordinat fazovogo vektora dinamicheskikh sistem: teoriya, metody i prilozhemiya* (Stability and Control of Dynamic systems by Part of Phase Vector Coordinates: Theory, Methods, and Applications), Moscow: Nauchnyi Mir, 2001.

2. Vorotnikov, V.I. and Prokop'ev, V.P., Partial Stability of Motion of Linear Systems, *Prikl. Mat. Mekh.*, 1978, vol. 42, no. 2, pp. 268–271.

3. Vorotnikov, V.I., Partial Stability and Control: The State-of-the-Art and Development Prospects, *Autom. Remote Control*, 2005, vol. 66, no. 24, pp. 511–561.

4. Vorotnikov, V.I., *Ustoichivost' dinamicheskikh sistem po otnosheniyu k chasti peremennykh* (Partial Stability of Dynamic Systems), Moscow: Nauka, 1991.

5. Zubov, V.I., *Matematicheskie metody issledovaniya sistem avtomaticheskogo regulirovaniya* (Mathematical Methods of Studying the Automatic Control Systems), Leningrad: Sudpromgiz, 1959.

6. Martynyuk, A.A. and Chernetskaya, L.N., Semistability of Linear Autonomous Systems, *Dokl. Akad. Nauk Ukrainy*, 1993, no. 8, pp. 17–19.

7. Oziraner, A.S., Investigation of Partial Stability by Quadratic Forms, *Prikl. Mat. Mekh.*, 1986, vol. 50, no. 1, pp. 163–167.

8. Rumyantsev, V.V. and Oziraner, A.S., *Ustoichivost' i stabilizatsiya dvizheniya po otnosheniyu k chasti peremennykh* (Partial Stability and Stabilization of Motion), Moscow: Nauka, 1987.

9. Chen, Y., Qu, X., Dai, G., and Aleksandrov, A.Yu., Linear-Transformation-Based Analysis and Design of State Consensus for Multi-Agent Systems with State Observers, *J. Frankl. Inst.*, 2015, vol. 352, no. 9, pp. 3447–3457.

10. Chen, Y., Qu, X., Aleksandrov, A.Y., and Dai, G., Consensus of Heterogeneous Linear Multi-Agent Systems: Linear-Transformation-Based Partial Stability Approach, *Transactions of the Institute of Measurement and Control.*, 2017, vol. 39, no. 11., pp. 1623–1630.

11. Chen, Y., Xu, G., and Zhan, J., Leader-Following Consensus of Heterogeneous Linear Multi-Agent Systems: New Results Based on Linear Transformation Method, *Transactions of the Institute of Measurement and Control.*, 2021, vol. 44(7), no. 7, pp. 1473–1483.

12. Abramov, S.A. and Bronstein, M., Solving Linear Systems of Differential and Difference Equations with Respect to a Part of The Unknowns, *Comput. Math. and Math. Phys.*, 2006, vol. 46(2), pp. 218–230.

13. Panferov, A.A., Differential Equation Systems with Selected Part Of The Unknowns, *Programming and Computer Software*, 2015, vol. 41, no. 2, pp. 90–97.

14. Panferov, A.A., Satellite Unknowns in Irreducible Differential Systems, *Programming and Computer Software*, 2018, vol. 44, no. 2, pp. 105–111.

15. Chudinov, K.M., Criterion of Stability with Respect to a Part of Variables of an Autonomous Systems, *Russ Math.*, 2003, vol. 47, pp. 64–69.

16. Chudinov, K.M., The Geometric Nature of Partial and Conditional Stability, *Russ Math.*, 2008, vol. 52, pp. 69–77.

17. Nikonov, V.I., Geometric Aspect of Partial Stability for Linear Systems, *Zhurnal Srednevolzhskogo matematicheskogo obshchestva*, 2011, vol. 13, no. 2, pp. 95–99.

18. Nikonov, V.I., Partial Stability of Linear Systems with Respect to a Given Component of The Phase Vector, *Zhurnal Srednevolzhskogo matematicheskogo obshchestva*, 2021, vol. 23, no. 1, pp. 43–57.

19. Postnikov, M.M., *Lektsii po geometrii. Semestr II. Lineynaya algebra* (Lectures on Geometry. Halfyear II. Linear Algebra), Moscow: Nauka, 1986.

20. Halmos, P.R., *Finite-Dimensional Vector Spaces*, New York: Springer-Verlag, 1974. Translated under the title *Konechnomernye vektornye prostranstva*, Moscow: Fizmatgiz,1963.

21. Kostrikin, A.I., *Vvedenie v algebru. Ch. 2* (Introduction of Algebra. Part 2), Moscow: Fiziko-Matematicheskaya Literatura, 2000.

22. Pontryagin, L.S., *Obyknovennye differentsial'nye uravneniya*(Ordinary Differential Equations), Moscow: Nauka, 1974.

This paper was recommended for publication by L.B. Rapoport, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 1, pp. 29–43.
© The Author(s), 2026 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2026.
Russian Text © The Author(s), 2026, published in Avtomatika i Telemekhanika, 2026, No. 1, pp. 42–61.

Boundary Control of a Rod Heating Process Using Current and Preceding Time Feedback

V. M. Abdullayev[*,**,***,a] **and V. A. Hashimov**[**,b]

[*]*Azerbaijan State Oil and Industry University, Baku, Azerbaijan*
[**]*Institute of Mathematics, Baku, Azerbaijan*
[***]*Azerbaijan University of Architecture and Construction, Baku, Azerbaijan*
e-mail: [a]*vaqif_ab@rambler.ru,* [b]*vugarhashimov@gmail.com*

Received March 20, 2025
Revised October 25, 2025
Accepted October 27, 2025

Abstract—This paper considers the problem of optimal boundary control with feedback for a rod heating process. Devices to measure the current temperature values are placed on the rod. To design the boundary control, a linear dependence of its value on the measured temperature values, both at the current and previous time instants, is proposed. The constant coefficients in this dependence are the feedback parameters under optimization. As a result, the original feedback control design problem is formulated as a parametric optimal control problem for a distributed parameter plant. The optimal values of the above parameters are determined via formulas derived for the components of the gradient of the objective functional. With these formulas, it is possible to use efficient first-order finite-dimensional optimization methods for parameter determination. Numerical results of computer experiments are presented and analyzed.

Keywords: boundary control, rod heating, feedback, measurement point, feedback parameters, the gradient of a functional

DOI: 10.7868/S1608303226010036

1. INTRODUCTION

In this paper, we investigate the problem of optimal control of a boundary value condition for a rod heating process. Feedback is implemented through temperature measurements at internal points of the rod, and the measured values are used to form the current value of the boundary control function. The peculiarity of the feedback considered here is that the dependence of the boundary control value includes the temperature values at the measurement points, both at the current and previous time instants. In this case, the boundary control design problem reduces to determining the optimal values of the coefficients in the dependence of the boundary control value on the temperature values at the measurement points.

Note that the number of publications devoted to the analysis and, especially, to the numerical solution of optimal control problems for plants described by partial differential equations (PDEs) [1–4] is substantially inferior to that of the publications related to control design problems for lumped parameter systems [5–9]. This is due to, first, the complexity of analyzing and numerically solving boundary value problems described by systems of PDEs and, second, the complexity of implementing control systems for distributed parameter plants [10–12]. With the development of measurement tools, remote control means, computer technologies, and methods of computational mathematics in recent decades, it has become possible to design and implement real-time feedback control and regulation systems for complex facilities and industrial processes, both with lumped and distributed parameters [13–18].

Note that the feedback control approach proposed below for the rod heating process using the history of previous measurements is, of course, more demanding in terms of implementation than the one using only the current measurement values. However, considering the state-of-the-art of technology and engineering, the implementation of such systems has become quite feasible.

In what follows, we derive explicit formulas for the components of the gradient of the objective functional with respect to the feedback parameters. With these formulas, efficient numerical methods and software packages for first-order finite-dimensional optimization [5, 19] can be used to determine the optimal parameter values. The results of computer experiments for test boundary control problems with feedback are provided and analyzed.

2. PROBLEM STATEMENT

Consider the problem of optimal boundary control (regulation) of a heating process for a rod heated from one end. The temperature at the points of the rod is described by a parabolic differential equation of the form

$$\frac{\partial u(x,t)}{\partial t} = a^2 \frac{\partial^2 u(x,t)}{\partial x^2} + \mu_1 \left[\theta - u(x,t)\right], \quad (x,t) \in \Omega = (0,l) \times (0,t_f], \tag{2.1}$$

with an initial condition

$$u(x,t) = \gamma(x;\varphi), \quad \varphi = \text{const} \in \mathbb{R}^m, \quad x \in (0,l), \quad t \le 0, \tag{2.2}$$

and boundary value conditions

$$\left.\frac{\partial u(x,t)}{\partial x}\right|_{x=0} = -\mu_2(\vartheta(t) - u(0,t)), \quad 0 \le t \le t_f, \tag{2.3}$$

$$\left.\frac{\partial u(x,t)}{\partial x}\right|_{x=l} = \mu_2(\theta - u(l,t)), \quad 0 \le t \le t_f. \tag{2.4}$$

Here, $u(x,t)$ is a function continuously differentiable with respect to $t \in (0,t_f]$ and twice continuously differentiable with respect to $x \in (0,l)$, which defines the temperature of the rod at a time instant $t \in [0,t_f]$ at a point $x \in [0,l]$; l is the length of the rod; t_f is the terminal time of the heating process; θ is an imprecisely known constant temperature of the environment defined by the set Θ of possible values of θ and a (probability) density function $\rho_\Theta(\theta)$ such that

$$\rho_\Theta(\theta) \ge 0, \quad \int\limits_\Theta \rho_\Theta(\theta) d\theta = 1;$$

finally, μ_1 and μ_2 are given heat transfer coefficients between the rod and the environment.

According to condition (2.2), the temperature at the points of the rod before the heating process is defined by a known parametrically given function $\gamma(x;\varphi)$. It depends on an imprecisely known constant m-dimensional vector of parameters φ defined by the set Φ of possible values of φ and a (probability) density function $\rho_\Phi(\varphi)$ such that

$$\rho_\Phi(\varphi) \ge 0, \quad \int\limits_\Phi \rho_\Phi(\varphi) d\varphi = 1. \tag{2.5}$$

In particular, the initial temperature may be constant along the length of the rod, i.e., $\gamma(x;\varphi) = \varphi \in \Phi \in R$.

The sets Θ and/or Φ may be finite sets

$$\Theta = \{\theta_i : i = 1, \ldots, N_\theta\}, \quad \Phi = \{\varphi_j : j = 1, \ldots, N_\varphi\}$$

with given probability values for their elements, i.e.,

$$p_i^\theta = p(\theta = \theta_i), \quad p_j^\varphi = p(\varphi = \varphi_j), \quad i = 1, \ldots, N_\theta, \quad j = 1, \ldots, N_\varphi,$$
$$\sum_{i=1}^{N_\theta} p_i^\theta = 1, \quad \sum_{j=1}^{N_\varphi} p_j^\varphi = 1. \tag{2.6}$$

The piecewise continuous function $\vartheta(t)$, $t \in [0, t_f]$, is the control action optimized for the rod heating process under study. It defines the boundary value condition (2.4) at the left end of the rod and is called the boundary control function [5]. Assume that, based on technical or technological considerations, it satisfies the constraint

$$\underline{\vartheta} \le \vartheta(t) \le \bar{\vartheta}, \quad t \in [0, t_f], \tag{2.7}$$

where $\underline{\vartheta}$ and $\bar{\vartheta}$ are given values.

The control problem for this heating process is to find a control function $\vartheta(t)$ that affects the temperature of the rod at the left end according to (2.3), satisfies condition (2.7), and minimizes the objective functional

$$J(\vartheta) = \int\limits_\Phi \int\limits_\Theta I(\vartheta; \varphi, \theta) \rho_\Phi(\varphi) d\theta d\varphi,$$
$$I(\vartheta; \varphi, \theta) = \int\limits_0^l \mu(x) \left(u\left(x, t_f; \vartheta, \varphi, \theta\right) - U(x)\right)^2 dx + \sigma \int\limits_0^{t_f} \vartheta^2(t) dt. \tag{2.8}$$

Here, $u(x,t) = u(x,t;\vartheta,\varphi,\theta)$ is a function continuously differentiable with respect to $t \in [0, t_f]$ and twice continuously differentiable with respect to $x \in (0, l)$ that represents the solution of the initial-boundary value problem (2.1)–(2.4) under arbitrarily given admissible control function $\vartheta(t)$, parameters φ of the initial function $\gamma(x;\varphi)$, and environment temperature θ; $U(x)$, $x \in (0,l)$, is a piecewise continuous function defining the desired temperature distribution on the rod at the end of the heating process; $\mu(x) \ge 0$, $x \in (0,l)$, is a given continuous weight function; finally, $\sigma > 0$ is the regularization parameter of the objective functional.

A specific feature of this setting is that the objective functional (2.8) assesses the control function $\vartheta(t)$ not by the behavior of a single phase trajectory $u(x,t)$ under any given admissible parameters φ of the initial function $\gamma(x;\varphi)$ and environment temperature θ, but in the whole for a pencil of phase trajectories. The pencil contains the phase trajectories obtained by solving the boundary value problem (2.1)–(2.5) under all admissible parameters φ and θ. Thus, in the optimal control problem under consideration, we seek for a boundary control function $\vartheta(t)$ minimizing the functional (2.8) on average over the entire set Φ of the parameters φ of the initial function $\gamma(x;\varphi)$ and the value set Θ of the environment temperature θ.

Let the heating process be controlled via a state feedback law. More precisely put, at some L points $\xi_i \in (0,l)$, $i = 1, \ldots, L$, the temperature values are measured, either in continuous time:

$$u_i(t) = u(\xi_i, t), \quad i = 1, \ldots, L, \quad t \in (0, t_f],$$

or at given discrete time instants $\bar{t}_j \in (0, t_f]$, $j = 1, \ldots, M$:

$$u_{ij} = u(\xi_i, \bar{t}_j), \quad i = 1, \ldots, L, \quad j = 1, \ldots, M. \tag{2.9}$$

The measured temperature values are employed to form the current value of the control function. In the case of continuous measurements, the following dependence is used:

$$\vartheta(t) = \sum_{i=1}^{L} \left[k_{1i} \left(u(\xi_i, t) - U_i\right) + k_{2i} \left(u(\xi_i, t-\tau) - U_i\right)\right], \quad t \in [0, t_f], \tag{2.10}$$
$$U_i = U(\xi_i), \quad i = 1, \ldots, L.$$

Here, $U_i = U(\xi_i)$, $i = 1, \ldots, L$, and $\tau > 0$ is a given value determined experimentally depending on the heating process dynamics. That is, under strong dynamics, the value τ should be small; under weak dynamics, the value τ should be chosen large enough so that the values $u(\xi_i, t)$ and $u(\xi_i, t-\tau)$ differ sufficiently.

The optimized constant coefficients k_{1i} and k_{2i}, $i = 1, \ldots, L$, will be called the feedback parameters.

In the case of the discrete feedback law (2.10), the control values $\vartheta(t)$ will be formed using the dependence

$$\vartheta(t) = \vartheta_{ij} = \sum_{i=1}^{L} [k_{1i}(u(\xi_i, \bar{t}_j) - U_i) + k_{2i}(u(\xi_i, \bar{t}_{j-1}) - U_i)], \quad t \in [\bar{t}_j, \bar{t}_{j+1}). \tag{2.11}$$

Clearly, the boundary control is, in general, a piecewise constant function with discontinuities at the discrete time instants of measurements.

Substituting the dependence (2.10) into boundary value condition (2.3) yields

$$\frac{\partial u(0,t)}{\partial x} = -\mu_2 \left(\sum_{i=1}^{L} [k_{1i}(u(\xi_i, t) - U_i) + k_{2i}(u(\xi_i, t-\tau) - U_i)] - u(0,t) \right), \tag{2.12}$$

where $t \in [0, t_f]$.

The new boundary value condition (2.12) has two features. First, it contains a time delay; second, it represents a nonlocal boundary value condition connecting the behavior of the unknown function on the boundary of the rod with its values inside the rod at the current and preceding time instants [20–23].

In the discrete feedback case, the boundary value condition (2.3) takes the form

$$\frac{\partial u(0,t)}{\partial x} = -\mu_2 \left(\sum_{i=1}^{L} \left[k_{1i} \left(u_{ij} - U_i\right) + k_{2i} \left(u_{ij-1} - U_i\right)\right] - u(0,t) \right), \quad t \in [t_{j-1}, t_j], \quad j = 1, \ldots, M.$$

Now we analyze the constraint (2.7) on the control function $\vartheta(t)$ taking the dependence (2.10) into account. Based on technological considerations, assume that the temperature at the points of the rod under all possible admissible heating modes, parameters of the initial conditions, and environment temperature values satisfies the condition

$$\underline{u} \le u(x, t; \vartheta, \varphi, \theta) \le \bar{u}, \quad \varphi \in \Phi, \quad \theta \in \Theta, \quad x \in (0, l), \quad t \in [0, t_f]. \tag{2.13}$$

Here, $\underline{u}$ and $\bar{u}$ are given values defining the range of possible temperature values at the points of the rod. In view of (2.7), from (2.10) we obtain the following conditions on the parameters of the continuous feedback law:

$$\underline{\vartheta} \le \sum_{i=1}^{L} \left[k_{1i} \left(u(\xi_i, t) - U_i\right) + k_{2i} \left(u(\xi_i, t-\tau) - U_i\right)\right] \le \bar{\vartheta}. \tag{2.14}$$

Let us choose τ sufficiently small so that for each $t \in [0, t_f]$, the values of $u_i(t) = u(\xi_i, t)$ and $u_i(t-\tau) = u(\xi_i, t-\tau)$ are close to each other and cannot, from practical considerations, lie at different end points of the interval $[\underline{u}, \bar{u}]$. Then from (2.14) we have

$$\underline{\vartheta} \leq \sum_{i=1}^{L} \left[(k_{1i} + k_{2i})\, u_i(t) - (k_{1i} + k_{2i})\, U_i \right] \leq \overline{\vartheta}, \quad t \in (0, t_f]. \tag{2.15}$$

In addition, temperature values at different measurement points can take values at different endpoints of the interval $[\underline{u}, \bar{u}]$. We introduce in the L-dimensional space a cube with $N = 2^L$ vertices at the points

$$T_1 = (\underline{u},\, \underline{u}, \ldots, \underline{u})^{\mathrm{T}} \in \mathbb{R}^L, \quad T_2 = (\underline{u},\, \underline{u}, \ldots, \bar{u})^{\mathrm{T}} \in \mathbb{R}^L, \ \ldots,$$
$$T_{N-1} = (\underline{u},\, \bar{u}, \ldots, \bar{u})^{\mathrm{T}} \in \mathbb{R}^L, \quad T_N = (\bar{u},\, \bar{u}, \ldots, \bar{u})^{\mathrm{T}} \in \mathbb{R}^L.$$

Note that inequalities (2.15) are linear with respect to the feedback coefficients. By denoting

$$U = (U_1, U_2, \ldots, U_L)^{\mathrm{T}}, \quad K_1 = (k_{11}, k_{12}, \ldots, k_{1L})^{\mathrm{T}}, \quad K_2 = (k_{21}, k_{22}, \ldots, k_{2L})^{\mathrm{T}},$$

from (2.15) we derive N linear constraints on the parameters K_1 and K_2 :

$$\underline{\vartheta} \leq \langle K_1 + K_2, T_s \rangle - \langle K_1 + K_2, U \rangle \leq \bar{\vartheta}, \quad s = 1, \ldots, N. \tag{2.16}$$

Here, $\langle a, b \rangle$ stands for the inner product of vectors a and b.

As is easily shown, the constraints (2.16) on the continuous feedback parameters will also be valid in the case of the discrete feedback law (2.11).

Let the optimized feedback coefficients and locations of measurement points be compiled into the vector

$$\mathbf{K} = (K_1,\ K_2) \in \mathbb{R}^{2L}.$$

For the feedback control problem, the objective functional (2.8) can be written as

$$J(\mathbf{K}) = \int\limits_{\Phi} \int\limits_{\Theta} I(\mathbf{K}; \varphi, \theta) \rho_{\Phi}(\varphi) d\theta d\varphi, \tag{2.17}$$

$$I(\mathbf{K}; \varphi, \theta) = \int\limits_0^l \mu(x) \left(u\left(x, t_f; \mathbf{K}, \varphi, \theta\right) - U(x) \right)^2 dx + \sigma \left\| \mathbf{K} \right\|^2_{\mathbb{R}^{2L}}. \tag{2.18}$$

Thus, the original optimal control problem for the boundary value condition, with both continuous and discrete feedback, has been reduced to a parametric optimal control problem.

According to (2.10) and (2.11), the feedback control approach proposed here involves not only current but also previous information about the process state at measurement points. Owing to this peculiarity of the approach, the process dynamics can be considered during control. Obviously, the above dependencies (2.10) and (2.11) can be further complicated, e.g., by adding more previous measurements.

An important feature of the resulting parametric optimal control problem is the possible nonconvexity (multi-extremality) of the objective functional (2.17), (2.18). Indeed, the dependencies (2.10) and (2.11) incorporate the product of the feedback coefficients and the temperature values of the rod at measurement points, which implicitly depend on the former coefficients. Hence, the objective functional (2.17), (2.18) should be minimized using nonlocal (global) optimization methods. We applied the multistart technique with a local optimization method launched from different starting points.

The peculiarity of the boundary value condition (2.12), i.e., the presence of a time delay, is considered without particular difficulties by the well-known method of steps [23]. When numerically solving the boundary value problem by the grid method, the nonlocality of this condition can be taken into account using sweep schemes [24, 25].

The gradient projection method will be used to solve the parametric optimal control problem with the linear constraints (2.16) on the optimized parameters numerically. For this purpose, in the next section, we prove the differentiability of the objective functional and derive formulas for the components of its gradient. The formulas will serve to formulate necessary conditions for the local optimality of the optimized parameters [5, 19].

3. THE NUMERICAL SOLUTION APPROACH TO THE CONTROL DESIGN PROBLEM

For the numerical solution of the parametric optimal control problem (2.1), (2.2), (2.4), (2.12), (2.17) with the linear constraints (2.13)–(2.16) on the optimized parameters (consequently, a convex admissible domain $\mathbf{K}$ of the parameters), we propose the gradient projection method [5]:

$$\mathbf{K}^{s+1} = \Pr_{(2.14)} \Big(\mathbf{K}^s - \alpha_s \, grad_K \; J(\mathbf{K}^s)\Big), \tag{3.1}$$

$$\alpha_s = \arg\min_{\alpha \geq 0} J\left(\Pr_{(2.14)} \Big(\mathbf{K}^s - \alpha \, grad_K \; J(\mathbf{K}^s)\Big)\right), \quad s = 0, 1, \ldots. \tag{3.2}$$

Here, $grad_K J(\mathbf{K})$ is the $2L$-dimensional vector formed by the components of the gradient of the objective functional (2.17), (2.18); α_s is the step size of one-dimensional optimization along the antigradient of the functional at the sth iteration, determined, e.g., by the golden section method; finally, $\mathbf{K}^0$ is some initial value of the sought parameter vector. (If this value violates the constraints (2.12), (2.14), it must first be projected onto the admissible domain.) There are known constructive formulas [5, 26] to obtain the projection operator $\Pr_{(2.14)}(\mathbf{K})$ of an arbitrary vector $\mathbf{K} \in \mathbb{R}^{2L}$ onto the admissible domain defined by the linear inequalities (2.12), (2.16).

As noted above, the parametric optimal control problem may be multi-extremal with respect to the optimized parameters $\mathbf{K}$, and the method (3.1), (3.2) yields only the local minimum point of the objective functional nearest to the point $\mathbf{K}^0 \in \mathbb{R}^{2L}$. Therefore, the multistart technique was selected for computations: the iterative procedure (3.1), (3.2) was applied several times for different admissible initial points $\mathbf{K}^0$. The solution is the local minimum vector $\mathbf{K}^*$ corresponding to the smaller value of the objective functional.

Clearly, the gradient $grad_{\mathbf{K}} \; J(\mathbf{K})$ of the objective functional plays an important role in implementing the procedure (3.1), (3.2). The next result establishes the differentiability of the objective functional $J(\mathbf{K})$ and provides formulas for the components of its gradient.

Theorem 1. *Under the above assumptions on the functions and parameters of the optimal control problem* (2.1), (2.2), (2.12), (2.4), (2.17), (2.18) *with the feedback law* (2.10), *the objective functional* (2.17), (2.18) *is differentiable with respect to the admissible values of the parameters* $\mathbf{K}$, *and the components of its gradient are given by*

$$\frac{\partial J(\mathbf{K})}{\partial k_{1i}} = \int\limits_{\Phi} \int\limits_{\Theta} \left[\int\limits_0^{t_f} a^2 \mu_2 \psi(0,t) \left(u(\xi_i, t) - U_i\right) \, dt + 2\sigma \, k_{1i} \right] \rho_\Phi(\varphi) d\theta d\varphi, \tag{3.3}$$

$$\frac{\partial J(\mathbf{K})}{\partial k_{2i}} = \int\limits_{\Phi} \int\limits_{\Theta} \left[\int\limits_0^{t_f - \tau} a^2 \mu_2 \psi(0,t+\tau) \, \left(u(\xi_i, t+\tau) - U_i\right) dt + 2\sigma \, k_{2i} \right] \rho_\Phi(\varphi) d\theta d\varphi, \tag{3.4}$$

where $i=1,\ldots,L$ *and* $u(x,t)=u(x,t;\mathbf{K},\varphi,\theta)$ *is the solution of the initial-boundary value problem* (2.1), (2.2), (2.4), (2.14); *an almost everywhere twice differentiable, with respect to* $x\in(0,l)$ *and* $t\in[0,t_f]$, *function* $\psi(x,t;\mathbf{K},\varphi,\theta)$ *is the solution of the following adjoint initial-boundary value problem under given admissible values of the parameters* $\mathbf{K},\varphi,\theta$:

$$\frac{\partial\psi(x,t)}{\partial t}=-a^2\frac{\partial^2\psi(x,t)}{\partial x^2}-\mu_1\psi(x,t),\quad (x,t)\in\Omega=(0,l)\times(0,t_f],\tag{3.5}$$

$$\psi(x,t_f)=-2\mu(x)\left(u\left(x,t_f;\mathbf{K},\varphi,\theta\right)-U(x)\right),\quad 0\le x\le l,\tag{3.6}$$

$$\psi_x(0,t)=\mu_2\psi(0,t),\quad 0\le t\le t_f,\tag{3.7}$$

$$\psi_x(l,t)=-\mu_2\psi(l,t),\quad 0\le t\le t_f,\tag{3.8}$$

and it satisfies the conditions

$$\begin{gathered}\frac{\partial\psi(\xi_i^-,t)}{\partial x}=\frac{\partial\psi(\xi_i^+,t)}{\partial x}+\frac{\partial\psi(\xi_i^+,t+\tau)}{\partial x}-\frac{\partial\psi(\xi_i^-,t+\tau)}{\partial x}\\ -k_{1i}\mu_2\psi(0,t)-k_{2i}\,\mu_2\psi(0,t+\tau),\quad 0\le t\le t_f-\tau,\end{gathered}\tag{3.9}$$

$$\psi_x(\xi_i^-,t)=\psi_x(\xi_i^+,t)-k_{1i}\mu_2\psi(0,t),\quad t_f-\tau\le t\le t_f,\tag{3.10}$$

$$\psi(\xi_i^-,t)=\psi(\xi_i^+,t),\quad 0\le t\le t_f,\tag{3.11}$$

at the points ξ_i, $i=1,\ldots,L$, *for* $t\in[0,t_f)$.

In the formulas presented, $\chi_{[0,t_f-\tau]}(t)$ *denotes the characteristic function:*

$$\chi_{[0,t_f-\tau]}(t)=\begin{cases}0, & t\notin[0,t_f-\tau],\\ 1, & t\in[0,t_f-\tau].\end{cases}$$

The proof of Theorem 1 is postponed to the Appendix.

In the case of the discrete-time feedback law (2.11), we have the following.

Theorem 2. *Under the above assumptions on the functions and parameters of problem* (2.1), (2.2), (2.14), (2.4), (2.17), (2.18), *the objective functional* (2.17), (2.18) *for the discrete feedback law* (2.11) *is differentiable with respect to the parameters* $\mathbf{K}$ *under their admissible values, and the components of the gradient of the objective functional are given by*

$$\frac{\partial J(\mathbf{K})}{\partial k_{1i}}=\int\limits_{\Phi}\int\limits_{\Theta}\left[\sum_{j=0}^{M-1}\int\limits_{\bar t_j}^{\bar t_j+\Delta t}a^2\mu_2\psi(0,t)\left(u(\xi_i,t)-U_i\right)dt+2\sigma\,k_{1i}\right]\rho_\Phi(\varphi)d\theta d\varphi,\ \ \bar t_j\le t\le\bar t_j+\Delta t,$$

$$\frac{\partial J(\mathbf{K})}{\partial k_{2i}}=\int\limits_{\Phi}\int\limits_{\Theta}\left[\sum_{j=0}^{M-1}\int\limits_{\bar t_j}^{\bar t_j+\Delta t-\tau}\Big(a^2\mu_2\psi(0,t+\tau)\left(u(\xi_i,t+\tau)-U_i\right)dt\Big)+2\sigma\,k_{2i}\right]\rho_\Phi(\varphi)d\theta d\varphi,$$

$$\bar t_j-\tau\le t\le\bar t_j+\Delta t-\tau,$$

where $i=1,\ldots,L$ *and, for given admissible values of the parameters* $\mathbf{K},\varphi,\theta$ *and the corresponding function* $u(x,t;\mathbf{K},\varphi,\theta)$ (*the solution of the initial-boundary value problem* (2.1), (2.2), (2.12), (2.4)), *the function* $\psi(x,t)=\psi(x,t;\mathbf{K},\varphi,\theta)$ *is the solution of the adjoint initial-boundary value problem*

$$\begin{gathered}\frac{\partial\psi(x,t)}{\partial t}=-a^2\frac{\partial^2\psi(x,t)}{\partial x^2}-\mu_1\psi(x,t),\quad \bar t_j\le t\le\bar t_j+\Delta t,\quad j=0,\ldots,M-1.\\ \psi(x,t_f)=-2\mu(x)\left(u\left(x,t_f;\mathbf{K},\varphi,\theta\right)-U(x)\right),\quad 0\le x\le l,\\ \psi_x(0,t)=\mu_2\psi(0,t),\quad 0\le t\le t_f,\quad \bar t_j\le t\le\bar t_j+\Delta t,\quad j=0,\ldots,M-1,\\ \psi_x(l,t)=-\mu_2\psi(l,t),\quad 0\le t\le t_f,\quad \bar t_j\le t\le\bar t_j+\Delta t,\quad j=0,\ldots,M-1,\end{gathered}$$

and it satisfies the conditions

$$\frac{\partial\psi(\xi_i^-,t)}{\partial x} = \frac{\partial\psi(\xi_i^+,t)}{\partial x} + \frac{\partial\psi(\xi_i^+,t+\tau)}{\partial x} - \frac{\partial\psi(\xi_i^-,t+\tau)}{\partial x}$$

$$-\sum_{j=1}^{M-1}\delta(t-t_j)\int\limits_{t_{j-1}}^{t_j}[k_{1i}\mu_2\psi(0,t)+k_{2i}\,\mu_2\psi(0,t+\tau)]dt,$$

$$\bar{t}_j \le t \le \bar{t}_j + \Delta t - \tau, \quad i=1,\ldots,L,$$

$$\frac{\partial\psi(\xi_i^-,t)}{\partial x} = \frac{\partial\psi(\xi_i^+,t)}{\partial x} - \sum_{j=1}^{M-1}\delta(t-t_j)\int\limits_{t_{j-1}}^{t_j}k_{1i}\mu_2\psi(0,t)dt,$$

$$\bar{t}_j + \Delta t - \tau \le t \le \bar{t}_j + \Delta t, \quad i=1,\ldots,L,$$

$$\psi(\xi_i^+,t) = \psi(\xi_i^-,t), \quad \bar{t}_j \le t \le \bar{t}_j + \Delta t, \quad i=1,\ldots,L,$$

at the points ξ_i, $i=1,\ldots,L$, *for* $\bar{t}_j \le t \le \bar{t}_j + \Delta t$ *and the conditions*

$$\psi(x,t_j^-) = \psi(x,t_j^+), \quad j=0,\ldots,M-1,$$

at the points $t_j, j=0,\ldots,M-1$ *for* $x\in[0,l]$.

Let us state necessary optimality conditions for the parameters $\mathbf{K}$ in variational form.

Theorem 3. *For the local optimality of the parameters* $\mathbf{K}^*\in\mathbb{R}^{2L}$ *in the control design problem* (2.1), (2.2), (2.4), (2.12), (2.17), (2.18) *with the continuous feedback law* (2.10), *it is necessary that*

$$(grad_{\mathbf{K}}J(\mathbf{K}),\mathbf{K}-\mathbf{K}^*) \ge 0$$

for an arbitrary admissible vector $\mathbf{K}\in\mathbb{R}^{2L}$ *satisfying conditions* (2.14), *where the gradient of the objective functional* (2.17), (2.18) *is given by formulas* (3.3), (3.4).

In the discrete feedback case, necessary optimality conditions for the parameters are formulated by analogy. The proofs of Theorems 2 and 3 can be found in the Appendix.

4. THE RESULTS OF COMPUTER EXPERIMENTS

In this section, we present the results of numerical experiments for the following functions and parameter values of the problem statement:

$$t_0=0, \quad t_f=2, \quad a=1, \quad \mu_1=0.1, \quad \mu(x)=0.1, \quad l=1,$$

$$\Phi=\{40,\,45,\,50\}, \quad \Theta=\{5,\,6,\,7\}, \quad \rho_\Phi(\varphi_i)=\rho_\Theta(\theta_i)=\frac{1}{3}, \quad i=1,2,3,$$

$$\underline{\vartheta}=10, \quad \bar{\vartheta}=750, \quad U(x)=70, \quad x\in[0,1].$$

The direct and adjoint initial-boundary value problems were numerically solved using an implicit scheme of the grid method [19]. The experiments were carried out for different values of the grid step. The experimental results below correspond to the steps $h_x=0.01$ and $h_t=0.05$.

The nonlocality in the left boundary value condition (2.12) was taken into account using an approach based on the boundary condition transfer scheme [20–22]. To consider the time delay in the boundary value condition, the method of steps with a step size of τ was applied [23].

Table 1. The components of the normalized gradients of the functional computed by formulas (3.5)–(3.7) and (4.1) for the parameter values given in Table 2

N	Formulas	$\nabla_{k_{11}}^{norma} J$	$\nabla_{k_{12}}^{norma} J$	$\nabla_{k_{21}}^{norma} J$	$\nabla_{k_{22}}^{norma} J$
1	(3.3), (3.4)	0.286391	0.437807	0.354151	0.775162
	(4.1)	0.286693	0.437915	0.354402	0.774879
2	(3.3), (3.4)	0.008363	−0.55936	−0.698830	−0.445737
	(4.1)	0.012542	−0.55928	−0.699544	−0.444614
3	(3.3), (3.4)	0.322336	−0.585921	0.276813	0.823887
	(4.1)	0.322786	−0.586143	0.278189	0.824126

Table 2. The initial and resulting values of the optimized parameters and the corresponding values of the functional for $L = 2$

N	$(k_{11}^0;\, k_{12}^0)$; $(k_{21}^0; k_{22}^0)$	$J(\mathbf{K})$	$(k_{11}^*;\, k_{12}^*)$; $(k_{21}^*;\, k_{22}^*)$	$J(\mathbf{K}^*)$
1	(2.00; 5.00); (4.00; 7.00)	2132.76	(1.497; 2.499); (3.496; 3.001)	0.00043142
2	(0.20; 0.70); (0.40; 0.90)	2714.32	(1.399; 2.600); (3.194; 2.800)	0.00056845
3	(2.50; 1.50); (0.40; 3.00)	2478.19	(1.502; 2.4563); (3.3212; 2.925)	0.00048367

The feedback control problem was solved for two different initial values of the optimized parameters.

For each set of the initial values of the optimized parameters, Table 1 provides the components of the normalized gradients computed by formulas (3.3), (3.4) and using the central finite-difference approximation of the derivative of the functional:

$$\partial J(\mathbf{K})/\partial \mathbf{K}_j \approx (J(\mathbf{K}+\varepsilon e_j) - J(\mathbf{K}-\varepsilon e_j))/2\varepsilon, \tag{4.1}$$

where $\mathbf{K}_j$ is the jth component of the n-dimensional optimized vector $\mathbf{K}$ (the aggregate of the optimized parameters k_{11}, k_{12}, k_{21}, and k_{22}); and e_j is the n-dimensional vector consisting of zeros, except for the jth component equal to one. The value of ε in the experiments was varied, and the table contains the most acceptable results.

In Table 2, the reader can find the initial and optimal values of the parameters $\mathbf{K}$ for the feedback control law (2.10) with $L = 2$ measurement points.

Numerical experiments were carried out in which the exact values $u(\xi_1, t)$ and $u(\xi_2, t)$ of the process states observed at the $L = 2$ measurement points were corrupted by random noise:

$$u(t;\mathbf{K}^*) = u(\xi_i, t;\mathbf{K}^*) = u(\xi_i, t;\mathbf{K}^*)(1 + \chi(2\sigma_i - 1)), \quad i = 1, 2,$$

where $u(\xi_i, t;\mathbf{K}^*)$ is the solution of the direct initial-boundary value problem computed at the point $x = \xi_i$ for $t \in [0, t_f]$, σ_i is a random variable with the uniform distribution on the interval $[0, 1]$, and χ is the noise level.

Figure 1 shows the graphs of the functions representing, along the rod, the deviations $\Delta u(x, t_f;\mathbf{K}^*, \chi) = u(x, t_f;\mathbf{K}^*, \chi) - U(x)$ of the resulting temperatures from the desired one under different noise levels: 0% (no noise), 1%, 3%, and 5%, corresponding to χ equal to 0, 0.01, 0.03, and 0.05.

Also, the control design approach proposed above was experimentally compared with the solution of the feedback control design problem without using previous measurements in the form [16]:

$$\vartheta(t;\mathbf{K}) = \sum_{i=1}^{L} k_{1j}(u(\xi_i, t) - U_i), \quad t \in [t_0, t_f]. \tag{4.2}$$

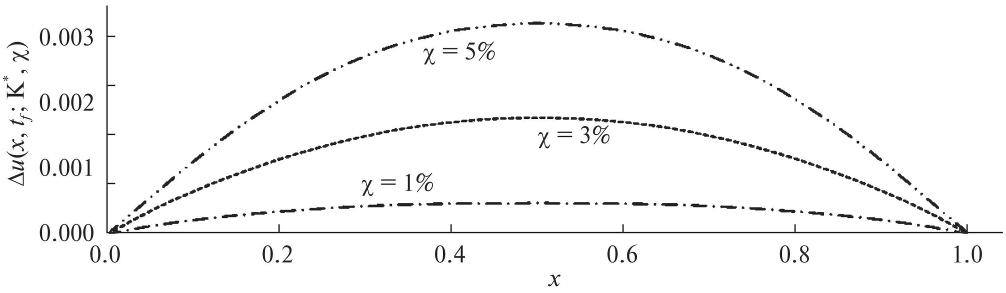

Fig. 1. The deviations of the temperature distribution functions along the rod from the desired one $U(x)$ under different noise levels χ.

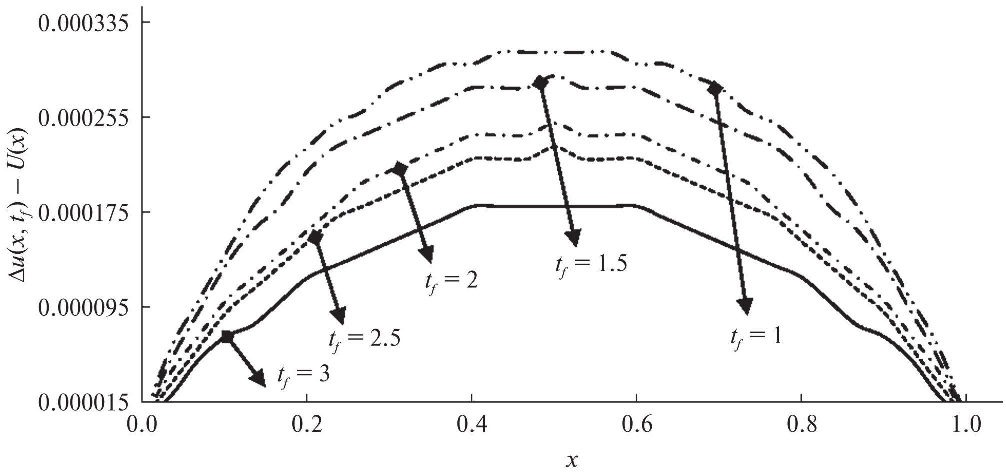

Fig. 2. Temperature distribution along the rod for different $t \in [1, 3]$. The ordinate axis has enlarged scale.

Clearly, the phase state at previous time instants is not used in (4.2) to form the current value of the control function $\vartheta(t)$.

Table 3 presents the numerical solution of the boundary control design problem using the feedback law (4.2). The experiments were carried out for three different initial values of the optimized parameters.

Table 3. The initial and resulting values of the feedback parameters (4.2) and the objective functional

N	$(k_{11}^0;\, k_{12}^0)$	$J(\mathbf{K^0})$	$(k_{11}^*;\, k_{12}^*)$	$J(\mathbf{K}^*)$
1	(2.00; 5.00)	2012.37	(1.2412; 4.4215)	0.0866534
2	(0.20; 0.70)	1987.13	(1.6512; 3.1265)	0.0872112
3	(2.50; 1.50)	2179.33	(2.2451; 1.1532)	0.0858423

Figure 2 demonstrates the graphs of the temperature deviations from the desired one for the obtained parameters and $t > t_f = 1$. According to the graphs, the heating process continues to settle, despite that the design problem was solved on the interval $[0, 1]$, i.e., $t_f = 1$.

5. CONCLUSIONS

In this paper, the optimal boundary control design problem for a rod heating process has been investigated. The problem statement includes a linear dependence of the boundary control value on the rod temperature values at measurement points, both at the current and previous time instants. The parameters optimized in this problem are the constant feedback parameters representing the coefficients in the dependence of the control value on the temperature values at the measurement points.

Determining the values of the feedback parameters has reduced the original problem to a parametric optimal control problem with distributed parameters. First-order numerical optimization methods have been applied to solve it. Computer experiments have been carried out.

The boundary control design approach and analysis method proposed in this paper can be extended to other types of PDEs with other initial-boundary value conditions.

APPENDIX

Proof of Theorem 1. Due to the mutual independence of all possible values of the parameters of the initial conditions $\gamma(x;\varphi)$ under different parameters φ and external influences θ, we have

$$grad_K\, J(\mathbf{K}) = grad_{\mathbf{K}} \int\limits_{\Phi}\int\limits_{\Theta} I(\mathbf{K};\varphi,\theta)\rho_\Phi(\varphi)d\theta d\varphi = \int\limits_{\Phi}\int\limits_{\Theta} grad_{\mathbf{K}} I(\mathbf{K};\varphi,\theta)\rho_\Phi(\varphi)d\theta d\varphi. \tag{A.1}$$

Therefore, let us study the differentiability of the functional $I(\mathbf{K};\varphi,\theta)$ and derive formulas for the components of its gradient with respect to the optimized parameters $\mathbf{K}$ under any given admissible parameters $\varphi \in \Phi$ and $\theta \in \Theta$.

To prove this theorem, we utilize the well-known method of incrementing the optimized parameters and estimate the corresponding increment of the functional [5, 27].

Let the parameter vector $\mathbf{K} = (k_1,\, k_2)$, associated with the solution $u(x,t) = u(x,t,\mathbf{K})$ of the boundary value problem (2.1), (2.2), (2.4), be incremented to $\tilde{\mathbf{K}} = \mathbf{K} + \Delta\mathbf{K} = (k_1 + \Delta k_1,\ k_2 + \Delta k_2)$, and the latter vector is associated with the solution $\tilde{u}(x,t) = \tilde{u}(x,t,\tilde{\mathbf{K}}) = u(x,t) + \Delta u(x,t)$ of the boundary value problem.

From (2.1), (2.2), (2.4), and (2.12) it follows that $\Delta u(x,t)$ is the solution of the boundary value problem

$$\frac{\partial \Delta u(x,t)}{\partial t} = a^2 \frac{\partial^2 \Delta u(x,t)}{\partial x^2} - \mu_1 \Delta u(x,t), \quad (x,t) \in \Omega = (0,l) \times (0,t_f], \tag{A.2}$$

$$\Delta u(x,0) = 0, \quad x \in [0,l], \tag{A.3}$$

$$\begin{aligned} \left.\frac{\partial \Delta u(x,t)}{\partial x}\right|_{x=0} &= \mu_2 \Delta u(0,t) - \sum_{i=1}^{L} \left[k_{1i}\Delta u(\xi_i,t) + k_{2i}\Delta u(\xi_i,t-\tau)\right] \\ &+ \sum_{i=1}^{L} \left[(u(\xi_i,t) - U_i)\,\Delta k_{1i} + (u(\xi_i,t-\tau) - U_i)\,\Delta k_{2i}\right], \quad t \in [0,t_f]\,, \end{aligned} \tag{A.4}$$

$$\left.\frac{\partial \Delta u(x,t)}{\partial x}\right|_{x=l} = -\mu_2 \Delta u(l,t), \quad t \in [0,t_f]\,. \tag{A.5}$$

The increment of the functional (2.17) can be, in a straightforward way, represented as

$$\begin{aligned} \Delta I(\mathbf{K};\varphi,\theta) &= I(\mathbf{K}+\Delta\mathbf{K};\varphi,\theta) - I(\mathbf{K};\varphi,\theta) \\ &= 2\int\limits_0^l \mu(x)\left[u\left(x,t_f;\mathbf{K},\varphi,\theta\right) - U(x)\right]\,\Delta u\left(x,t_f\right)dx \\ &+ 2\sigma \sum_{i=1}^{L} \left[k_{1i}\Delta k_{1i} + k_{2i}\Delta k_{2i}\right] + R\left(\|\Delta u\|_{L^2(\Omega)},\ \|\Delta\mathbf{K}\|_{\mathbb{R}^{2L}}\right), \end{aligned} \tag{A.6}$$

where the residual $R\left(\|\Delta u\|_{L^2(\Omega)},\ \|\Delta \mathbf{K}\|_{\mathbb{R}^{2L}}\right)$ includes all terms of the second order of smallness with respect to $\|\Delta \mathbf{K}\|_{\mathbb{R}^{2L}}$ and $\|\Delta u\|_{L^2(\Omega)}$. Some estimates of the increment of the solution of an initial-boundary value problem for a parabolic equation depending on the variations of the problem parameters were obtained in [5, 27]. According to these estimates,

$$\|\Delta u\|_{L^2(\Omega)} \leq \eta\ \|\Delta \mathbf{K}\|_{\mathbb{R}^{2L}},$$

where the value of $\eta > 0$ is independent of the parameters $\mathbf{K}$. We obtain formulas for the components of the gradient of the objective functional with respect to the parameters $\mathbf{K}$. Let $\psi(x,t)$ be some, for now arbitrary, function continuous everywhere in Ω, twice differentiable with respect to $x \in (\xi_i, \xi_{i+1})$, $i = 0, \ldots, L$, $\xi_0 = 0$, $\xi_{L+1} = l$, and differentiable with respect to $t \in (0, t_f)$. We multiply (A.2) by $\psi(x,t)$ and integrate the result over the rectangle Ω. Under the above assumptions and conditions (A.4)–(A.5), the result is

$$\int_0^{t_f}\int_0^{l} \psi(x,t)\, \frac{\partial \Delta u(x,t)}{\partial t}\, dxdt - a^2 \sum_{i=0}^{L} \int_{\xi_i}^{\xi_{i+1}}\int_0^{t_f} \psi(x,t)\, \frac{\partial^2 \Delta u(x,t)}{\partial x^2}\, dtdx \tag{A.7}$$
$$- \mu_1 \int_0^{t_f}\int_0^{l} \psi(x,t) \Delta u(x,t)\, dxdt = 0.$$

Let us split the interval $[0, l]$ into the subintervals $[\xi_i, \xi_{i+1}]$, $i = 0, \ldots, L$, and perform integration by parts separately for the first and second terms of (3.5). In view of (A.3)–(A.5), we have

$$\int_0^{t_f}\int_0^{l} \psi(x,t)\, \frac{\partial \Delta u(x,t)}{\partial t}\, dxdt = \int_0^{l} \psi(x,t_f) \Delta u(x,t_f)\, dx - \int_0^{t_f}\int_0^{l} \frac{\partial \psi(x,t)}{\partial t}\, \Delta u(x,t)\, dxdt, \tag{A.8}$$

$$a^2 \sum_{i=0}^{L} \int_{\xi_i}^{\xi_{i+1}}\int_0^{t_f} \psi(x,t)\, \frac{\partial^2 \Delta u(x,t)}{\partial x^2}\, dtdx \tag{A.9}$$
$$= a^2 \int_0^{t_f} \left[\left(\frac{\partial \psi(0,t)}{\partial x} - \mu_2 \psi(0,t)\right) \Delta u(0,t) - \left(\frac{\partial \psi(l,t)}{\partial x} + \mu_2 \psi(l,t)\right) \Delta u(l,t)\right] dt$$
$$+ a^2 \sum_{i=1}^{L} \int_0^{t_f} \left[\frac{\partial \psi(\xi_i^-,t)}{\partial x} - \frac{\partial \psi(\xi_i^+,t)}{\partial x} + k_{1i}\mu_2 \psi(0,t)\right] \Delta u(\xi_i,t)\, dt$$
$$+ a^2 \sum_{i=1}^{L} \int_0^{t_f-\tau} \left[\frac{\partial \psi(\xi_i^-,t+\tau)}{\partial x} - \frac{\partial \psi(\xi_i^+,t+\tau)}{\partial x} + k_{2i}\, \mu_2 \psi(0,t+\tau)\right] \Delta u(\xi_i,t)\, dt$$
$$+ a^2 \sum_{i=1}^{L} \int_0^{t_f} \left[(\psi(\xi_i^+,t) - \psi(\xi_i^-,t))\, \frac{\partial \Delta u(\xi_i,t)}{\partial x}\right] dt - a^2 \sum_{i=1}^{L} \Delta k_{1i} \int_0^{t_f} \mu_2 \psi(0,t)\, (u(\xi_i,t) - U_i)\, dt$$
$$- a^2 \sum_{i=1}^{L} \Delta k_{2i} \int_0^{t_f-\tau} \mu_2 \psi(0,t+\tau)\, (u(\xi_i,t+\tau) - U_i)\, dt + a^2 \int_0^{t_f}\int_0^{l} \frac{\partial^2 \psi(x,t)}{\partial x^2}\, \Delta u(x,t)\, dxdt.$$

Based on (A.6)–(A.9), the increment of the functional takes the form

$$\Delta I(\mathbf{K};\varphi,\theta)=\int_0^l\Big[\psi(x,t_f)+2\mu(x)(u(x,t_f;\mathbf{K},\varphi,\theta)-U(x))\Big]\Delta u(x,t_f)\,dx \tag{A.10}$$

$$-a^2\int_0^{t_f}\Big[\Big(\frac{\partial\psi(0,t)}{\partial x}-\mu_2\psi(0,t)\Big)\Delta u(0,t)-\Big(\frac{\partial\psi(l,t)}{\partial x}+\mu_2\psi(l,t)\Big)\Delta u(l,t)\Big]\,dt$$

$$+\int_0^{t_f}\int_0^l\Big[-\frac{\partial\psi(x,t)}{\partial t}-a^2\frac{\partial^2\psi(x,t)}{\partial x^2}-\mu_1\psi(x,t)\Big]\Delta u(x,t)\,dxdt$$

$$+a^2\sum_{i=1}^L\int_0^{t_f-\tau}\Big[\frac{\partial\psi(\xi_i^-,t)}{\partial x}-\frac{\partial\psi(\xi_i^+,t)}{\partial x}+\frac{\partial\psi(\xi_i^-,t+\tau)}{\partial x}$$

$$-\frac{\partial\psi(\xi_i^+,t+\tau)}{\partial x}+k_{1i}\mu_2\psi(0,t)+k_{2i}\,\mu_2\psi(0,t+\tau)\Big]\Delta u(\xi_i,t)\,dt$$

$$+a^2\sum_{i=1}^L\int_{t_f-\tau}^{t_f}\Big[\frac{\partial\psi(\xi_i^-,t)}{\partial x}-\frac{\partial\psi(\xi_i^+,t)}{\partial x}+k_{1i}\mu_2\psi(0,t)\Big]\Delta u(\xi_i,t)\,dt$$

$$-\sum_{i=1}^L\Bigg\{a^2\int_0^{t_f}\mu_2\psi(0,t)\ (u(\xi_i,t)-U_i)\,dt+2\sigma k_{1i}\Bigg\}\Delta k_{1i}$$

$$-\sum_{i=1}^L\Bigg\{a^2\int_0^{t_f-\tau}\psi(0,t+\tau)\ (u(\xi_i,t+\tau)-U_i)\ dt+2\sigma\,k_{2i}\Bigg\}\Delta k_{2i}$$

$$+R\left(\|\Delta u\|,\ \|\Delta\mathbf{K}\|_{\mathbb{R}^{2L}}\right).$$

Due to the arbitrariness of the function $\psi(x,t)$, we require it to be almost everywhere the solution of the initial-boundary value problem (3.5)–(3.11).

Note that the components of the gradient of the functional are determined by the linear part of the increment of the functional at the increments of the corresponding arguments. Therefore,

$$\frac{\partial I(\mathbf{K};\varphi,\theta)}{\partial k_{1i}}=\int_0^{t_f}\Big(a^2\mu_2\psi(0,t)(u(\xi_i,t)-U_i)\Big)\ dt+2\sigma k_{1i},\ \ i=1,\dots,L,\quad 0\le t\le t_f,$$

$$\frac{\partial I(\mathbf{K};\varphi,\theta)}{\partial k_{2i}}=\int_0^{t_f-\tau}\Big(a^2\mu_2\psi(0,t+\tau)\,(u(\xi_i,t+\tau)-U_i)\Big)\,dt+2\sigma\,k_{2i},\ \ i=1,\dots,L,\quad 0\le t\le t_f-\tau.$$

By the expression (A.1), we finally arrive at the desired formulas of Theorem 2. The proof of Theorem 1 is complete.

Theorem 2 is proved similarly to Theorem 1. The main difference is related to formulas (2.11) and (2.12), which necessitate replacing integration over t on the entire interval $[0,t_f]$ with a sum of integrals over the intervals $[t_{j-1},t_j]$, $j=1,\dots,M$, $t_0=0$, $\bar{t}_M=t_f$.

Theorem 3 follows from the well-known theorem on integral necessary conditions for optimality in optimization problems [5, 27].

REFERENCES

1. Lions, J.L., *Optimal Control of Systems Governed by Partial Differential Equations*, Berlin–Heidelberg: Springer, 1971.

2. Butkovskii, A.G., *Metody upravleniya sistemami s raspredelennymi parametrami* (Control Methods for Distributed Parameter Systems), Moscow: Nauka, 1984.

3. Sirazetdinov, T.K., *Optimizatsiya sistem s raspredelennymi parametrami* (Optimization of Distributed Parameter Systems), Moscow: Nauka, 1977.

4. Sergienko, I.V. and Deineka, V.S., *Chislennyi sistemnyi analiz mnogokomponentnykh raspredelennykh sistem* (Numerical Systemic Analysis of Multicomponent Distributed Systems), Kiev: Naukova Dumka, 2009.

5. Vasil'ev, F.P., *Metody optimizatsii* (Optimization Methods), Moscow: Faktorial Press, 2008.

6. Egorov, A.I., *Osnovy teorii upravleniya* (Fundamentals of Control Theory), Moscow: Fizmatlit, 2004.

7. Polyak, B.T. and Shcherbakov, P.S., *Robastnaya ustoichivost' i upravlenie* (Robust Stability and Control), Moscow: Nauka, 2002.

8. Emelyanov, S.V. and Korovin, S.K., *Control of Complex and Uncertain Systems: New Types of Feedback*, London: Springer, 2000.

9. Khlebnikov, M.V., and Shcherbakov, P.S., Optimal Feedback Design under Bounded Control, *Autom. Remote Control*, 2014, vol. 75, no. 2, pp. 320–322.

10. Alessandri, A., Gaggero, M., and Zoppoli, R., Feedback Optimal Control of Distributed Parameter Systems by Using Finite-Dimensional Approximation Schemes, *IEEE Trans. Neural Netw. Learn. Syst.*, 2012, vol. 23, pp. 984–996.

11. Ray, W.H., *Advanced Process Control*, McGraw-Hill, 1981.

12. Smyshlyaev, A., and Krstic, M., Explicit State and Output Feedback Boundary Controllers for Partial Differential Equations, *Journal of Automatic Control*, 2003, vol. 13, no. 2, pp. 1–9.

13. Liu, W., Boundary Feedback Stabilization of an Unstable Heat Equation, *SIAM Journal on Control and Optimization*, 2003, vol. 42, pp. 1033–1043.

14. Afifi, L., Lasri, K., Joundi, M., and Amimi, N., Feedback Controls for Exact Remediability in Disturbed Dynamical Systems, *IMA Journal of Mathematical Control and Information*, 2018, vol. 35, no. 2, pp. 411–425.

15. Abdullayev, V.M., and Aida-zade, K.R., Rod Temperature Regulation Using Current and Time-Delayed Feedback, *Quaestiones Mathematicae*, 2023, vol. 46, no. 10, pp. 1991–2011.

16. Aida-zade, K.R. and Abdullayev, V.M., Optimizing Placement of the Control Points at Synthesis of the Heating Process Control, *Autom. Remote Control*, 2017, vol. 78, no. 9, pp. 1585–1599.

17. Aida-zade, K.R. and Abdullayev, V.M., Controlling the Heating of a Rod Using the Current and Preceding Time Feedback, *Autom. Remote Control*, 2022, vol. 83, no. 1, pp. 106–122.

18. Aida-zade, K.R. and Hashimov, V.A., Optimization of Measurement Points Positioning in a Border Control Synthesis Problem for the Process of Heating a Rod, *Autom. Remote Control*, 2018, vol. 79, no. 9, pp. 1643–1660.

19. Polyak, B.T., *Introduction to Optimization*, New York: Optimization Software, 1987.

20. Assanova, A.T., Dzhobulaeva, Z.K., and Imanchiyev, A.E., A Multi-Point Initial Problem for a Non-Classical System of a Partial Differential Equations, *Lobachevskii J. Math.*, 2020, vol. 41, no. 6, pp. 1031–1042.

21. Aida-zade, K.R., and Abdullayev, V.M., Optimization of the Right-Hand Sides of Nonlocal Conditions of a Controllable System with Multipoint and Integral Objective Functional, *Optimization*, 2024, vol. 73, no. 1, pp. 205–228.

22. Abdullayev, V.M., and Aida-zade, K.R., Approach to the Numerical Solution of Optimal Control Problems for Loaded Differential Equations with Nonlocal Conditions, *Comput. Math. Math. Phys.*, 2019, vol. 59, no. 5, pp. 696–707.

23. Elsgolts, L.E. and Norkin, S.B., *Introduction to the Theory and Application of Differential Equations with Deviating Arguments*, New York: Academic Press, 1973.

24. Samarskii, A.A. and Nikolaev, E.S., *Numerical Methods for Grid Equations. Vol. 1: Direct Methods; vol. 2: Iterative Methods*, Basel–Boston–Berlin: Birkhauser Verlag, 1989.

25. Abdullayev, V.M., and Aida-zade, K.R., Finite-Difference Methods for Solving Loaded Parabolic Equation, *Comput. Math. Math. Phys.*, 2016, vol. 56, no. 1, pp. 93–105.

26. Polak, E., *Computational Methods in Optimization: A Unified Approach*, New York–London: Academic Press, 1971.

27. Pshenichnyi, B.N., *Necessary Conditions for an Extremum*, New York: Marcel Dekker, 1971.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 1, pp. 44–64.
© The Author(s), 2026 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2026.
Russian Text © The Author(s), 2026, published in Avtomatika i Telemekhanika, 2026, No. 1, pp. 62–88.

Admissible and Optimal Control Design for Nonlinear Continuous–Discrete Dynamic Systems

S. V. Akmanova

Nosov Magnitogorsk State Technical University, Magnitogorsk, Russia
e-mail: svet.akm_74@mail.ru

Received April 5, 2025
Revised October 17, 2025
Accepted October 21, 2025

Abstract—This paper considers nonlinear continuous–discrete (hybrid) dynamic systems with a constant sampling step whose continuous and discrete parts operate within a single loop. Sufficient conditions for the complete controllability and existence of admissible and optimal control laws of such systems are established. Algorithms for designing admissible program control laws, as well as an algorithm for designing an optimal program control law, for such systems are presented. Some examples are provided to confirm the effectiveness of these control design algorithms for nonlinear hybrid systems with a constant sampling step on a finite horizon.

Keywords: continuous–discrete system, hybrid system, discrete system, admissible control, admissible process, optimal control, optimal process

DOI: 10.7868/S1608303226010047

1. INTRODUCTION. PROBLEM STATEMENT

Solving control problems for nonlinear dynamic systems is topical in many fields since these systems serve as mathematical models of various technical, mechanical, economic, and other processes; for example, see [1–4].

This paper addresses control problems for a dynamic system described by a combination of differential and difference equations; the latter include control, and the continuous and discrete parts of the system operate within a single loop. Such systems are commonly called continuous–discrete or hybrid dynamic control systems [2, 5].

Thus, we consider a nonlinear continuous–discrete control system of the form

$$\begin{cases} x'(t) = f(x(t), y(t_k)), \quad t_k \leqslant t < t_{k+1}, \\ y(t_{k+1}) = g(x(t_{k+1}), y(t_k), u(t_k)), \quad k = \overline{0, l-1}, \end{cases} \tag{1}$$

with initial conditions

$$x(t_0) = x_0, \quad y(t_0) = y_0, \tag{2}$$

where $x(t) \in R^n$ and $y(t_k) \in R^m$ are the state vectors of system (1) characterizing the behavior of its continuous and discrete parts, respectively; $u(t_k) \in R^q$ is the control vector (input) of the system; the time instants t_k define a uniform grid on the closed interval $[t_0, t_l]$ with a constant step $h = t_{k+1} - t_k > 0$ and nodes $t_k = t_0 + kh$, $0 \leq t_0 < t_l$, $k = 0, 1, \ldots, l$ ($l \in N$ is a fixed number); the functions $f(x, y)$ and $g(x, y, u)$ are continuously differentiable in the aggregate of variables on the sets $R^n \times R^m$ and $R^n \times R^m \times R^q$, respectively, and

$$f(0, 0) = 0, \quad g(0, 0, 0) = 0. \tag{3}$$

In other words, i.e., system (1) with $u = 0$ has the zero (trivial) equilibrium $x = 0,\ y = 0$.

According to (1) and the initial conditions (2), on each closed interval $[t_k, t_{k+1}]$ for a chosen control law u (i.e., for a given finite sequence of $u(t_k)$, $k = \overline{0, l-1}$), the solution $x = \varphi_k(t)$ of the Cauchy problem $x'(t) = f(x(t), y_k)$, $x(t_k) = x_k$ is found, and then the vectors $x_{k+1} = \varphi_k(t_{k+1})$ and $y_{k+1} = g(x_{k+1}, y_k, u_k)$, where $y_k = y(t_k)$ and $u_k = u(t_k)$, with $y(t) = y(t_k)$ for $t \in [t_k, t_{k+1})$, are constructed. The state of system (1) at a time instant t is the pair $z(t) = (x(t), y(t))$, and hence the solution of this system with the initial conditions $z_0 = (x_0, y_0)$ for a chosen control law u is the function

$$z(t) = (x(t), y(t)) = \begin{cases} (\varphi_0(t), y_0), & t_0 \leqslant t < t_1, \\ (\varphi_1(t), y_1), & t_1 \leqslant t < t_2, \\ \vdots & \\ (\varphi_{l-1}(t), y_{l-1}), & t_{l-1} \leqslant t < t_l, \\ (\varphi_{l-1}(t), y_l), & t = t_l. \end{cases} \tag{4}$$

In the solution (4), the component $x(t)$ is continuous for all $t \in [t_0, t_l]$, continuously differentiable on each open interval (t_k, t_{k+1}), $k = \overline{0, l-1}$, but not necessarily differentiable at time instants $t = t_k$, $k = \overline{0, l}$; the component $y(t)$ is piecewise constant and changes its values only at time instants $t = t_k$, $k = \overline{0, l}$.

An example of system (1) is the mathematical model of aircraft flying in a plane [4].

We pose the following problems: 1) the transfer of system (1) from a given initial state $z(t_0) = z^{\{0\}}$ to a given terminal state $z(t_l) = z^{\{1\}}$ ($t_l = t_0 + lh$) using a suitable control law; 2) optimization, i.e., finding a control law that transfers system (1) from $z(t_0) = z^{\{0\}}$ to $z(t_l) = z^{\{1\}}$ and minimizes the control performance index (cost functional)

$$I(u) = \sum_{k=0}^{l-1} u^\top(t_k) u(t_k). \tag{5}$$

The solution of the posed problems is related to the grid step $h > 0$ since the value of h affects the existence of a control law transferring system (1) from an initial state $z^{\{0\}}$ to a terminal one $z^{\{1\}}$. Therefore, we adopt the following definitions.

Definition 1. The hybrid system (1) is said to be completely controllable for a given $h > 0$ if for any vectors $z^{\{0\}}$, $z^{\{1\}} \in R^{n+m}$, there exists a control law $u(t_k)$, $k = \overline{0, l-1}$, such that $z(t_l) = z^{\{1\}}$ for the solution $z = z(t)$ of system (1) with the initial condition $z(t_0) = z^{\{0\}}$.

Definition 2. A control law $u(t_k)$, $k = \overline{0, l-1}$, that satisfies system (1) for a given $h > 0$ and transfers this system from a given initial state $z(t_0) = z^{\{0\}}$ to a given terminal state $z(t_l) = z^{\{1\}}$ is called an admissible (program) control law of system (1) for $h > 0$.

Definition 3. An admissible control law $u(t_k)$, $k = \overline{0, l-1}$, of system (1) for a given $h > 0$ that minimizes the cost functional (5) is called an optimal (program) control law of system (1) for $h > 0$.

Note that in [6, 7], the author established necessary and sufficient conditions for complete controllability, as well as sufficient conditions and stabilization methods, for a nonlinear hybrid system described by equalities (1) but operating on an infinite horizon (i.e., for $k = 0, 1, 2, \ldots$), in contrast to system (1) that operates on a finite horizon. Therefore, the concepts of complete controllability for system (1) and the one in [6, 7] are defined differently.

The generalizations of system (1) are the hybrid systems considered in [5, 8], where optimization problems with a non-fixed terminal state were posed. Note that in [5], V.F. Krotov's optimality principle [9] was applied jointly to a combination of differential and difference equations to establish sufficient conditions for optimal control of a nonlinear system; and in [8], necessary conditions of optimal control were obtained to solve problems with non-fixed switching instants, which can be chosen during process optimization. The above optimization problem for system (1) assumes a fixed terminal state of the system, with switching instants and their number known in advance for a given $h > 0$; moreover, the continuous subsystem of system (1) contains no control input. As a result, the optimization approaches presented in [5, 8] are not applicable for solving the optimization problem of system (1). In addition, the problem of transferring hybrid systems from one state to another based on designing admissible control laws was not considered in [5, 8].

We also emphasize that scientific literature covers the issues of admissible program control design for nonlinear discrete systems with the cost functional (5) [10, 11] as well as various aspects of admissible and optimal control design for nonlinear continuous systems via the first approximation [12–16], the issues of optimal control for discrete systems [17, 18], and the issues of optimality for continuous and discrete processes [9, 19, 20]. There are also studies on optimal control for certain types of nonlinear hybrid systems (e.g., see [21–24]). As a rule, however, they are devoted to systems with a non-fixed terminal state, and the states of such systems are characterized by only one phase vector, changing continuously between switching instants and discretely at switching instants. In contrast, the state of system (1) is characterized by a set of two different phase vectors, one changing continuously according to differential equations and the other discretely according to the difference equations of the system that contain control. Thus, the issues related to the posed problems are still open.

To solve these problems, we perform a transition from system (1) to an equivalent, in a natural sense, nonlinear discrete system [7]. Transferring the equivalent discrete system from initial to terminal states involves an iterative process of designing an admissible control law, which may be convergent or divergent. Therefore, two admissible control design approaches are considered for the nonlinear hybrid system (1).

The primary objectives of this paper are to establish sufficient conditions for the admissibility and optimality of a control law for the nonlinear hybrid system (1) and design corresponding admissible and optimal control laws for this system.

The novelty of this research is as follows:

1) Sufficient conditions for the complete controllability of system (1) are established. Under these conditions, there exist admissible program control laws transferring the system from any initial state to any terminal one on a finite horizon $[t_0, t_l]$.

2) A sufficient condition for the existence of an optimal control law for system (1) is established.

3) Algorithms for designing admissible and optimal control laws for system (1) are presented.

2. TRANSITION TO A DISCRETE SYSTEM. COMPLETE CONTROLLABILITY CRITERION FOR SYSTEM (1)

Let $x(t_k) = x_k$, $y(t_k) = y_k$, and $u(t_k) = u_k$; using [7] and the relations (3), we pass from system (1) to an equivalent nonlinear discrete system of the form

$$\begin{cases} x_{k+1} = e^{A_1 h} x_k + A_1^{-1}(e^{A_1 h} - I)B_1 y_k + \varepsilon(x_k, y_k; h), \quad k = \overline{0, l-1}, \\ y_{k+1} = A_2 e^{A_1 h} x_k + (A_2 A_1^{-1}(e^{A_1 h} - I)B_1 + B_2) y_k + C u_k + \delta(x_k, y_k, u_k; h), \end{cases}$$

where $A_1 = f'_x(0,0)$, $B_1 = f'_y(0,0)$, $A_2 = g'_x(0,0,0)$, $B_2 = g'_y(0,0,0)$, and $C = g'_u(0,0,0)$ are matrices of appropriate dimensions, and $det A_1 \neq 0$; in addition,

$$f(x,y) = A_1 x + B_1 y + a(x,y), \quad g(x,y,u) = A_2 x + B_2 y + Cu + b(x,y,u),$$

and the smooth nonlinearities $a(x,y)$ and $b(x,y,u)$ start with quadratic terms in x, y and x, y, u, respectively, satisfying the conditions $a(x,y) = o(\|x\| + \|y\|)$ as $\|x\| + \|y\| \to 0$ and $b(x,y,u) =$

$o(\|x\|+\|y\|+\|u\|)$ as $\|x\|+\|y\|+\|u\| \to 0$ (from this point onwards, $\|\cdot\|$ means the Euclidean norm of a vector in the corresponding space of real numbers or the norm of a matrix);

$$\varepsilon(x_k, y_k; h) = e^{(t_k+h)A_1} \int\limits_{t_k}^{t_k+h} e^{-sA_1} a(p(s, x_k, y_k), y_k) ds, \tag{6}$$

where $x = p(t, x_k, y_k)$ is the solution of the Cauchy problem $x' = A_1 x + B_1 y_k + a(x, y_k)$, $x(t_k) = x_k$; finally,

$$\delta(x_k, y_k, u_k; h) = A_2 \varepsilon(x_k, y_k; h) + b(x_{k+1}, y_k, u_k). \tag{7}$$

The equivalent nonlinear discrete system can be compactly written as

$$z_{k+1} = A(h) z_k + B u_k + \xi(z_k, u_k, h), \quad k = \overline{0, l-1}, \tag{8}$$

where

$$z_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}, \quad B = \begin{bmatrix} O \\ C \end{bmatrix}, \quad \xi(z_k, u_k, h) = \begin{bmatrix} \varepsilon(x_k, y_k; h) \\ \delta(x_k, y_k, u_k; h) \end{bmatrix},$$

and $A(h)$ is a block matrix (of order $n+m$) given by

$$A(h) = \begin{bmatrix} e^{A_1 h} & A_1^{-1}(e^{A_1 h} - I)B_1 \\ A_2 e^{A_1 h} & A_2 A_1^{-1}(e^{A_1 h} - I)B_1 + B_2 \end{bmatrix}. \tag{9}$$

System (8) with $u = 0$ has the zero (trivial) equilibrium, and the function $\xi(z, u, h)$ satisfies the condition $\xi(z, u, h) = o(\|z\| + \|u\|)$ as $\|z\| + \|u\| \to 0$ [7, 25].

Definition 4. The discrete system (8) is said to be completely controllable for a given $h > 0$ if for any states $z^{\{0\}}$, $z^{\{1\}} \in R^{n+m}$ there exists a control law u_k, $k = \overline{0, l-1}$, such that $z_l = z^{\{1\}}$, $z_l = z(t_l)$, for the solution z_k, $k = \overline{0, l}$, of system (8) with the initial condition $z_0 = z^{\{0\}}$.

We have the following result.

Theorem 1. *The hybrid system* (1) *is completely controllable for a given* $h > 0$ *if and only if the discrete system* (8) *is completely controllable for* $h > 0$.

The proof of Theorem 1 and other main statements is provided in the Appendix.

3. OPTIMAL CONTROL OF THE LINEAR HYBRID SYSTEM

Let the first approximation system of the nonlinear discrete system (8) be completely controllable for $h = h_0 > 0$. In other words, the discrete system

$$z_{k+1} = A_0 z_k + B u_k, \quad k = \overline{0, l-1}, \tag{10}$$

where $A_0 = A(h_0)$ due to (9), is completely controllable. Then, according to the complete controllability criterion of linear discrete systems [10, 11],

$$\operatorname{rank}[B, A_0 B, A_0^2 B, \ldots, A_0^{n+m-1} B] = n + m. \tag{11}$$

System (10) is equivalent, in a natural sense [7], to the linear hybrid system

$$\begin{cases} x'(t) = A_1 x(t) + B_1 y(t_k), \quad t_k \leqslant t < t_{k+1}, \quad t_{k+1} - t_k = h_0, \\ y(t_{k+1}) = A_2 x(t_{k+1}) + B_2 y(t_k) + C u(t_k), \quad k = \overline{0, l-1}. \end{cases} \tag{12}$$

Based on [11], we arrive at the following result.

Theorem 2. *If system* (10) *is completely controllable, then there exists an optimal control law* $u^{(0)}(t_k) = u_k^{(0)}$ *transferring the hybrid system* (12) *from any initial state* $z(t_0) = z^{\{0\}}$ *to any terminal state* $z(t_l) = z^{\{1\}}$ *and minimizing the cost functional* (5), *with*

$$u_k^{(0)} = S^\top(k)F^+(0)d(0, z^{\{0\}}), \quad S(k) = A_0^{l-k-1}B, \quad k = \overline{0, l-1},$$
$$d(0, z^{\{0\}}) = z^{\{1\}} - A_0^l z^{\{0\}}, \quad F(0) = \sum_{k=0}^{l-1} S(k)S^\top(k), \tag{13}$$

and $F^+(0)$ *as the pseudoinverse of the matrix* $F(0)$.

This optimal control law of system (12) leads to the optimal motion $z^{(0)}(t) = (x^{(0)}(t), y^{(0)}(t))$, $t \in [t_0, t_l]$, of this system and the optimal motion $z_k^{(0)} = [x_k^{(0)} \;\; y_k^{(0)}]^\top$ $(k = \overline{0, l})$ of system (10), with $x^{(0)}(t_k) = x_k^{(0)}$ and $y^{(0)}(t_k) = y_k^{(0)}$.

Example 1. For the linear hybrid system

$$\begin{cases} x'(t) = x(t) + 2y(t_k), & t_k \leqslant t < t_{k+1}, \\ y(t_{k+1}) = -x(t_{k+1}) + y(t_k) + 3u(t_k), & k = 0, 1, 2, \end{cases} \tag{14}$$

it is required to design an optimal control law $u^{(0)}(t_k)$ transferring this system in three steps $(l = 3)$ from the initial state $z^{\{0\}} = [2 \;\; 1]^\top$ to the terminal one $z^{\{1\}} = [1 \;\; 2]^\top$.

The hybrid system (14) is equivalent to the linear discrete system

$$\begin{cases} x_{k+1} = e^h x_k + 2(e^h - 1)y_k \\ y_{k+1} = -e^h x_k + (3 - 2e^h)y_k + 3u_k, & k = 0, 1, 2. \end{cases}$$

For $h_0 = ln2$, it takes the form (10) with the matrices $A_0 = \begin{bmatrix} 2 & 2 \\ -2 & -1 \end{bmatrix}$, and $B = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$. Condition (11) holds, i.e., this system is completely controllable for $h_0 = ln2$. Using formulas (13), we find

$$S(0) = A_0^2 B = [6 \;\; -9]^\top, \quad S(1) = A_0 B = [6 \;\; -3]^\top, \quad S(2) = B = [0 \;\; 3]^\top,$$
$$d(0, z^{\{0\}}) = z^{\{1\}} - A_0^3 z^{\{0\}} = [11 \;\; -1]^\top,$$
$$F(0) = \sum_{k=0}^{2} S(k)S^\top(k) = \begin{bmatrix} 72 & -72 \\ -72 & 99 \end{bmatrix}, \quad F^+(0) = F^{-1}(0) = \begin{bmatrix} 11/216 & 1/27 \\ 1/27 & 1/27 \end{bmatrix}.$$

For $t_{k+1} - t_k = h_0$ and $t_0 = 0$, system (14) has the optimal control law $u^{(0)}(t_0) = u^{(0)}(0) = -7/36$, $u^{(0)}(t_1) = u^{(0)}(ln2) = 73/36$, $u^{(0)}(t_2) = u^{(0)}(ln4) = 10/9$, and the corresponding optimal motion

$$z(t) = (x(t), y(t)) = \begin{cases} (4e^t - 2, \;\; 1), & 0 \leqslant t < ln2, \\ \left(-\dfrac{31}{12}e^t + \dfrac{67}{6}, -\dfrac{67}{12}\right), & ln2 \leqslant t < ln4, \\ \left(\dfrac{1}{24}e^t + \dfrac{2}{3}, -\dfrac{1}{3}\right), & ln4 \leqslant t < ln8, \\ \left(\dfrac{1}{24}e^t + \dfrac{2}{3}, 2\right), & t = ln8. \end{cases}$$

At $t_0 = 0$, $t_1 = ln2$, $t_2 = ln4$, and $t_3 = ln8$, this motion coincides with the solution of the equivalent discrete system $z_0 = [2 \;\; 1]^\top$, $z_1 = [6 \;\; -67/12]^\top$, $z_2 = [5/6 \;\; -1/3]^\top$, $z_3 = [1 \;\; 2]^\top$. That is, the control law designed transfers system (14) in three steps from the initial state $z^{\{0\}}$ to the terminal one $z^{\{1\}}$ and, by Theorem 2, minimizes the cost functional (5).

4. THE COMPLETE CONTROLLABILITY OF SYSTEMS (1) AND (8) ON A CLOSED SUBSET OF THE EUCLIDEAN SPACE. FEATURES OF THE NONLINEARITY $\xi(Z,U,H)$

Consider the interconnection between the complete controllability of systems (1) and (8) for some $h=h_0>0$ on closed subsets of the Euclidean space R^{n+m}.

Let $S\subset R^{n+m}$ be the set of states of system (1). We introduce several notions as follows.

Definition 5. System (1) is said to be completely controllable on the set S for $h=h_0>0$ if all states $z^*\in S$ are reachable from any state $z^{\{0\}}\in S$.

Definition 6. A state $z^*\in S$ of system (1) for $h=h_0>0$ is said to be reachable from a state $z^{\{0\}}\in S$ if there exists an admissible control law $u(t_k)$, $k=\overline{0,l-1}$, such that the solution $z=z(t)$ of system (1) with the initial condition $z(t_0)=z^{\{0\}}$ satisfies the following conditions:
1) $z(t)\in S$ for $t\in[t_0,t_l]$; 2) $z(t_l)=z^*$, where $t_l=t_0+lh_0$.

Similar definitions can be formulated for the nonlinear discrete system (8), taking its specifics into account.

Assume that for $h=h_0>0$, system (8) is completely controllable on a set $K\subset R^{n+m}$, and system (1) is completely controllable on a set $S\subset R^{n+m}$.

Theorem 3. *If for $h=h_0>0$ systems (1) and (8) are completely controllable on sets S and K, respectively, then $K\subset S$.*

Let $U_0=S_0\times S^0$, where S_0 and S^0 are closed subsets of the neighborhoods of the points $z=0$ and $u=0$, respectively. In view of the features of the functions $f(x,y)$ and $g(x,y,u)$ in system (1), the next result can be proved for any $h>0$.

Theorem 4. *The nonlinearity $\xi(z,u,h)$ in system (8) satisfies the Lipschitz condition in the variables z, u on any closed subset U_0 of the neighborhood of the point $z=0$, $u=0$, i.e.,*

$$\left\|\xi\left(z_k^{(1)},u_k^{(1)},h\right)-\xi\left(z_k^{(2)},u_k^{(2)},h\right)\right\|\leqslant L\left(\left\|z_k^{(1)}-z_k^{(2)}\right\|+\left\|u_k^{(1)}-u_k^{(2)}\right\|\right), \tag{15}$$

where $(z_k^{(i)},u_k^{(i)})\in U_0$, $k=\overline{0,l-1}$, $i=1,2$; and the Lipschitz constant L is sufficiently small if the diameter of the subset U_0 is small.

5. THE EXISTENCE AND DESIGN OF AN ADMISSIBLE CONTROL LAW FOR SYSTEM (1)

Let system (10) be completely controllable. We pose the problem of designing an admissible control law $u(t_k)$, $k=\overline{0,l-1}$, transferring system (1) from a given initial state $z(t_0)=z^{\{0\}}\in R^{n+m}$ to a given terminal state $z(t_l)=z^{\{1\}}\in R^{n+m}$ and investigate the issue of the complete controllability of system (1) for $h=h_0>0$.

Let

$$m=\max\left\{\max_{0\le k\le l-1}\|A_0^{l-k-1}B\|\max_{0\le k\le l-1}\|S^\top(k)F^+(0)\|\max_{0\le k\le l-1}\|A_0^{l-k-1}\|,\right.$$
$$\left.\max_{0\le k\le l-1}\|A_0^{l-k-1}\|,\max_{0\le k\le l-1}\|S^\top(k)F^+(0)\|\max_{0\le k\le l-1}\|A_0^{l-k-1}\|\right\},$$

and let the function $\xi(z,u,h)$, $h=h_0>0$, in system (8) satisfy inequality (15) on the set $R^{n+m}\times R^q$, with a Lipschitz constant L such that

$$0<L<\frac{1}{ml(l+2)}. \tag{16}$$

The following result is true.

Theorem 5. *Assume that system* (10) *is completely controllable and the function* $\xi(z, u, h_0)$ *in system* (8) *satisfies the Lipschitz condition in the variables* z, u *on the set* $R^{n+m} \times R^q$*, with a Lipschitz constant obeying* (16)*. Then the hybrid system* (1) *is completely controllable for* $h = h_0$.

Consider again the closed subset $U_0 = S_0 \times S^0$ of the neighborhood of the point $z = 0$, $u = 0$. According to Section 4, we arrive at the following statement.

Corollary 1. *Assume that system* (10) *is completely controllable on the set* S_0 *and the Lipschitz constant for the function* $\xi(z, u, h_0)$ *of system* (8) *satisfies inequality* (16) *on the set* U_0*. Then for* $h = h_0$ *there exists an admissible program control law transferring the hybrid system* (1) *from any initial state* $z(t_0) = z^{\{0\}} \in S_0$ *to any terminal state* $z(t_l) = z^{\{1\}} \in S_0$.

Based on the proof of Theorem 5, we describe an algorithm for designing an admissible program control law for system (1) with $h = h_0 > 0$.

Algorithm 1.

1. Verify the complete controllability of system (10), i.e., check condition (11).

2. Using formulas (13), construct the optimal control law $u_k^{(0)}$, $k = \overline{0, l-1}$, for system (10) and the corresponding optimal motion $z_k^{(0)}$, $k = \overline{0, l}$.

3. Apply the simple iteration method to construct the sequences of optimal control actions $\{u_k^{(q)}\}$ and optimal motions $\{z_k^{(q)}\}$ of the systems:

$$z_{k+1}^{(q)} = A_0 z_k^{(q)} + B u_k^{(q)} + \xi\left(z_k^{(q-1)}, u_k^{(q-1)}, h_0\right), \quad k = \overline{0, l-1}, \quad q = 1, 2, \ldots,$$

where $z_k^{(0)}$ and $u_k^{(0)}$ are obtained at Step 2 above, using formulas (A.2)–(A.4) (see the Appendix).

4. Find the admissible program control law of system (1) with $h = h_0 > 0:$

$$u^*(t_k) = u_k^* = \lim_{q \to \infty} u_k^{(q)}, \quad k = \overline{0, l-1}.$$

Note that in a small neighborhood of the point $x = 0$, $y = 0$, $u = 0$, condition (16) surely holds for the function $\xi(z, u, h_0)$ (see Theorem 4). Therefore, for system (1) with $h = h_0 > 0$, it is always possible to design an admissible control law if the states $z^{\{0\}}$ and $z^{\{1\}}$ lie in a small neighborhood of the point $x = 0$, $y = 0$ and the condition of Step 1 of Algorithm 1 is valid.

Example 2. For the nonlinear hybrid system

$$\begin{cases} x'(t) = x(t) + 2y(t_k), \quad t_k \leqslant t < t_{k+1}, \\ y(t_{k+1}) = -x(t_{k+1}) + y(t_k) + 3u(t_k) + x(t_{k+1})y(t_k), \quad k = 0, 1, 2, \end{cases} \tag{17}$$

it is required to design an admissible program control law transferring this system in three steps ($l = 3$) from the initial state $z^{\{0\}} = [2 \;\; 1]^\top$ to the terminal one $z^{\{1\}} = [1 \;\; 2]^\top$.

The first approximation system of the equivalent nonlinear discrete system for $h_0 = ln2$ is completely controllable, and its optimal control law is given by $u_0^{(0)} = -7/36$, $u_1^{(0)} = 73/36$, $u_2^{(0)} = 10/9$. The corresponding motion of the system is $z_0^{(0)} = [2 \;\; 1]^\top$, $z_1^{(0)} = [6 \;\; -67/12]^\top$, $z_2^{(0)} = [5/6 \;\; -1/3]^\top$, $z_3^{(0)} = [1 \;\; 2]^\top$ (see Example 1).

The discrete system equivalent to (17) for $h = h_0$ has the form

$$z_{k+1} = A_0 z_k + B u_k + \xi(z_k, u_k, h_0), \quad k = 0, 1, 2,$$

where

$$A_0 = \begin{bmatrix} 2 & 2 \\ -2 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad \xi(z_k, u_k, h_0) = \begin{bmatrix} 0 \\ 2x_k y_k + 2y_k^2 \end{bmatrix}.$$

We obtain the following result by applying Algorithm 1: on the 49th iteration with accuracy up to 10^{-15}, the iterative process converges to the control law

$$u^*(t_0) = -1.421310674166737, \quad u^*(t_1) = 7.208203932499369, \quad u^*(t_2) = 4.314757303333053.$$

For the sake of comparison, note that when transferring system (17) with $h = h_0$ from the state $z^{\{0\}} = [0.1 \ \ -0.02]^\top$ to the state $z^{\{1\}} = [0.03 \ \ 0.2]^\top$ in three steps, the iterative process with accuracy up to 10^{-15} converges on the 11th iteration. That is, if the states $z^{\{0\}}$ and $z^{\{1\}}$ of system (17) lie in a neighborhood of the point $x = 0$, $y = 0$, the iterative process will converge faster.

However, when transferring system (17) with $h = h_0$ from the state $z^{\{0\}} = [-2 \ \ 5]^\top$ to the state $z^{\{1\}} = [3 \ \ 4]^\top$ in three steps, the iterative process diverges.

Consider the existence and design of an admissible program control law for system (1) when the conditions of Theorem 5 fail. For this purpose, we introduce an auxiliary system of the form

$$z_{k+1} = A_0 z_k + B u_k, \quad k = 0, 1, 2, \dots, \tag{18}$$

differing from (10) in the range of k, assuming that the former system is completely controllable [7]. This is true if system (10) is completely controllable [26].

6. THE EXISTENCE AND DESIGN OF AN ADMISSIBLE CONTROL LAW FOR SYSTEM (1) CONSIDERING STABILIZATION OF SYSTEM (18)

According to [10], the iterative process of designing an admissible control law for system (8) with $h = h_0 > 0$ is convergent if the function $\xi(z, u, h_0)$ satisfies the global Cauchy–Lipschitz condition in the variables z, u with a sufficiently small Lipschitz constant L. When this requirement fails, there may arise rapidly growing motions tending to infinity at a high rate; in this case, the iterative process will diverge, which is confirmed by the analysis of system (17) being transferred from the state $z^{\{0\}} = [-2 \ \ 5]^\top$ to the state $z^{\{1\}} = [3 \ \ 4]^\top$ in three steps.

One way to improve the convergence of the iterative process is the preliminary stabilization [27] of the system defined by (8) with $h = h_0 > 0$, or, equivalently, of the system defined by (1) with $h = h_0 > 0$, assuming $k = 0, 1, 2, \dots$ in these systems [7]. Any of these systems is stabilized by stabilizing the completely controllable system (18), as the three systems mentioned have the same stabilizing control law [7].

Let us design a stabilizing control law $w(z_k) = D z_k$, $k = 0, 1, 2, \dots$, where D is a constant matrix of dimensions $q \times (n + m)$, for system (18), under which the initial $z(t_0) = z^{\{0\}}$ and terminal $z(t_l) = z^{\{1\}}$ states will fall into the attraction domain of the equilibrium $z = 0$ of the system

$$z_{k+1} = A z_k + B v_k + \xi(z_k, D z_k + v_k, h_0), \quad k = \overline{0, l-1}, \tag{19}$$

with $v_k \equiv 0$, where $A = A_0 + BD$ and $v_k \in R^q$ is the control input. This ensures the existence of an admissible control law for system (8) with $h = h_0$ [10].

Assume that the function $\xi(z, Dz + v, h_0)$ in system (19) satisfies the Lipschitz condition in the variables z, $Dz + v$ on the set $R^{n+m} \times R^q$. If an estimate of the Lipschitz constant L for the function $\xi(z, Dz + v, h_0)$ on $R^{n+m} \times R^q$ is found (e.g., see [28]), we compare it with the condition

$$0 < L < \frac{1}{ml(1 + (\|D\| + 1)(l + 1))}, \tag{20}$$

where

$$m = \max\left\{\max_{0\le k\le l-1}\|S^\top(k)F^+(0)\|\max_{0\le k\le l-1}\|A^{l-k-1}\|,\max_{0\le k\le l-1}\|A^{l-k-1}\|,\right.$$

$$\left.\max_{0\le k\le l-1}\|S^\top(k)F^+(0)\|\max_{0\le k\le l-1}\|A^{l-k-1}\|\max_{0\le k\le l-1}\|A^{l-k-1}B\|\right\},$$

$$S(k) = A^{l-k-1}B.$$

These considerations are based on the following result.

Theorem 6. *Assume that system* (18) *is completely controllable and the stabilizing control law* $w(z_k) = Dz_k$ *of system* (18) *is chosen so that the Lipschitz constant for the function* $\xi(z, Dz+v, h_0)$ *in system* (19) *satisfies condition* (20) *for* $(z, Dz+v) \in R^{n+m} \times R^q$. *Then the hybrid system* (1) *is completely controllable for* $h = h_0$, *and an admissible control* $u^*(t_k)$ *transferring it from any initial state* $z(t_0) = z^{\{0\}}$ *to any terminal state* $z(t_l) = z^{\{1\}}$ *is given by*

$$u^*(t_k) = w(z^*(t_k)) + v^*(t_k), \quad k = \overline{0, l-1},$$

where $v^*(t_k) = v_k^*$ *and* $z^*(t_k) = z_k^*$ *are an admissible control law and the corresponding motion of system* (19), *respectively.*

In this case, according to the proof of Theorem 6, we present the following algorithm for designing an admissible control law for system (1) with $h = h_0 > 0$.

Algorithm 2.

1. Verify the complete controllability of system (18) and design the stabilizing control law $w(z_k) = Dz_k$ for this system.

2. Using formulas (13), construct the optimal control law $v_k^{(0)}$, $k = \overline{0, l-1}$, for the system $z_{k+1} = Az_k + Bv_k$, where $A = A_0 + BD$, $k = \overline{0, l-1}$, and the corresponding optimal motion $z_k^{(0)}$, $k = \overline{0, l}$.

3. Apply the simple iteration method to construct the sequences of optimal control actions $\{v_k^{(q)}\}$ and optimal motions $\{z_k^{(q)}\}$ of the systems $z_{k+1}^{(q)} = Az_k^{(q)} + Bv_k^{(q)} + \xi\left(z_k^{(q-1)}, Dz_k^{(q-1)} + v_k^{(q-1)}, h_0\right)$, $k = \overline{0, l-1}$, $q = 1, 2, \ldots$, where $z_k^{(0)}$ and $v_k^{(0)}$ are obtained at Step 2 above, using formulas (A.12)–(A.14) (see the Appendix).

4. Find $v^*(t_k) = \lim_{q\to\infty} v_k^{(q)}$, $k = \overline{0, l-1}$; $z_k^* = \lim_{q\to\infty} z_k^{(q)}$, $k = \overline{0, l}$.

5. Design the admissible program control law of system (1) with $h = h_0 > 0$: by formula $u^*(t_k) = w(z_k^*) + v^*(t_k)$, $k = \overline{0, l-1}$.

Example 3. a) It is required to transfer system (17) with $h_0 = ln2$ from the state $z(t_0) = z^{\{0\}} = [-2\ \ 5]^\top$ to the state $z(t_3) = z^{\{1\}} = [3\ \ 4]^\top$ (see Example 2).

Recall that Algorithm 1 has failed to solve this problem. Now we apply Algorithm 2 and design, for the corresponding system (18), a stabilizing control law such that the matrix A has eigenvalues $\lambda_{1,2} = 0.001$. Then, on the 27th iteration with accuracy up to 10^{-15}, the admissible control law of system (17) becomes

$u_0^* = -2.459959701325314$, $u_1^* = -404.806942255234989$, $u_2^* = 73.986989056064154$.

Note that for $\lambda_{1,2} = 0.0001$, the iterative process with accuracy up to 10^{-15} will converge on the 12th iteration.

b) Consider the nonlinear hybrid system

$$\begin{cases} x'(t) = x(t) - y(t_k) + y^2(t_k), \quad t_k \leqslant t < t_{k+1}, \\ y(t_{k+1}) = -x(t_{k+1}) - y(t_k) + 2u(t_k) + u(t_k)y(t_k), \quad k = 0, 1, 2, 3, 4. \end{cases} \tag{21}$$

This system for $h_0 = ln\frac{3}{2}$ is equivalent to the nonlinear discrete system

$$\begin{cases} x_{k+1} = \dfrac{3}{2}x_k - \dfrac{1}{2}y_k + \dfrac{1}{2}y_k^2 \\ y_{k+1} = -\dfrac{3}{2}x_k - \dfrac{1}{2}y_k + 2u_k - \dfrac{1}{2}y_k^2 + u_k y_k, \quad k = 0, 1, 2, 3, 4. \end{cases}$$

The first approximation system for this discrete system is completely controllable. Algorithm 1 will not help to transfer system (21) from the state $z(t_0) = z^{\{0\}} = [0.1 \ \ -0.2]^\top$ to the state $z(t_5) = z^{\{1\}} = [0.4 \ \ 0.6]^\top$. Application of Algorithm 2 leads to the following results:

1) For $\lambda_1 = 0.1$ and $\lambda_2 = 0.2$, the iterative process converges on the 80th iteration, and with accuracy up to 10^{-18} we obtain the admissible control law

$u^*(t_0) = 0.324385159473078927$, $u^*(t_1) = 0.501724268550236387$,
$u^*(t_2) = 0.502341831544914188$, $u^*(t_3) = 0.399479636515260206$,
$u^*(t_4) = 0.544220425724104726$.

2) For $\lambda_1 = 0.2$ and $\lambda_2 = 0.3$, the iterative process diverges.

3) For $\lambda_1 = 0.1$ and $\lambda_2 = 0.3$, the iterative process with accuracy up to 10^{-18} converges on the 88th iteration.

4) For $\lambda_1 = 0.01$ and $\lambda_2 = 0.1$, as well as for $\lambda_1 = 0.01$ and $\lambda_2 = -0.001$, the iterative process diverges, meaning that the matrices D corresponding to the above $\lambda_{1,2}$ violate condition (20).

Thus, in particular cases (see Example 3a)), stabilization of the corresponding system (18) with $\lambda_{1,2}$ being near zero accelerates the convergence of the iterative process; however, in the general case, this is false.

7. SUFFICIENT CONDITIONS FOR OPTIMAL CONTROL OF THE HYBRID SYSTEM (1)

The algorithms presented in Sections 5 and 6 can be used to design an admissible control law for system (1). The following questions arise accordingly. Is an admissible control law optimal? What are the sufficient conditions for optimal control of system (1)? How to design an optimal control law for such a system?

To answer these questions, we will assume the complete controllability of system (1) for some $h = h_0 > 0$ and, due to the equivalence of systems (1) and (8), utilize sufficient optimality conditions for controlled discrete processes [19], with an appropriate modification to the problem of finding an optimal process with a fixed right endpoint of the trajectory.

Let $M = \{(z_k, u_k) : k = \overline{0, l-1}\}$ be the set of *admissible processes* of system (8) with $h = h_0 > 0$, i.e., pairs (z_k, u_k) satisfying, for $h = h_0 > 0$, this system and the boundary conditions $z_0 = z^{\{0\}}$, $z_l = z^{\{1\}}$.

Assume that the performance of a control process of system (8) with $h = h_0 > 0$ is generally evaluated by the cost functional

$$I(z_k, u_k) = \sum_{k=0}^{l-1} f^0(z_k, u_k), \tag{22}$$

where the function $f^0(z, u)$ is differentiable in z and u on the set M.

A process $(z_k, u_k) \in M$ is said to be *optimal* if $I(z_k, u_k) \to \min$.

By denoting $\tilde{f}(z_k, u_k) = A_0 z_k + B u_k + \xi(z_k, u_k, h_0)$, we write system (8) with $h = h_0 > 0$ as $z_{k+1} = \tilde{f}(z_k, u_k)$, $k = \overline{0, l-1}$, where the function $\tilde{f}(z, u)$ is continuously differentiable in z and u on the set $R^{n+m} \times R^q$.

Let D be the set of *possible processes* [19] of system $z_{k+1} = \tilde{f}(z_k, u_k)$, $k = \overline{0, l-1}$, i.e., pairs (z_k, u_k) satisfying this system; then $M \subset D$. Following [19], we introduce auxiliary mathematical constructs:

1) the function

$$R(z_k, u_k) = \varphi(\tilde{f}(z_k, u_k)) - \varphi(z_k) - f^0(z_k, u_k), \tag{23}$$

where $\varphi(z)$ is some function;

2) the functional

$$L(z_k, u_k, \varphi) = -\sum_{k=0}^{l-1} R(z_k, u_k) + \varphi(z^{\{1\}}) - \varphi(z^{\{0\}}). \tag{24}$$

Consider several auxiliary statements.

Lemma 1. *For any function $\varphi(z)$, the values of the cost functional* (22) *and functional* (24) *coincide on the set M, i.e., $L(z_k, u_k, \varphi) = I(z_k, u_k)$ for $(z_k, u_k) \in M$.*

Lemma 2. *If an admissible process $(z_k^*, u_k^*) \in M$ and some function $\varphi(z)$ satisfy the condition*

$$R(z_k^*, u_k^*) = \max_{(z_k, u_k) \in D} R(z_k, u_k), \quad k = \overline{0, l-1}, \tag{25}$$

then

$$L(z_k^*, u_k^*, \varphi) = \min_{(z_k, u_k) \in D} L(z_k, u_k, \varphi), \quad k = \overline{0, l-1}.$$

Let the cost functional (22) take the form (5), i.e., $I(z_k, u_k) = I(u_k) = \sum_{k=0}^{l-1} u_k^\top u_k$. We arrive at the following result.

Theorem 7. *Assume that an admissible process $(z_k^*, u_k^*) \in M$ and some function $\varphi(z)$ satisfy condition* (25)*, with $f^0(z_k, u_k) = u_k^\top u_k$. Then $u^*(t_k) = u_k^*$ is the optimal control law of system* (1) *with $h = h_0$.*

The optimization problem of the cost functional can be formulated in two settings [19]:

1) It is required to determine the optimal process (z_k^*, u_k^*) minimizing the cost functional (22); in this setting, if $f^0(z_k, u_k) = u_k^\top u_k$, then $u^*(t_k) = u_k^*$ is the optimal control law of system (1) with $h = h_0 > 0$.

2) It is required to determine a minimizing sequence of admissible processes $\{z_k^{(s)}, u_k^{(s)}\} \subset M$, $s = 1, 2, \ldots$, on which the cost functional (22) tends to its greatest lower bound, i.e.,

$$I\left(z_k^{(s)}, u_k^{(s)}\right) \to \inf_{(z_k, u_k) \in M} I(z_k, u_k) \quad \text{as} \quad s \to \infty;$$

in this setting, if $f^0(z_k, u_k) = u_k^\top u_k$, then $\{u^{(s)}(t_k)\}$ is a minimizing sequence of admissible control actions of system (1) with $h = h_0 > 0$ on which the cost functional (5) tends to its greatest lower bound.

As a rule, the optimal control problem is formulated in the first setting, but does not always have a solution since the cost functional (22) may generally possess no minimum on the set M. Meanwhile, the problem in the second setting always has a solution if the cost functional (22) is bounded below [19].

The cost functional (5) is nonnegative, i.e., bounded below. Hence, there exists a minimizing sequence of admissible control actions of system (1) for the cost functional (5). And the following result is valid.

Theorem 8. *Let* $\{z_k^{(s)}, u_k^{(s)}\} \subset M$, $s = 1, 2, \ldots$, *be a sequence of admissible processes of the system* $z_{k+1} = \tilde{f}(z_k, u_k)$, $k = \overline{0, l-1}$. *Assume the existence of a function* $\varphi(z)$ *such that, as* $s \to \infty$,

$$R\left(z_k^{(s)}, u_k^{(s)}\right) \to \sup_{(z_k, u_k) \in D} R(z_k, u_k), \quad k = \overline{0, l-1},$$

with $f^0(z_k, u_k) = u_k^\top u_k$. *Then* $\{u^{(s)}(t_k)\}$ *is a minimizing sequence of admissible control actions of system* (1) *with* $h = h_0$ *on which the cost functional* (5) *tends to its greatest lower bound.*

Suppose that system (1) with $h = h_0 > 0$ is known to have an optimal control. Let us study the issue of designing such a control.

8. OPTIMAL CONTROL DESIGN FOR THE HYBRID SYSTEM (1)

Due to the equivalence of systems (1) and (8), let us pose the following problem: for the system $z_{k+1} = \tilde{f}(z_k, u_k)$, $k = \overline{0, l-1}$, it is required to find an admissible process (z_k^*, u_k^*) on which the cost functional (22) (or the corresponding functional (5)) is minimized without imposing additional constraints on the control input $u_k \in R^q$.

To solve this problem, we use Lagrange's method of multipliers for discrete control processes with a one-dimensional (scalar) argument and an unbounded control input [19] as well as the sufficient optimality conditions provided by Theorem 7.

Let $\varphi(z)$ be a function differentiable in z on the set R^{n+m}. Then the necessary conditions for the maximum of the function $R(z, u)$ in z, u are given by

$$\frac{\partial R(z_k^*, u_k^*)}{\partial z_{k,i}} = 0, \quad i = \overline{1, n+m}, \quad k = \overline{0, l-1}, \tag{26}$$

$$\frac{\partial R(z_k^*, u_k^*)}{\partial u_{k,j}} = 0, \quad j = \overline{1, q}, \quad k = \overline{0, l-1}. \tag{27}$$

Let a vector function $\psi(k) = (\psi_1(k), \psi_2(k), \ldots, \psi_{n+m}(k))$, $k = \overline{0, l-1}$, be such that

$$\psi_i(k) = \frac{\partial \varphi(z_k^*)}{\partial z_{k,i}}, \quad i = \overline{1, n+m}.$$

We compile the Hamiltonian of the problem:

$$H(\psi(k), z_k, u_k) = \sum_{i=1}^{n+m} \psi_i(k) \tilde{f}_i(z_k, u_k) - f^0(z_k, u_k).$$

In view of (23), the equality $z_{k+1} = \tilde{f}(z_k, u_k)$, and the last two equalities above, it can be shown that conditions (26) and (27) take the form

$$\psi_i(k) = \frac{\partial H(\psi(k+1), z_k^*, u_k^*)}{\partial z_{k,i}}, \quad i = \overline{1, n+m}, \tag{28}$$

and

$$\frac{\partial H(\psi(k+1), z_k^*, u_k^*)}{\partial u_{k,j}} = 0, \quad j = \overline{1, q}, \tag{29}$$

respectively. Equalities (28) and (29) are necessary optimality conditions for discrete control processes with a scalar argument [19], with equality (28) called the adjoint system of finite-difference equations [29].

These conditions allow selecting, from the set of all admissible processes of the system $z_{k+1} = \tilde{f}(z_k, u_k)$, $k = \overline{0, l-1}$, a subset of potentially optimal processes [19]. If an optimal process is known to exist, and the resulting subset consists of a single element, it will be the desired optimal process. In this case, we obtain the following algorithm for designing an optimal control law for system (1) with $h = h_0 > 0$.

Algorithm 3.

1. Perform the transition from system (1) with $h = h_0 > 0$ to the system $z_{k+1} = \tilde{f}(z_k, u_k)$, $k = \overline{0, l-1}$.
2. Compile the Hamiltonian of the problem and conditions (29).
3. Find from conditions (29) the control law u_k^* with the parameters $\psi(k+1)$ and z_k^*, $k = \overline{0, l-1}$.
4. Compile the adjoint system (28).
5. Find the solutions $z_1^*, z_2^*, \ldots, z_{l-1}^*$ of the system $z_{k+1} = \tilde{f}(z_k, u_k)$ considering the condition $z_0^* = z^{\{0\}}$ and the control law $u_k^* = \{u_{k,j}^*(\psi(k+1), z_k^*)\}$, $k = \overline{0, l-1}$, $j = \overline{1, q}$.
6. Solve the adjoint system together with the system $z_l^* = z^{\{1\}}$, considering the solutions z_1^*, $z_2^*, \ldots, z_{l-1}^*$, and find $\psi_i(k+1)$, z_k^*, and u_k^*, $i = \overline{1, n+m}$, $k = \overline{0, l-1}$.
7. If the resulting solution is unique, then $u^*(t_k) = u_k^*$ is the optimal control law of system (1) with $h = h_0$; otherwise, find the value of the cost functional (5) on the found solutions and choose the control law corresponding to the solution that minimizes this functional.

Example 4. It is required to design an optimal control law that transfers system (17) with a sampling step of $h_0 = ln2$ from the initial state $z(t_0) = [2 \;\; 1]^\top$ to the terminal one $z(t_3) = [1 \;\; 2]^\top$ and minimizes the cost functional (5).

Recall that (Example 2) system (17) with $h_0 = ln2$ is equivalent to the nonlinear discrete system

$$\begin{cases} x_{k+1} = 2x_k + 2y_k \\ y_{k+1} = -2x_k - y_k + 3u_k + 2x_k y_k + 2y_k^2, \quad k = 0, 1, 2, \end{cases} \tag{30}$$

and the problem is to design an optimal control law transferring system (30) from the state $z_0 = [x_0 \;\; y_0]^\top = [2 \;\; 1]^\top$ to the state $z_3 = [x_3 \;\; y_3]^\top = [1 \;\; 2]^\top$.

According to (5), $f^0(z_k, y_k) = u_k^2$ since $u_k \in R^1$, and the set of admissible processes M for system (30) is compact. Therefore, the posed problem has a solution as the continuous cost functional (5) will reach its minimum value on M.

Note that $z_{k,1} = x_k$ and $z_{k,2} = y_k$; the Hamiltonian of the problem takes the form

$$H(\psi(k), z_k, u_k) = \psi_1(k)(2x_k + 2y_k) + \psi_2(k)(-2x_k - y_k + 3u_k + 2x_k y_k + 2y_k^2) - u_k^2,$$

and condition (29) yields

$$u_k^* = \frac{3}{2}\psi_2(k+1), \quad k = 0, 1, 2. \tag{31}$$

Due to (28), the adjoint system is given by

$$\begin{cases} \psi_1(k) = 2\psi_1(k+1) + 2\psi_2(k+1)(y_k^* - 1) \\ \psi_2(k) = 2\psi_1(k+1) + \psi_2(k+1)(4y_k^* + 2x_k^* - 1), \quad k = 0, 1, 2. \end{cases} \tag{32}$$

Considering the first boundary-value condition as well as equalities (30) and (31), we find the solutions

$$\begin{cases} x_1^* = 6 \\ y_1^* = 1 + 4.5\psi_2(1), \end{cases} \qquad \begin{cases} x_2^* = 14 + 9\psi_2(1) \\ y_2^* = 40.5\psi_2^2(1) + 67.5\psi_2(1) + 4.5\psi_2(2) + 1. \end{cases}$$

Then the system $z_3^* = [1 \;\; 2]^\top$ turns into

$$\begin{cases} 29 + 153\psi_2(1) + 81\psi_2^2(1) + 9\psi_2(2) = 0 \\ 729\psi_2^3(1) + 2308.5\psi_2^2(1) + 1822.5\psi_2(1) + 121.5\psi_2(2) + 4.5\psi_2(3) + 81\psi_2(1)\psi_2(2) + 2y_2^{*2} = 3. \end{cases}$$

Based on the first equality in (32), (31), and the obtained solutions, we find

$$\psi_1(2) = 2\psi_1(3) + 2\psi_2(3)(40.5\psi_2^2(1) + 67.5\psi_2(1) + 4.5\psi_2(2)); \tag{33}$$

and based on the second equality in (32), $\psi_2(1) = 2\psi_1(2) + \psi_2(2)(15 + 18\psi_2(1))$. Then

$$\psi_1(2) = 0.5\psi_2(1) - \psi_2(2)(7.5 + 9\psi_2(1)) \tag{34}$$

and $\psi_2(2) = 2\psi_1(3) + \psi_2(3)(162\psi_2^2(1) + 288\psi_2(1) + 18\psi_2(2) + 31)$, consequently,

$$\psi_1(3) = 0.5\psi_2(2) - \psi_2(3)(81\psi_2^2(1) + 153\psi_2(1) + 15.5). \tag{35}$$

Let us substitute conditions (34) and (35) into (33). In view of the first equation in the system $z_3^* = [1 \;\; 2]^\top$, we get

$$\psi_2(2)(8.5 + 9\psi_2(1)) - 0.5\psi_2(1) = 2\psi_2(3). \tag{36}$$

Solving (36) together with the system $z_3^* = [1 \;\; 2]^\top$ yields three solutions and, accordingly, three control laws satisfying equality (31), system (30), and the boundary value conditions. Only one of them is optimal, particularly for system (17) with $h_0 = ln2$, as it minimizes the cost functional (5). With accuracy up to 10^{-15}, the optimal control law is given by $u^*(t_0) = -2.498184215120807$, $u^*(t_1) = 0.190252087647811$, $u^*(t_2) = 0.007263139516772$.

9. CONCLUSIONS

In this paper, we have established sufficient conditions for the complete controllability and existence of admissible and optimal control laws for nonlinear continuous–discrete dynamic systems with a constant sampling step. Algorithms for designing admissible and optimal program control laws for the above systems on a finite horizon have been presented. The effectiveness of these algorithms has been confirmed by numerical examples.

The conditions and algorithms will serve for control and stable operation of real systems, most topical in aviation, engineering, and economics.

APPENDIX

Proof of Theorem 1. Systems (1) and (8) are equivalent, and there exists a connection between their solutions [7, p. 867]. Based on this connection, the desired result follows from Definitions 1 and 4.

The proof of Theorem 1 is complete.

Proof of Theorem 2. According to [11], by the complete controllability of system (10), there exists the optimal control law (13) transferring system (10) from the state $z_0 = z^{\{0\}}$ to the state $z_l = z^{\{1\}}$ and minimizing the cost functional (5). And since systems (10) and (12) are equivalent and have the same control performance index, the optimal control of system (10) will also be optimal for system (12).

The proof of Theorem 2 is complete.

Proof of Theorem 3. In view of Definitions 5 and 6 and the equivalence of systems (1) and (8), the sets S and K can be described as $S = \{(x(t), y_k) : t \in [t_k, t_{k+1}),\ k = \overline{0, l-1}\} \cup \{(x(t_l), y_l)\}$, where $x(t_k) = x_k$, and $K = \{(x_k, y_k) : k = \overline{0, l}\}$. Consequently, $K \subset S$, and the proof of Theorem 3 is complete.

Proof of Theorem 4. The smooth nonlinearities $a(x, y)$ and $b(x, y, u)$ in the expansions of the functions $f(x, y)$ and $g(x, y, u)$ start with the quadratic terms in x, y and in x, y, u, respectively. Therefore, on the sets S_0 and U_0 of small diameter, they satisfy the Lipschitz condition with sufficiently small Lipschitz constants L_1 and L_2, respectively. The function $\xi(z, u, h)$ is continuously differentiable on the set U_0, and condition (15) holds for it. Since the components $\varepsilon(x, y; h)$ and $\delta(x, y, u; h)$ of the function $\xi(z, u, h)$ are linearly expressed through the functions $a(x, y)$ and $b(x, y, u)$ (see (6) and (7)), the Lipschitz constant L in condition (15) will be expressed through the constants L_1 and L_2 and will be sufficiently small.

The proof of Theorem 4 is complete.

Proof of Theorem 5. According to [11], by the complete controllability of system (10), there exists the optimal control law $u_k^{(0)}$ (13) transferring system (10) from $z_0 = z^{\{0\}} \in R^{n+m}$ to any state $z_l = z^{\{1\}} \in R^{n+m}$ and minimizing the cost functional (5). Then $z_k^{(0)}$ $(k = \overline{0, l})$ is the optimal motion of system (10) generated by the control law $u_k^{(0)}$, $k = \overline{0, l-1}$.

We design an admissible control law for system (8) with $h = h_0$ by the simple iteration method applied to solving the optimal control problems

$$z_{k+1}^{(q)} = A_0 z_k^{(q)} + B u_k^{(q)} + \xi\left(z_k^{(q-1)}, u_k^{(q-1)}, h_0\right), \quad k = \overline{0, l-1}, \quad q = 1, 2, \ldots, \tag{A.1}$$

$$I(u^{(q)}) = \sum_{k=0}^{l-1} u_k^{(q)\top} u_k^{(q)} \to \min,$$

intended to transfer systems (A.1) from the state $z_0 = z^{\{0\}}$ to the state $z_l = z^{\{1\}}$, where $z_k^{(0)}$ and $u_k^{(0)}$ are the optimal motion and control law of system (10), respectively. Then, using [11], we obtain a sequence of optimal controls $\{u_k^{(q)}\}$, $\quad q = 1, 2, \ldots$, given by the formulas

$$u_k^{(q)} = S^\top(k) F^+(0) d^{(q)}\left(0, z^{\{0\}}\right), \quad k = \overline{0, l-1}, \tag{A.2}$$

$$d^{(q)}\left(0, z^{\{0\}}\right) = z^{\{1\}} - A_0^l z^{\{0\}} - \sum_{i=0}^{l-1} A_0^{l-i-1} \xi\left(z_i^{(q-1)}, u_i^{(q-1)}, h_0\right), \tag{A.3}$$

and the corresponding sequence of optimal motions $\{z_k^{(q)}\}$, $\quad q = 1, 2, \ldots$,

$$z_k^{(q)} = A_0^k z^{\{0\}} + \sum_{i=0}^{k-1} A_0^{k-i-1} B u_i^{(q)} + \sum_{i=0}^{k-1} A_0^{k-i-1} \xi\left(z_i^{(q-1)}, u_i^{(q-1)}, h_0\right), \quad k = \overline{0, l}, \tag{A.4}$$

with $z_0^{(q)} = z^{\{0\}}$ and $z_l^{(q)} = z^{\{1\}}$.

Indeed, from (A.4) it follows that $z_0^{(q)} = z^{\{0\}}$, and based on [11, pp. 13–14], the control law $u_k^{(q)}$ transfers systems (A.1) from the state $z_0 = z^{\{0\}}$ to the state $z_l = z^{\{1\}}$ if and only if $\sum_{k=0}^{l-1} S(k) u_k^{(q)} = d^{(q)}(0, z^{\{0\}})$, where $S(k) = A_0^{l-k-1} B$. Then, taking (A.4) and (A.3) into account, we get $z_l^{(q)} = z^{\{1\}}$.

To prove the convergence of the iterative process, let us treat the pair of functions $\{z_k^{(q)}, u_k^{(q)}\}$ as an element of the Euclidean space R^{n+m+q} with the norm

$$\rho\left(\left\{z_k^{(q)}, u_k^{(q)}\right\}\right) = \max_{0 \le k \le l} \left\|z_k^{(q)}\right\| + \max_{0 \le k \le l-1} \left\|u_k^{(q)}\right\|. \tag{A.5}$$

By utilizing (A.2), (A.3), the properties of the norm, inequality (15), and (A.5), we derive the upper bound

$$\max_{0\le k\le l-1}\left\|u_k^{(q)}-u_k^{(q-1)}\right\|$$
$$\le lL\max_{0\le k\le l-1}\left\|S^\top(k)F^+(0)\right\|\max_{0\le k\le l-1}\left\|A_0^{l-k-1}\right\|\rho\left(\left\{z_k^{(q-1)},u_k^{(q-1)}\right\}-\left\{z_k^{(q-2)},u_k^{(q-2)}\right\}\right).$$

Similarly, by utilizing (A.4), the properties of the norm, inequality (15), the above estimate, and (A.5), we obtain the upper bound

$$\max_{0\le k\le l}\left\|z_k^{(q)}-z_k^{(q-1)}\right\|$$
$$\le lL\left(l\max_{0\le k\le l-1}\left\|A_0^{l-k-1}B\right\|\max_{0\le k\le l-1}\left\|S^\top(k)F^+(0)\right\|\max_{0\le k\le l-1}\left\|A_0^{l-k-1}\right\|+\max_{0\le k\le l-1}\left\|A_0^{l-k-1}\right\|\right)$$
$$\times\rho\left(\left\{z_k^{(q-1)},u_k^{(q-1)}\right\}-\left\{z_k^{(q-2)},u_k^{(q-2)}\right\}\right).$$

Let

$$m=\max\Big\{\max_{0\le k\le l-1}\|A_0^{l-k-1}B\|\max_{0\le k\le l-1}\|S^\top(k)F^+(0)\|\max_{0\le k\le l-1}\|A_0^{l-k-1}\|,$$
$$\max_{0\le k\le l-1}\|A_0^{l-k-1}\|,\max_{0\le k\le l-1}\|S^\top(k)F^+(0)\|\max_{0\le k\le l-1}\|A_0^{l-k-1}\|\Big\}.$$

Summing these bounds, according to (A.5), yields

$$\rho\left(\left\{z_k^{(q)},u_k^{(q)}\right\}-\left\{z_k^{(q-1)},u_k^{(q-1)}\right\}\right)\le lLm(l+2)\rho\left(\left\{z_k^{(q-1)},u_k^{(q-1)}\right\}-\left\{z_k^{(q-2)},u_k^{(q-2)}\right\}\right). \quad \text{(A.6)}$$

Due to condition (16), we have $0<lLm(l+2)<1$; then the sequence of vector functions $\{z_k^{(q)},u_k^{(q)}\}$ converges, uniformly in k, to some vector function $\{z_k^*,u_k^*\}$, and the operator $F(z,u):R^{n+m+q}\to R^{n+m+q}$ (A.2)–(A.4) is a contraction. According to the principle of contractive mappings, it has a unique fixed point, i.e., the limit vector function $\{z_k^*,u_k^*\}$ with

$$z_k^*=\lim_{q\to\infty}z_k^{(q)},\quad u_k^*=\lim_{q\to\infty}u_k^{(q)}, \quad \text{(A.7)}$$

$$u_k^*=S^T(k)F^+(0)\left(z^{\{1\}}-A_0^lz^{\{0\}}-\sum_{i=0}^{l-1}A_0^{l-i-1}\xi(z_i^*,u_i^*,h_0)\right),\quad k=\overline{0,l-1}, \quad \text{(A.8)}$$

$$z_k^*=A_0^kz^{\{0\}}+\sum_{i=0}^{k-1}A_0^{k-i-1}Bu_i^*+\sum_{i=0}^{k-1}A_0^{k-i-1}\xi(z_i^*,u_i^*,h_0),\quad k=\overline{0,l}. \quad \text{(A.9)}$$

In addition, (A.7) implies $z_0^*=z_0$ and $z_l^*=z_l$.

Direct substitution of (A.8) with $k=0,1,\ldots,l-1$ into (A.9) with $k=l$ shows that $z_l^*=z_l$ if the product $F(0)F^+(0)$ acts as an identity matrix, which is true because $\sum_{k=0}^{l-1}S(k)u_k^{(q)}=d^{(q)}(0,z^{\{0\}})$.

Due to the equivalence of systems (1) and (8), the admissible control law $u_k^*=u^*(t_k)$ for system (8) with $h=h_0$ is also an admissible control law for system (1) with $h=h_0$. For any states $z^{\{0\}}$, $z^{\{1\}}\in R^{n+m}$, one can design an admissible control law for system (8) with $h=h_0$; therefore, system (8) with $h=h_0$ is completely controllable. By Theorem 1, system (1) is completely controllable for $h=h_0$, and the proof of Theorem 5 is finished.

Proof of Theorem 6. The complete controllability of system (18) implies its stabilizability [7]. Let us choose a stabilizing control law $w(z_k) = Dz_k$ $(k = 0, 1, 2, \dots)$ for system (18) so that all eigenvalues of the matrix $A = A_0 + BD$ lie inside the unit circle [7]. Then the linear system

$$z_{k+1} = Az_k + Bv_k, \quad k = \overline{0, l-1}, \tag{A.10}$$

is completely controllable [10, 13].

By employing formulas (13) of Theorem 2, we design an optimal control law $v_k^{(0)}$ transferring the linear system (A.10) from any state $z_0 = z^{\{0\}}$ to any state $z_l = z^{\{1\}}$ and ensuring $I(v^{(0)}) \to \min$, where $I(v^{(0)})$ is the cost functional (5) of the control variable $v_k^{(0)}$; then $z_k^{(0)}$ is the corresponding optimal motion. Next, we design an admissible control law transferring system (19) from the state $z_0 = z^{\{0\}}$ to the state $z_l = z^{\{1\}}$ by the simple iteration method applied to solving the optimal control problems

$$\begin{gathered} z_{k+1}^{(q)} = Az_k^{(q)} + Bv_k^{(q)} + \xi\left(z_k^{(q-1)}, Dz_k^{(q-1)} + v_k^{(q-1)}, h_0\right), \quad k = \overline{0, l-1}, \quad q = 1, 2, \dots, \\ I(v^{(q)}) = \sum_{k=0}^{l-1} v_k^{(q)\top} v_k^{(q)} \to \min, \end{gathered} \tag{A.11}$$

where $z_k^{(0)}$ and $v_k^{(0)}$ are the optimal motion and control law of system (A.10), respectively.

As a result, we obtain the following sequences of optimal control actions $\{v_k^{(q)}\}$ and optimal motions $\{z_k^{(q)}\}$, $q = 1, 2, \dots$:

$$v_k^{(q)} = S^\top(k) F^+(0) d^{(q)}\left(0, z^{\{0\}}\right), \text{ where } S(k) = A^{l-k-1}B, \ k = \overline{0, l-1}, \tag{A.12}$$

$$d^{(q)}\left(0, z^{\{0\}}\right) = z^{\{1\}} - A^l z^{\{0\}} - \sum_{i=0}^{l-1} A^{l-i-1} \xi\left(z_i^{(q-1)}, Dz_i^{(q-1)} + v_i^{(q-1)}, h_0\right), \tag{A.13}$$

$$z_k^{(q)} = A^k z^{\{0\}} + \sum_{i=0}^{k-1} A^{k-i-1} B v_i^{(q)} + \sum_{i=0}^{k-1} A^{k-i-1} \xi\left(z_i^{(q-1)}, Dz_i^{(q-1)} + v_i^{(q-1)}, h_0\right), \quad k = \overline{0, l}, \tag{A.14}$$

with $z_0^{(q)} = z^{\{0\}} = z_0$ and $z_l^{(q)} = z^{\{1\}} = z_l$ for $q = 1, 2, \dots$.

To justify the convergence of the iterative process, let us treat the pair of functions $\{z_k^{(q)}, v_k^{(q)}\}$ as an element of the Euclidean space R^{n+m+q} with the norm

$$\rho\left(\left\{z_k^{(q)}, v_k^{(q)}\right\}\right) = (\|D\| + 1) \max_{0 \le k \le l} \left\|z_k^{(q)}\right\| + \max_{0 \le k \le l-1} \left\|v_k^{(q)}\right\|. \tag{A.15}$$

Similar to the proof of Theorem 5, we derive the following upper bounds:

$$\begin{gathered} \max_{0 \le k \le l-1} \left\|v_k^{(q)} - v_k^{(q-1)}\right\| \\ \le lL \max_{0 \le k \le l-1} \left\|S^\top(k) F^+(0)\right\| \max_{0 \le k \le l-1} \left\|A^{l-k-1}\right\| \rho\left(\left\{z_k^{(q-1)}, v_k^{(q-1)}\right\} - \left\{z_k^{(q-2)}, v_k^{(q-2)}\right\}\right), \end{gathered}$$

$$\begin{aligned} &\max_{0 \le k \le l} \left\|z_k^{(q)} - z_k^{(q-1)}\right\| \\ &\le lL \left(l \max_{0 \le k \le l-1} \left\|A^{l-k-1}B\right\| \max_{0 \le k \le l-1} \left\|S^\top(k) F^+(0)\right\| \max_{0 \le k \le l-1} \left\|A^{l-k-1}\right\| \right. \\ &\qquad \left. + \max_{0 \le k \le l-1} \left\|A^{l-k-1}\right\|\right) \rho\left(\left\{z_k^{(q-1)}, v_k^{(q-1)}\right\} - \left\{z_k^{(q-2)}, v_k^{(q-2)}\right\}\right). \end{aligned}$$

Let

$$m = \max\left\{\max_{0\le k\le l-1}\|S^\top(k)F^+(0)\|\max_{0\le k\le l-1}\|A^{l-k-1}\|, \max_{0\le k\le l-1}\|A^{l-k-1}\|,\right.$$
$$\left.\max_{0\le k\le l-1}\|S^\top(k)F^+(0)\|\max_{0\le k\le l-1}\|A^{l-k-1}\|\max_{0\le k\le l-1}\|A^{l-k-1}B\|\right\}.$$

Adding the first bound to the second one multiplied by $(\|D\|+1)$, in view of (A.15), gives

$$\begin{aligned}&\rho\left(\left\{z_k^{(q)}, v_k^{(q)}\right\} - \left\{z_k^{(q-1)}, v_k^{(q-1)}\right\}\right)\\ &\le Llm(1+(l+1)(\|D\|+1))\rho\left(\left\{z_k^{(q-1)}, v_k^{(q-1)}\right\} - \left\{z_k^{(q-2)}, v_k^{(q-2)}\right\}\right).\end{aligned} \tag{A.16}$$

Under condition (20) we establish $0 < Llm(1+(l+1)(\|D\|+1)) < 1$; then the sequence of vector functions $\{z_k^{(q)}, v_k^{(q)}\}$ converges, uniformly in k, to some vector function $\{z_k^*, v_k^*\}$, and the operator $F(z,v) : R^{n+m+q} \to R^{n+m+q}$ (A.12)–(A.14) is a contraction. According to the principle of contractive mappings, it has a unique fixed point, i.e., the limit vector function $\{z_k^*, v_k^*\}$ with $z_k^* = \lim_{q\to\infty} z_k^{(q)}$, $v_k^* = \lim_{q\to\infty} v_k^{(q)}$. In addition,

$$v_k^* = S^\top(k)F^+(0)\left(z^{\{l\}} - A^l z^{\{0\}} - \sum_{i=0}^{l-1} A^{l-i-1}\xi(z_i^*, Dz_i^* + v_i^*, h_0)\right), \quad k = \overline{0, l-1},$$

$$z_k^* = A^k z^{\{0\}} + \sum_{i=0}^{k-1} A^{k-i-1}Bv_i^* + \sum_{i=0}^{k-1} A^{k-i-1}\xi(z_i^*, Dz_i^* + v_i^*, h_0), \ k = \overline{0, l},$$

are the admissible control law and the corresponding motion of system (19), respectively.

For $h = h_0$, system (8) has the form $z_{k+1} = A_0 z_k + Bu_k + \xi(z_k, u_k, h_0)$, $k = \overline{0, l-1}$. We substitute the control law $u_k^* = Dz_k^* + v_k^*$ and the motion z_k^*, $k = \overline{0, l-1}$, into this system to obtain

$$z_{k+1}^* = Az_k^* + Bv_k^* + \xi(z_k^*, Dz_k^* + v_k^*, h_0), \quad k = \overline{0, l-1}.$$

Actually, this is system (19) with $z_k = z_k^*$ and $v_k = v_k^*$; in addition, $z_0^* = z^{\{0\}} = z_0$ and $z_l^* = z^{\{1\}} = z_l$. Hence, $u_k^* = w(z_k^*) + v_k^*$ $(k = \overline{0, l-1})$ is an admissible control law for system (8) with $h = h_0$; and due to the equivalence of systems (1) and (8), it is the desired admissible control law for system (1) with $h = h_0$. Moreover, system (1) is completely controllable for $h = h_0$, and the proof of Theorem 6 is finished.

Proof of Lemma 1. Let $(z_k, u_k) \in M$. Due to equalities (24), (23), $z_{k+1} = \tilde f(z_k, u_k)$, and the boundary value conditions $z_0 = z^{\{0\}}$, $z_l = z^{\{1\}}$, we obtain

$$\begin{aligned}L(z_k, u_k, \varphi) &= -\sum_{k=0}^{l-1}\left(\varphi\left(\tilde f(z_k, u_k)\right) - \varphi(z_k) - f^0(z_k, u_k)\right) + \varphi(z^{\{1\}}) - \varphi(z^{\{0\}})\\ &= -\sum_{k=0}^{l-1}\left(\varphi(z_{k+1}) - \varphi(z_k) - f^0(z_k, u_k)\right) + \varphi(z_l) - \varphi(z_0)\\ &= -\varphi(z_1) + \varphi(z_0) + f^0(z_0, u_0) - \varphi(z_2) + \varphi(z_1) + f^0(z_1, u_1) - \ldots - \varphi(z_{l-1})\\ &+ \varphi(z_{l-2}) + f^0(z_{l-2}, u_{l-2}) - \varphi(z_l) + \varphi(z_{l-1}) + f^0(z_{l-1}, u_{l-1}) + \varphi(z_l) - \varphi(z_0)\\ &= \sum_{k=0}^{l-1} f^0(z_k, u_k) = I(z_k, u_k).\end{aligned}$$

The proof of Lemma 1 is complete.

Proof of Lemma 2. It rests on Theorem 5.1 from [19], stating (as applied to minimization of the cost functional (22)) that a process (z_k^*, u_k^*) is optimal if and only if

$$f^0(z_k^*, u_k^*) = \min_{(z_k,u_k)\in D} f^0(z_k, u_k), \quad k = \overline{0, l-1}.$$

Since $\varphi(z^{\{1\}}) - \varphi(z^{\{0\}}) \equiv \text{const}$, it follows from (24) that only the expression $\sum_{k=0}^{l-1} R(z_k, u_k)$ affects the minimization of the functional $L(z_k, u_k, \varphi)$. Then, by taking (25) into account and applying Theorem 5.1 from [19] to the functional (24), we arrive at

$$-\sum_{k=0}^{l-1} R(z_k^*, u_k^*) = \min_{(z_k,u_k)\in D} \left(-\sum_{k=0}^{l-1} R(z_k, u_k)\right) \text{ and}$$

$$\min_{(z_k,u_k)\in D} L(z_k, u_k, \varphi) = \min_{(z_k,u_k)\in D} \left(-\sum_{k=0}^{l-1} R(z_k, u_k)\right) + \varphi(z^{\{1\}}) - \varphi(z^{\{0\}})$$
$$= -\sum_{k=0}^{l-1} R(z_k^*, u_k^*) + \varphi(z^{\{1\}}) - \varphi(z^{\{0\}}) = L(z_k^*, u_k^*, \varphi), \quad k = \overline{0, l-1}.$$

The proof of Lemma 2 is complete.

Proof of Theorem 7. Letting

$$l_\varphi = \min_{(z_k,u_k)\in D} L(z_k, u_k, \varphi)$$

gives

$$L(z_k, u_k, \varphi) \geq l_\varphi \text{ for } (z_k, u_k)\in D, \tag{A.17}$$

and, by Lemma 2,

$$L(z_k^*, u_k^*, \varphi) = l_\varphi. \tag{A.18}$$

According to (22) and Lemma 1, we have $L(z_k, u_k, \varphi) = I(z_k, u_k)$ for $(z_k, u_k)\in M$. Then from (A.17) and (A.18) it follows that $I(z_k, u_k) \geq l_\varphi$ for $(z_k, u_k)\in M$, and $I(z_k^*, u_k^*) = l_\varphi$. Hence, $I(z_k^*, u_k^*) \leq I(z_k, u_k)$ for $(z_k, u_k)\in M$, and (z_k^*, u_k^*) is the optimal process for system (8) with $h = h_0$; moreover, u_k^* is the optimal control law for this system. Since $I(z_k, u_k) = I(u_k)$ and the equivalent systems (1) and (8) have a common control performance index, $u^*(t_k) = u_k^*$ is the optimal control law for system (1) with $h = h_0$. The proof of Theorem 7 is complete.

Proof of Theorem 8. Consider the cost functional (24) on the set D and let

$$l_\varphi = \inf_{(z_k,u_k)\in D} L(z_k, u_k, \varphi).$$

According to Lemma 1, we have $L(z_k, u_k, \varphi) = I(z_k, u_k)$ for $(z_k, u_k)\in M$, $M \subset D$. Then

$$\inf_{(z_k,u_k)\in D} L(z_k, u_k, \varphi) \leq \inf_{(z_k,u_k)\in M} L(z_k, u_k, \varphi)$$

and

$$\inf_{(z_k,u_k)\in M} I(z_k, u_k) \geq l_\varphi. \tag{A.19}$$

On the other hand, the condition of this theorem implies

$$-\sum_{k=0}^{l-1} R\left(z_k^{(s)}, u_k^{(s)}\right) \to \inf_{(z_k,u_k)\,\in\, D} \left(-\sum_{k=0}^{l-1} R(z_k, u_k)\right) \quad \text{as} \quad s \to \infty,$$

and because $\varphi(z^{\{1\}}) - \varphi(z^{\{0\}}) \equiv \text{const}$, $L(z_k^{(s)}, u_k^{(s)}, \varphi) \to l_\varphi$ as $s \to \infty$. Due to (A.19) and the definition of the infimum of a functional, we establish the convergence

$$I\left(z_k^{(s)}, u_k^{(s)}\right) \to \inf_{(z_k,u_k)\,\in\, M} I(z_k, u_k),$$

i.e., $\{z_k^{(s)}, u_k^{(s)}\}$ is a minimizing sequence of admissible processes for the functional $I(z_k, u_k) = \sum_{k=0}^{l-1} u_k^\top u_k$. Then $\{u^{(s)}(t_k)\}$ is a minimizing sequence of admissible control actions of system (1) with $h = h_0$ on which the functional (5) tends to its greatest lower bound. The proof of Theorem 8 is complete.

REFERENCES

1. Logunova, O.S., Agapitov, E.B., Barankova, I.I., et al., Mathematical Models for Investigation of the Heat Condition of Objects and Heat Processes Control, *Electrotechnical Systems and Complexes*, 2019, no. 2(43), pp. 25–34.

2. Maksimov, V.P. and Chadov, A.L., Hybrid Models in Problems of Economic Dynamics, *Bulletin of Perm University*, 2011, no. 2(9), pp. 13–23.

3. Shpilevaya, O.Ya. and Kotov, K.Yu., Switched Systems: Stability and Design, *Optoelectronics, Instrumentation and Data Processing*, 2008, vol. 44, no. 5, pp. 439–449.

4. Bortakovsky, A.S. and Uryupin, I.P., Routes Optimization of Continuous-Discrete Movement of Controlled Objects in the Presence of Obstacles, *Trudy MAI*, 2020, vol. 113. https://doi.org/10.34759/trd-2020-113-17

5. Bortakovskii, A.S. and Panteleev, A.V., Sufficient Conditions for Optimal Control of Continuously Discrete Systems, *Avtom. Telemekh.*, 1987, no. 7, pp. 57–66.

6. Akmanova, S.V., On Controllability and Stabilization of Nonlinear Continuous–discrete Dynamic Systems, *The Bulletin of Irkutsk State University. Series Mathematics*, 2025, vol. 52, pp. 3–20.

7. Akmanova, S.V., Stabilization of Nonlinear Continuous-Discrete Dynamic Systems with a Constant Sampling Step, *Autom. Remote Control*, 2024, vol. 85, no. 9, pp. 864–878.

8. Bortakovskii, A.S., Necessary Optimality Conditions for Switching Systems, *Trudy Inst. Mat. i Mekh. Ural. Otd. Ross. Akad. Nauk*, 2021, vol. 27, no. 2, pp. 67–78.

9. Krotov, V.F. and Gurman, V.I., *Metody i zadachi optimal'nogo upravleniya* (Methods and Problems of Optimal Control), Moscow: Nauka, 1973.

10. Albrekht, E.H. and Sazanova, L.A., On Constructing the Admissible Control in Globally Controllable Nonlinear Discrete Systems, *Izv. Ural. Gos. Univ.*, 2003, no. 26, pp. 11–23.

11. Sazanova, L.A., On Control of Discrete Systems, *Cand. Sci. (Phys.–Math.) Dissertation*, Yekaterinburg: Ural State University, 2002.

12. Krasovskii, N.N., On a Problem of Optimum Control of Nonlinear Systems, *Journal of Applied Mathematics and Mechanics*, 1959, vol. 23, no. 2, pp. 303–332.

13. Krasovskii, N.N., *Teoriya upravleniya dvizheniem* (Theory of Motion Control), Moscow: Nauka, 1968.

14. Subbotin, A.I., Control of Motion of a Quasilinear System, *Differ. Uravn.*, 1967, vol. 3, no. 7, pp. 1113–1118.

15. Al'brekht, E.H., On Control over the Motion of Nonlinear Systems, *Differ. Uravn.*, 1966, vol. 2, no. 3, pp. 324–334.

16. Al'brekht, E.H., The Optimal Control of the Motion of Quasilinear Systems, *Differ. Uravn.*, 1969, vol. 5, no. 3, pp. 430–442.

17. Boltyansky, V.G., *Mathematical Methods of Optimal Control*, Holt, Rinehart and Winston, 1971.

18. Nevliudov, I., Omarov, M., and Romashov, Y., One Approach to Find Optimal Controls for Discrete Dynamic Systems with Numerical Methods Application, *Advanced Mathematical Models & Applications*, 2023, vol. 8, no. 3, pp. 548–564.

19. Krotov, V.F., Lagosha, B.A., Lobanov, S.M., et al., *Osnovy teorii optimal'nogo upravleniya* (Foundations of Optimal Control Theory), Moscow: Vysshaya Shkola, 1990.

20. Nikol'skii, M.S., On Linear Multistep Controlled Processes, *Moscow Univ. Comput. Math. Cybern.*, 2020, vol. 44, no. 4, pp. 185–189.

21. Rasina, I.V., Iterative Optimization Algorithms for Discrete-Continuous Processes, *Autom. Remote Control*, 2012, vol. 73, no. 10, pp. 1591–1603.

22. Wigström, O. and Lennartson, B., An Integrated CP/OR Method for Optimal Control of Modular Hybrid Systems, *Proceedings of the 12th IFAC/IEEE Workshop on Discrete Event Systems*, 2014, pp. 486–491.

23. Bortakovskii, A.S., Sufficient Optimality Conditions for Hybrid Systems of Variable Dimension, *Proceedings of the Steklov Institute of Mathematics*, 2020, vol. 308, pp. 79–91.

24. Uryupin, I.V., Optimization of Switching of Continuous–Discrete Controlled Processes, *Cand. Sci. (Phys.–Math.) Dissertation*, Moscow: Moscow Aviation Institute, 2022.

25. Yumagulov, M.G. and Akmanova, S.V., On Stability of Equilibria of Nonlinear Continuous-Discrete Dynamical Systems, *Ufa Mathematical Journal*, 2023, vol. 15, no. 2, pp. 85–99.

26. Kwakernaak, H. and Sivan, R., *Linear Optimal Control Systems*, 1st ed., Wiley-Interscience, 1972.

27. Krasovskii, N.N., Stabilization Problems for Controlled Motions. Supplement IV, in *Teoriya ustoichivosti dvizheniya* (Theory of Stability of Motion), Malkin, I.G., Ed., Moscow: Nauka, 1966, pp. 475–514.

28. Digo, G.B. and Digo, N.B., Evalution Definition of Unknown Lipshits Constant in Optimization of Algorithmic Certain Function, *Information Science and Control Systems*, 2007, no. 1(13), pp. 92–97.

29. Lagosha, B.A. and Apal'kova, T.G., *Optimal'noe upravlenie v ekonomike: teoriya i prilozheniya* (Optimal Control in Economics: Theory and Applications), Moscow: Finansy i Statistika, 2008.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 1, pp. 65–76.
© The Author(s), 2026 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2026.
Russian Text © The Author(s), 2026, published in Avtomatika i Telemekhanika, 2026, No. 1, pp. 89–103.

NONLINEAR SYSTEMS

On Optimality of Singular Controls in One Optimal Control Problem of the First Order Stochastic Hyperbolic Equations

K. B. Mansimov[*,**,a] **and R. O. Mastaliyev**[***,**,b,c]

[*]*Baku State University, Baku, Azerbaijan*
[**]*Institute of Mathematics, Baku, Azerbaijan*
[***]*Azerbaijan University, Baku, Azerbaijan*
e-mail: [a]*kamilbmansimov@gmail.com,*
[b]*rashad.mastaliyev@au.edu.az,* [c]*mastaliyevrashad@gmail.com*
Received August 16, 2024
Revised September 20, 2025
Accepted September 23, 2025

Abstract—The optimal control problem described by a system of the first order stochastic hyperbolic equations and arising in the modeling of a number of chemical-technological processes under the influence of random noise is considered. The optimality of singular controls in the sense of the Pontryagin maximum principle is investigated.

Keywords: stochastic first order equations hyperbolic system, two parameter Wiener process, optimality, Pontryagin maximum principle, singular control, second order optimality conditions

DOI: 10.7868/S1608303226010058

1. INTRODUCTION

Various quality aspects, including derivation of necessary optimality conditions of first order and study of singular cases of optimal control problems for determinate dynamical systems described by the first order hyperbolic equations, have been studied in [1–6] and others.

Such an optimal control problem in the stochastic case has been considered in [7] and first-order necessary optimality conditions (the analogue of Pontryagin's maximum principle, linearized maximum principle, analogue of the Euler's equation [8]) have been established there.

The relevance of research in this direction is determined by the need for the most accurate description, for example, of automatic control systems, a number of chemical technological processes [9–14], whose realistic option is a stochastic description that takes into account the impact of random factors on the controlled object.

The present work is devoted to the study of the singular case in the sense of Pontryagin's maximum principle and derivation of the second order necessary optimality conditions for singular controls in the stochastic control problem described by the system of nonlinear stochastic first order hyperbolic equations in canonical form.

The applied scheme of the study is a modification and development of the schemes from [5, 6], allowing for stochastic properties of the problem under consideration.

2. PROBLEM STATEMENT

Assume that the control process in the given domain $D = [t_0, t_1][x_0, x_1]$ is described by the following system of stochastic nonlinear partial differential equations of first order

$$\begin{aligned} \frac{\partial z(t,x)}{\partial t} &= f(t,x,z,y,u) + p(t,x,z)\frac{\partial W_1(t,x)}{\partial t}, \\ \frac{\partial y(t,x)}{\partial x} &= g(t,x,z,y,u) + q(t,x,y)\frac{\partial W_2(t,x)}{\partial x}, \quad (t,x) \in D, \end{aligned} \tag{1}$$

with Goursat type boundary conditions

$$z(t_0, x) = a(x), \quad x \in [x_0, x_1], \quad y(t, x_0) = b(t), \quad t \in [t_0, t_1]. \tag{2}$$

Here $(z(t,x), y(t,x))$ is $(n+m)$-dimensional sought vector-function: $f(t,x,z,y,u)$ $(g(t,x,z,y,u))$ is a given $n(m)$-dimensional vector-function continuous in totality of variables together with partial derivative with respect to (z, y) up to the second order inclusively; $p(t,x,z)$ $(q(t,x,y))$ is $(n \times n)$ $((m \times m))$-dimensional measurable and bounded matrix-function; white noises $\frac{\partial W_1(t,x)}{\partial t}$, $\frac{\partial W_2(t,x)}{\partial x}$ are the derivatives with respect to t and x, respectively, of the two-parameter Wiener's process $W_1(t,x)$, $W_2(t,x)$ [10, 13]; $a(x)$, $b(t)$ are the given measurable and bounded on $[x_0, x_1]$, $[t_0, t_1]$ vector-function of appropriate dimensions.

As admissible controls we take a class of measurable with respect to non-decreasing Borel σ algebra $\mathcal{F} = \overline{s}(W(t,s), t_0 \leqslant t \leqslant t_1, x_0 \leqslant s \leqslant x_1)$ and bounded in D r-dimensional vector-functions $u(t,x)$ with the values from the given non-empty and bounded set $U \subset R^r$ $(u(t,x) \in L_\infty(D,U))$.

The solution $(z(t,x), y(t,x))$ of system (1)–(2) corresponding to the definite admissible control $u(t,x)$ is understood in the sense [15].

Everywhere it is assumed that for each admissible control the corresponding solution to the boundary value problem (1)–(2) exists and is unique in the domain D.

Let us consider the problem of the minimization of the functional

$$\begin{aligned} S(u) = E\Bigg\{ &\int_{t_0}^{t_1}\int_{x_0}^{x_1} F_3(t,x,z(t,x),y(t,x),u(t,x))\,dxdt \\ &+ \int_{x_0}^{x_1} F_1(x, z(t_1,x))\,dx + \int_{t_0}^{t_1} F_2(t, y(t,x_1))\,dt \Bigg\}, \end{aligned} \tag{3}$$

determined on the solutions of the boundary value problem (1)–(2), generated by all admissible controls.

Here $F_1(x,z)$, $F_2(t,y)$, $F_3(t,x,z,y,u)$ are the given scalar functions continuous in totality of variables together with partial derivatives with respect to the state vector (i.e. with respect to (z,y)) up to the second order inclusively, E stands for mathematical expectation.

An admissible control $u(t,x)$ that minimizes the functional (3) subject to (1) and (2) is called an optimal control. The corresponding process $(u(t,x), z(t,x), y(t,x))$ is then called an optimal process.

It is assumed that in the considered stochastic problem there exists an optimal control.

The main goal of the paper is to derive second order necessary conditions for optimality (the case of singular controls) in the studied stochastic control problem with distributed parameters.

3. SECOND ORDER INCREMENT FORMULA OF THE QUALITY FUNCTIONAL

Let $(u(t,x), z(t,x), y(t,x))$ be fixed, $(\overline{u}(t,x) = u(t,x) + \Delta u(t,x), \overline{z}(t,x) = z(t,x) + \Delta z(t,x), \overline{y}\,(t,x) = y\,(t,x) + \Delta y\,(t,x))$ be arbitrary admissible processes.

We introduce the analogue of the Hamilton–Pontryagin function

$$H\,(t,x,z,y,u,\psi,\xi) = -F_3\,(t,x,z,y,u) + \psi' f\,(t,x,z,y,u) + \xi' g\,(t,x,z,y,u)\,,$$

and the denotation of type:

$$\Delta_v f\,[t,x] = f\,(t,x,z\,(t,x)\,, y\,(t,x)\,, v) - f\,(t,x,z\,(t,x)\,, y\,(t,x)\,, u\,(t,x))\,,$$
$$H_z\,[t,x] = H_z(t,x,z(t,x),y(t,x),u(t,x),\psi(t,x),\xi(t,x)).$$

Here $(\psi(t,x), \xi(t,x), \alpha(t,x), \beta(t,x)) \in L_\infty(D,R^n) \times L_\infty(D,R^m) \times L_\infty(D,R^{n\times n}) \times L_\infty(D,R^{m\times m})$ are the solutions of the following stochastic conjugate problem

$$\psi_t\,(t,x) = -\frac{\partial H\,(t,x,z,y,u,\psi,\xi)}{\partial z} + \alpha\,(t,x)\,\frac{\partial W_1\,(t,x)}{\partial t},\quad \psi\,(t_1,x) = \frac{\partial F_1\,(x,z\,(t_1,x))}{\partial z},$$
$$\xi_x\,(t,x) = -\frac{\partial H\,(t,x,z,y,u,\psi,\xi)}{\partial y} + \beta\,(t,x)\,\frac{\partial W_2\,(t,x)}{\partial x},\quad \xi\,(t,x_1) = \frac{\partial F_2\,(t,y\,(t,x_1))}{\partial y}.$$

Now, applying the Taylor formula and taking into account the introduced denotations, we can write increment of quality criterion (3) corresponding to the admissible controls $u\,(t,x)$ and $\overline{u}\,(t,x)$ in the form

$$\begin{aligned}
\Delta S\,(u) = -E\Bigg\{ &\int_{t_0}^{t_1}\int_{x_0}^{x_1} \Delta_{\overline{u}} H\,[t,x]\,dxdt \\
&+ \frac{1}{2}\int_{x_0}^{x_1} \Delta z'\,(t_1,x)\,\frac{\partial^2 F_1\,(x,z\,(t_1,x))}{\partial z^2}\Delta z\,(t_1,x)dx \\
&+ \frac{1}{2}\int_{t_0}^{t_1} \Delta y'\,(t,x_1)\,\frac{\partial^2 F_2\,(t,y\,(t,x_1))}{\partial y^2}\Delta y\,(t,x_1)dt \\
&+ \int_{t_0}^{t_1}\int_{x_0}^{x_1} \Big[\Delta_{\overline{u}} H_z'\,[t,x]\,\Delta z\,(t,x) + \Delta_{\overline{u}} H_y'\,[t,x]\,\Delta y\,(t,x)\Big]\,dxdt \\
&+ \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1} \Big[\Delta z'(t,x) H_{zz}[t,x]\Delta z(t,x) + \Delta z'(t,x) H_{zy}[t,x]\Delta y(t,x) \\
&+ \Delta y'\,(t,x)\,H_{yy}\,[t,x]\,y\,(t,x)\,\Big]dxdt + \eta_1\,(t,x;\Delta u)\,,
\end{aligned} \tag{4}$$

where $\eta_1(t,x;\Delta u)$ is determined by the formula

$$\eta_1(t,x;\Delta u) = E\left\{ \int_{x_0}^{x_1} o_1\left(\|\Delta z(t_1,x)\|^2\right)dx + \int_{t_0}^{t_1} o_2\left(\|\Delta y(t,x_1)\|^2\right)dt \right.$$

$$- \int_{t_0}^{t_1}\int_{x_0}^{x_1} o_3\left([\|\Delta z(t,x)\| + \|\Delta y(t,x)\|]^2\right)dxdt$$

$$-\frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1}\Big[\Delta z'(t,x)\,\Delta_{\overline{u}}H_{zz}[t,x]\,\Delta z(t,x) + \Delta z'(t,x)\,\Delta_{\overline{u}}H_{zy}[t,x]\,\Delta y(t,x)$$

$$\left. + \Delta y'(t,x)\Delta_{\overline{u}}H_{yz}[t,x]\Delta z(t,x) + \Delta y'(t,x)\Delta_{\overline{u}}H_{yy}[t,x]\Delta y(t,x)\Big]dxdt \right\},$$

$\|\Delta z\|$ is a norm of the vector $\Delta z = (\Delta z_1, \Delta z_2, \ldots, \Delta z_n)$ in R^n defined by the formula $\|\Delta z\| = \sum_{i=1}^{n}|\Delta z_i|$, and everywhere the condition $\frac{o_i(\propto^2)}{a^2} \to 0$, holds as $\propto \to 0$.

On the other hand, by linearization from boundary value problem (1)–(2), we obtain that $(\Delta z(t,x), \Delta y(t,x))$ is the solution of the following stochastic linearized systems of equations

$$\begin{aligned}\Delta z_t &= f_z[t,x]\,\Delta z(t,x) + \Delta_{\overline{u}}f_y[t,x]\,\Delta y(t,x) + \Delta_{\overline{u}}f[t,x] \\ &\quad + p_z[t,x]\Delta z(t,x)\,\frac{\partial W_1(t,x)}{\partial t} + \eta_2(t,x;\Delta u), \\ \Delta y_x &= g_z[t,x]\,\Delta z(t,x) + \Delta_{\overline{u}}g_y[t,x]\,\Delta y(t,x) + \Delta_{\overline{u}}g[t,x] \\ &\quad + q_y[t,x]\,\Delta y(t,x)\,\frac{\partial W_2(t,x)}{\partial x} + \eta_3(t,x;\Delta u)\end{aligned} \tag{5}$$

with boundary conditions

$$\Delta z(t_0,x) = 0, \quad x \in [x_0,x_1]; \quad \Delta y(t,x_0) = 0, \quad t \in [t_0,t_1], \tag{6}$$

where by definition

$$\eta_2(t,x;\Delta u) = \Delta_{\overline{u}}f_z[t,x]\,\Delta z + \Delta_{\overline{u}}f_y[t,x]\,\Delta y + o_4(\|\Delta z + \Delta y\|) + o_5(\|\Delta z(t,x)\|)\,\frac{\partial W_1(t,x)}{\partial t},$$

$$\eta_3(t,x;\Delta u) = \Delta_{\overline{u}}g_z[t,x]\,\Delta z + \Delta_{\overline{u}}g_y[t,x]\,\Delta y + o_6(\|\Delta z + \Delta y\|) + o_7(\|\Delta y(t,x)\|)\,\frac{\partial W_1(t,x)}{\partial t}.$$

Here the quantities $o_i(.)$, $i = \overline{4,7}$ are determined from the expansion of functions $f(.)$, $p(.)$, $g(.)$, $q(.)$, respectively by the first order Taylor formula corresponding to the controls $\overline{u}(t,x)$ and $u(t,x)$.

Interpreting equations (5) as linear inhomogeneous stochastic equations with respect to $\Delta z(t,x)$, $\Delta y(t,x)$, and taking into account boundary conditions (6) based on the analogue of the Cauchy formula from [15], we obtain

$$\Delta z(t,x) = \int_{t_0}^{t} V_{11}(t,x;\tau,x)\,\Delta_{\overline{u}}f[\tau,x]dt + \eta_4(t,x;\Delta u), \tag{7}$$

$$\Delta y(t,x) = \int_{x_0}^{x} V_{22}(t,x;t,s)\,\Delta_{\overline{u}}g[t,s]ds + \eta_5(t,x;\Delta u). \tag{8}$$

Here, by definition

$$\eta_4(t,x;\Delta u) = \int_{t_0}^{t}\int_{x_0}^{x}\left\{\frac{\partial V_{11}(t,x;\tau,s)}{\partial x}\left[\Delta_{\overline{u}} f[\tau,s] + \eta_2(\tau,s;\Delta u)\right]\right.$$
$$\left. + \frac{\partial V_{12}(t,x;\tau,s)}{\partial x}\left[\Delta_{\overline{u}} g[\tau,s] + \eta_3(\tau,s;\Delta u)\right]\right\} ds d\tau$$
$$+ \int_{t_0}^{t} V_{11}(t,x;\tau,x)\,\eta_2(\tau,x;\Delta u)\,d\tau,$$

$$\eta_5(t,x;\Delta u) = \int_{t_0}^{t}\int_{x_0}^{x}\left\{\frac{\partial V_{21}(t,x;\tau,s)}{\partial t}\left[\Delta_{\overline{u}} f[\tau,s] + \eta_2(\tau,s;\Delta u)\right]\right.$$
$$\left. + \frac{\partial V_{22}(t,x;\tau,s)}{\partial t}\left[\Delta_{\overline{u}} g[\tau,s] + \eta_3(\tau,s;\Delta u)\right]\right\} ds d\tau$$
$$+ \int_{t_0}^{t} V_{22}(t,x;\tau,s)\,\eta_3(t,s;\Delta u) ds,$$

while $V_{ij}(t,x;\tau,s)$, $(t_0 \leqslant t \leqslant t \leqslant t_1, x_0 \leqslant s \leqslant x \leqslant x_1)$, $i,j = 1,2$ are matrix functions that are the solutions of the following stochastic problems [15]:

$$\frac{\partial V_{11}(t,x;\tau,s)}{\partial \tau} = -V_{11}(t,x;\tau,s)\,f_z[\tau,s] - V_{12}(t,x;\tau,s)\,g_z[\tau,s] - V_{11}(t,x;\tau,s)\,p[\tau,s]\frac{\partial W_1(\tau,s)}{\partial \tau},$$
$$\frac{\partial V_{12}(t,x;\tau,s)}{\partial s} = -V_{11}(t,x;\tau,s)\,f_y[t,s] - V_{12}(t,x;\tau,s)\,g_y[\tau,s] - V_{12}(t,x;\tau,s)\,q[\tau,s]\frac{\partial W_2(\tau,s)}{\partial s},$$
$$\frac{\partial V_{21}(t,x;\tau,s)}{\partial t} = -V_{21}(t,x;\tau,s)\,f_z[\tau,s] - V_{22}(t,x;\tau,s)\,g_z[\tau,s] - V_{21}(t,x;\tau,s)\,p[\tau,s]\frac{\partial W_1(\tau,s)}{\partial \tau},$$
$$\frac{\partial V_{22}(t,x;\tau,s)}{\partial s} = -V_{21}(t,x;\tau,s)\,f_z[\tau,s] - V_{21}(t,x;\tau,s)\,g_z[\tau,s] - V_{22}(t,x;\tau,s)\,q[\tau,s]\frac{\partial W_2(\tau,s)}{\partial s},$$
$$V_{11}(t,x;t,s) = E_1, \quad V_{12}(t,x;\tau,x) = 0, \quad t_0 \leqslant \tau \leqslant t,$$
$$V_{21}(t,x;\tau,s) = 0, \quad V_{22}(t,x;t,s) = E_2,$$

$x_0 \leqslant s \leqslant x$, where E_1, E_2 are unit matrices of corresponding dimensions.

For further presentations, we introduce into consideration the $(n \times n)$-dimensional matrix function $R(x,\tau,s)$ and $(m \times m)$-dimensional function $Q(t,\tau,s)$ defined as:

$$R(x,\tau,s) = \int_{\max(\tau,s)}^{t_1} V'_{11}(t,x;\tau,x) H_{zz}[t,x]\,V_{11}(t,x;s,x)\,dt$$
$$- V'_{11}(t_1,x;\tau,x)\,\frac{\partial^2 F_1(x,z(t_1,x))}{\partial z^2} V_{11}(t_1,x;s,x),$$

$$Q(t,\tau,s) = \int_{\max(\tau,s)}^{x_1} V'_{22}(t,x;t,x) H_{yy}[t,x]\,V_{22}(t,x;t,s)\,dx$$
$$- V'_{22}(t,x_1;\tau,t)\,\frac{\partial^2 F_2(t,y(t,x_1))}{\partial y^2} V_{22}(t,x_1;t,s).$$

Taking into account the introduced denotations and following [5, 6], using representations (7), (8) of the solutions Δz and Δy, we can represent the second order increment formula (4) of the quality criterion (3) in the following form:

$$
\begin{aligned}
\Delta S(u) = E\Bigg\{ &-\int_{t_0}^{t_1}\int_{x_0}^{x_1} \Delta_{\overline{u}} H[t,x]\,dxdt - \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1}\int_{t_0}^{t_1} \Delta_{\overline{u}} f'[\tau,x] R(x,\tau,s)\,\Delta_{\overline{u}} f[s,x]\,dsdxd\tau - \\
&- \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1}\int_{x_0}^{x_1} \Delta_{\overline{u}} g'[\tau,x] Q(t,\tau,s)\,\Delta_{\overline{u}} g[t,s]\,dtdsd\tau - \\
&- \int_{t_0}^{t_1}\int_{x_0}^{x_1}\left[\int_{t}^{t_1} \Delta_{\overline{u}} H_z'[\tau,x]\,V_{11}(\tau,x;t,x)\,dt\right]\Delta_{\overline{u}} f[t,x]\,dxdt - \\
&- \int_{t_0}^{t_1}\int_{x_0}^{x_1}\left[\int_{x}^{x_1} \Delta_{\overline{u}} H_y'[t,s]\,V_{22}(t,s;t,x)\,ds\right]\Delta_{\overline{u}} g[t,x]\,dxdt \Bigg\} + \eta(t,x;\Delta u),
\end{aligned}
\tag{9}
$$

where

$$
\begin{aligned}
\eta(t,x;\Delta u) = E\Bigg\{ &\eta_1(t,x;\Delta u) - \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1}\Big[\Delta z'(t,x) H_{zy}[t,x]\,\Delta y(t,x) \\
&+ \Delta y'(t,x)\,H_{yz}[t,x]\,\Delta z(t,x) + \eta_4(t,x)\,H_{zz}[t,x]\,\Delta z(t,x)\Big]dxdt \\
&- \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1}\left(\int_{t_0}^{t} V_{11}(t,x;\tau,x)\,\Delta_{\overline{u}} f[t,x]\,dt\right)' H_{zz}[t,x]\,\eta_4(t,x)\,dxdt \\
&+ \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1} V_{11}(t_1,x;\tau,x)\,\Delta_{\overline{u}} f[t,x]\,\frac{\partial^2 F_1(x,z(t_1,x))}{\partial z^2}\eta_4(t_1,x;\Delta u)\,dxdt \\
&+ \frac{1}{2}\int_{x_0}^{x_1} \eta_4'(t_1,x;\Delta u)\,\frac{\partial^2 F_1(x,z(t_1,x))}{\partial z^2}\Delta z(t_1,x)\,dx - \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1} \eta_5'(t,x;\Delta u)\,H_{yy}[t,x]\,\Delta y(t,x)\,dxdt \\
&- \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1}\left(\int_{x_0}^{x} V_{22}(t,x;\tau,s)\,\Delta_{\overline{u}} g[t,s]\,ds\right)' H_{yy}[t,x]\,\eta_5(t,x;\Delta u)\,dxdt \\
&+ \frac{1}{2}\int_{t_0}^{t_1}\int_{x_0}^{x_1} V_{22}(t,x_1;\tau,s)\,\Delta_{\overline{u}} g[t,s]\,\frac{\partial^2 F_2(t,y(t,x_1))}{\partial y^2}\eta_5(t,x_1;\Delta u) dsdt \\
&+ \frac{1}{2}\int_{t_0}^{t_1} \eta_5'(t,x_1)\,\frac{\partial^2 F_2(t,y(t,x_1))}{\partial y^2}\Delta y(t,x_1)\,dt \\
&- \int_{t_0}^{t_1}\int_{x_0}^{x_1} \Delta_{\overline{u}} H_z'[t,x]\eta_4(t,x;\Delta u)\,dxdt + \int_{t_0}^{t_1}\int_{x_0}^{x_1} \Delta_{\overline{u}} H_y'[t,x]\eta_5(t,x;\Delta u)\,dxdt \Bigg\}.
\end{aligned}
\tag{10}
$$

Note that the constructed increment formula (9) allows one to get Pontryagin's maximum principle type first order necessary optimality conditions and to investigate the case of degeneration of the maximum principle and its consequences from a unified standpoint.

4. STOCHASTIC ANALOGUE OF PONTRYAGIN'S MAXIMUM PRINCIPLE

First we give an auxiliary statement.

Lemma 1. *For almost all* $(t,x) \in D$ *the following estimations are valid*

$$\|\Delta z(t,x)\| \leqslant K_1 E\left(\int\limits_{t_0}^{t} \|\Delta_{\overline{u}} f\,[\tau,x]\|\, d\tau + \int\limits_{t_0}^{t}\int\limits_{x_0}^{x} \|\Delta_{\overline{u}} g\,[\tau,s]\|\, dsd\tau\right), \tag{11}$$

$$\|\Delta y(t,x)\| \leqslant K_2 E\left(\int\limits_{x_0}^{x} \|\Delta_{\overline{u}} g\,[t,s]\|\, ds + \int\limits_{t_0}^{t}\int\limits_{x_0}^{x} \|\Delta_{\overline{u}} f\,[\tau,s]\|\, dsd\tau\right), \tag{12}$$

where $K_i = \text{const} > 0,\ i = 1,2.$

The proof of the lemma is in the Appendix.

The validity of the statement [7] follows from the increment formula (9).

Theorem 1. *For the admissible control* $u\,(t,x)$ *be optimal in the studied distributed parameter stochastic control problem* (1)–(3) *it is necessary that the inequality*

$$E\Delta_v H\,[\theta,\gamma] \leqslant 0 \tag{13}$$

be fulfilled for all $(\theta,\gamma) \in [t_0,t_1) \times [x_0,x_1)$ *and for* $v \in U$.

Here

$$\begin{aligned}\Delta_v H\,[\theta,\gamma] &= H\left(\theta,\gamma,z\left(\theta,\gamma\right),y\left(\theta,\gamma\right),v,\psi\left(\theta,\gamma\right),\xi\left(\theta,\gamma\right)\right)\\ &\quad - H\left(\theta,\gamma,z\left(\theta,\gamma\right),y\left(\theta,\gamma\right),u\left(\theta,\gamma\right),\psi\left(\theta,\gamma\right),\xi\left(\theta,\gamma\right)\right),\end{aligned}$$

while $(\theta,\gamma) \in [t_0,t_1) \times [x_0,x_1)$ *is an arbitrary Lebesgue point* (*a regular control point* [16] $u\,(t,x)$).

The proof of Theorem 1 is given in the Appendix.

Inequality (13) is a first-order necessary condition for optimality and is a stochastic analogue of Pontryagin's maximum principle for control problem (1)–(3) under consideration.

5. NECESSARY CONDITIONS FOR OPTIMALITY OF CONTROLS SINGULAR IN THE SENSE OF PONTRYAGIN'S MAXIMUM PRINCIPLE

Although Pontryagin's maximum principle (13) is the strongest first-order necessary condition for optimality in the control problem (1)–(3) under consideration, in some cases it holds trivially (the singular case [3–6, 17]).

Following [3–6, 17], we introduce the definition of singular control for the problem under consideration.

Definition 1. The admissible control $u\,(t,x)$ is said to be singular in the sense of Pontryagin's maximum principle if for all $(\theta,\gamma) \in [t_0,t_1) \times [x_0,x_1)$ and $v \in U$

$$E\Delta_v H\,[\theta,\gamma] = 0. \tag{14}$$

In the singular case, i.e., if relation (14) is fulfilled, studying the increment formula (9), we prove the validity of the following statement.

Theorem 2. *For optimality of the control $u(t,x)$, singular in the sense of Pontryagin's maximum principle, in the considered distributed parameters stochastic control problem* (1)–(3), *it is necessary that for any $v \in U$, $(\theta,\gamma) \in [t_0,t_1) \times [x_0,x_1)$ the following relations be fulfilled:*

$$A(\theta,\gamma,v) = E\left[\Delta_v f'[\theta,\gamma] R(\gamma,\theta,\theta) + \Delta_v H_z^{'}[\theta,\gamma]\right] \Delta_v f[\theta,\gamma] \leqslant 0, \tag{15}$$

$$B(\theta,\gamma,v) = E\left[\Delta_v g^{'}[\theta,\gamma] Q(\theta,\gamma,\gamma) + \Delta_v H_y^{'}[\theta,\gamma]\right] \Delta_v g[\theta,\gamma] \leqslant 0. \tag{16}$$

The proof of Theorem 2 is given in the Appendix.

Now let us consider one special case that requires separate study.

Let in equation (1)

$$\begin{aligned} f(t,x,z,y,u) &= A_1(t,x)z + B_1(t,x)y + f_1(t,x,u), \\ g(t,x,z,y,u) &= A_2(t,x)z + B_2(t,x)y + f_2(t,x,u). \end{aligned} \tag{17}$$

And the minimizing functional is in the form:

$$S(u) = E\left\{\int_{x_0}^{x_1} C^{'}(x) z(t_1,x)\,dx + \int_{t_0}^{t_1} F_2(t,y(t,x_1))\,dt\right\}. \tag{18}$$

We introduce the denotations

$$K(t,\gamma,v) = V_{22}(t,x_1;t,\gamma)\Delta_{\overline{u}} g[t,\gamma]$$

$$+\int_{t_0}^{t}\left[\frac{\partial V_{21}(t,x_1;\tau,\gamma)}{\partial t}\Delta_v f[\tau,\gamma] + \frac{\partial V_{22}(t,x_1;\tau,\gamma)}{\partial t}\Delta_{\overline{u}} g[\tau,\gamma]\right] d\tau. \tag{19}$$

The following theorem is valid

Theorem 3. *For optimality of the control $u(t,x)$ singular in the sense of Pontryagin's maximum principle, in the considered distributed parameters stochastic control problem* (1), (2), (17), (18) *it is necessary that the inequality*

$$E\int_{t_0}^{t_1} K(t,\gamma,v)\,\frac{\partial^2 F_2(t,y(t,x_1))}{\partial y^2} K(t,\gamma,v)\,dt \geqslant 0$$

be fulfilled for all $\gamma \in [x_0,x_1)$, $v(t) \in U, t \in [t_0,t_1)$.

The proof of Theorem 3 is given in the Appendix.

Note that the similar result holds in the case of the functional of the form

$$S(u) = E\left\{\int_{x_0}^{x_1} F_1(x,z(t_1,x))\,dx + \int_{t_0}^{t_1} D^{'}(t) y(t,x_1)\,dt\right\}.$$

6. CONCLUSION

In order to develop a stochastic analogue of the increment method, in the considered problem optimal control of the systems of the first order hyperbolic equations, we set a second-order increment formula for the quality functional. Based on the obtained incremental formula, we proved Pontryagin's maximum principle type necessary condition for optimality.

Further, using special variations of admissible controls, we proved necessary conditions of optimality for the optimality of the singular (in the sense of Pontryagin's maximum principle) controls for the considered distributed parameters stochastic control problem.

APPENDIX

Proof of Lemma. From system of equations (5) considering (6) we obtain

$$\Delta z\,(t,x)=\int_{t_0}^{t}[f\,(\tau,x,\overline{z},\overline{y},\overline{u})-f\,(\tau,x,z,y,u)]d\tau+\int_{t_0}^{t}[p\,(\tau,x,\overline{z})\,-p(\tau,x,z\,)]\,\frac{\partial W_1\,(\tau,x)}{\partial\tau}d\tau,$$

$$\Delta y\,(t,x)=\int_{x_0}^{x}[g\,(t,s,\overline{z},\overline{y},\overline{u})-f\,(t,s,z,y,u)]ds+\int_{x_0}^{x}[q\,(t,s,\overline{z})\,-q(t,s,z\,)]\,\frac{\partial W_2\,(t,s)}{\partial s}ds.$$

Hence, passing to the norm and using the Lipschitz condition, and also accepting the mathematical expectations from both parts of the obtained inequalities, we obtain

$$E\,\|\Delta z(t,x)\|\leqslant E\left\{\int_{t_0}^{t}\|\Delta_{\overline{u}}f[t,x]\|dt+L_1\int_{t_0}^{t}\left(\|\Delta z(t,x)\|+\|\Delta y(t,x)\|\right)dt\right\},$$

$$E\,\|\Delta y(t,x)\|\leqslant E\left\{\int_{x_0}^{x}\|\Delta_{\overline{u}}g\,[t,s]\|ds+L_2\int_{x_0}^{x}\left(\|\Delta z\,(t,s)\|+\|\Delta y\,(t,s)\|\right)ds\right\}.$$

Applying the Gronwall–Vendroff's lemma to the last two inequalities (see, e.i. [18]), we have

$$E\,\|\Delta z(t,x)\|\leqslant E\left\{\int_{t_0}^{t}\|\Delta_{\overline{u}}f[t,x]\|dt+L_3\int_{t_0}^{t}\|\Delta y\,(t,x)\|\,dt\right\},\tag{A.1}$$

$$E\,\|\Delta y(t,x)\|\leqslant E\left\{\int_{x_0}^{x}\|\Delta_{\overline{u}}g\,[t,s]\|ds+L_4\int_{x_0}^{x}\|\Delta z\,(t,s)\|\,ds\right\},\tag{A.2}$$

where $L_i=\text{const}>0$, $i=\overline{1,4}$. Enhancing inequalities (A.1), (A.2) with each other and again using the Gronwall–Vendroff lemma, we obtain the validity of the estimation (11) for $\|\Delta z(t,x)\|$, and estimate (12) for $\|\triangle y(t,x)\|$.

The lemma is proved.

Proof of Theorem 1. Assuming that $u\,(t,x)$ is a fixed admissible control, we define its special increment in the form

$$\Delta u_\varepsilon\,(t,x)=\begin{cases} v-u\,(t,x)\,,\ (t,x)\in D_\varepsilon=[\theta,\theta+\varepsilon)\times\left[\gamma,\gamma+\varepsilon^2\right), \\ 0,(t,x)\in D/D_\varepsilon.\end{cases}\tag{A.3}$$

Here and in the sequel, $(\theta,\gamma)\in[t_0,t_1)\times[x_0,x_1)$ is an arbitrary regular point of the control $u\,(t,x)$, $v\in U$ is an arbitrary vector, while $\varepsilon>0$ is an arbitrary, rather small number.

By $(\triangle z_\varepsilon\,(t,x)\,,\Delta y_\varepsilon\,(t,x))$ we denote a special increment of the system (1)–(2), corresponding to the special increment (A.3) of the control $u\,(t,x)$.

Using estimations (11), (12) for $(\Delta z_\varepsilon(t,x), \Delta\, y_\varepsilon(t,x))$, we obtain

$$E\left\|\Delta z_\varepsilon(t,x)\right\| \leqslant \begin{cases} 0, (t,x) \in D \setminus D_\varepsilon, \\ K\varepsilon,\ (t,x) \in D_\varepsilon, \end{cases} \qquad E\left\|\Delta y_\varepsilon(t,x)\right\| \leqslant \begin{cases} 0, (t,x) \in D \setminus D_\varepsilon, \\ K\varepsilon,\ (t,x) \in D_\varepsilon. \end{cases}$$

Taking into account these estimation in formula (10), we see that the remainder term $\eta(\triangle u_\varepsilon(t,x))$ is a quantity of order $o\left(\varepsilon^4\right)$.

Consequently,

$$\begin{aligned} \Delta S_\varepsilon(u) = E\Bigg\{ &- \int\limits_{\theta}^{\theta+\varepsilon}\int\limits_{\gamma}^{\gamma+\varepsilon^2} \Delta_v H[t,x]\,dxdt \\ &-\frac{1}{2}\int\limits_{\theta}^{\theta+\varepsilon}\int\limits_{\theta}^{\gamma+\varepsilon^2}\int\limits_{\theta}^{\theta+\varepsilon} \Delta_v f'[\tau,x]R(x,\tau,s)\Delta_v f[s,x]\,dsdxd\tau \\ &-\frac{1}{2}\int\limits_{\theta}^{\theta+\varepsilon}\int\limits_{\gamma}^{\gamma+\varepsilon^2}\int\limits_{\gamma}^{\varepsilon^2} \Delta_v g'[\tau,x]Q(t,\tau,s)\Delta_v g[t,s]\,d\tau dsdt \\ &-\int\limits_{\theta}^{\theta+\varepsilon}\int\limits_{\gamma}^{\gamma+\varepsilon^2}\left[\int\limits_{t}^{t_1}\Delta_v H_z'[\tau,x]V_{11}(\tau,x;t,x)\,dt\right]\Delta_v f[t,x]\,dxdt \\ &-\int\limits_{\theta}^{\theta+\varepsilon}\int\limits_{\gamma}^{\gamma+\varepsilon^2}\left[\int\limits_{x}^{x_1}\Delta_v H_y'[t,s]V_{22}(t,s;t,x)\,ds\right]\Delta_v g[t,x]\,dxdt\Bigg\} + o\left(\varepsilon^4\right). \end{aligned} \tag{A.4}$$

Hence, after applying the mean value theorem and allowing for $\triangle S_\varepsilon(u) = S(u+\Delta u_\varepsilon) - S(u) \geqslant 0$, the validity of the theorem statement follows. Theorem 1 is proved.

Proof of Theorem 2. Considering (14), from expansion (A.4) we obtain

$$\Delta S_\varepsilon(u) = S(u+\Delta u_\varepsilon) - S(u)$$
$$= E\varepsilon^4\left[\Delta_v f'[\theta\,,\gamma]R(\gamma,\theta,\theta) + \Delta_v H_z'[\theta,\gamma]\right]\Delta_v f[\theta,\gamma] + o(\varepsilon^4) \geqslant 0.$$

Hence the statement (15) follows.

Now we determine the special increment of the singular control by the formula

$$u_\varepsilon(t,x) = \begin{cases} v - u(t,x),\ (t,x) \in D_\varepsilon = [\theta,\theta+\varepsilon^2) \times [\gamma,\gamma+\varepsilon), \\ 0, (t,x) \in D/D_\varepsilon. \end{cases}$$

Then expansion (9) yields that condition (16) is fulfilled along the process $(u(t,x), z(t,x), y(t,x))$, singular in the sense of Pontryagin's maximum principle. Thus, Theorem 2 is proved.

Proof of Theorem 3. In the case of problem (1), (2), (17), (18) formulas (7)–(9) yield that

$$\begin{aligned}
\Delta S(u) = &- E\int_{t_0}^{t_1}\int_{x_0}^{x_1} \Delta_{\overline{u}} H[t,x]\,dxdt \\
&+ \frac{1}{2}\int_{t_0}^{t_1} \Delta y'(t,x_1)\frac{\partial^2 F_2(t,y(t,x_1))}{\partial y^2}\Delta y(t,x_1)dt + \int_{t_0}^{t_1} o_2\left(\|\Delta y(t,x_1)\|^2\right)dt, \\
\Delta z(t_1,x) = &\int_{t_0}^{t_1} V_{11}(t_1,x;\tau,x)\,\Delta_{\overline{u}} f[t,x]dt \\
&+ \int_{t_0}^{t}\int_{x_0}^{x}\left[\frac{\partial V_{11}(t_1,x;\tau,s)}{\partial x}\Delta_{\overline{u}} f[t,s] + \frac{\partial V_{12}(t_1,x;\tau,s)}{\partial x}\Delta_{\overline{u}} g[t,s]\right]dsdt, \\
\Delta y(t,x_1) = &\int_{x_0}^{x_1} V_{22}(t,x_1;\tau,s)\,\Delta_{\overline{u}} g[t,s]ds \\
&+ \int_{t_0}^{t}\int_{x_0}^{x}\left[\frac{\partial V_{21}(t,x_1;\tau,s)}{\partial t}\Delta_{\overline{u}} f[t,s] + \frac{\partial V_{22}(t,x_1;\tau,s)}{\partial t}\Delta_{\overline{u}} g[t,s]\right]dsdt.
\end{aligned} \tag{A.5}$$

We determine the special increment of the control $u(t,x)$ by the formula

$$\Delta u_\varepsilon(t,x) = \begin{cases} v(t) - u(t,x), \ (t,x) \in D_\varepsilon = [t_0,t_1) \times [\gamma,\gamma+\varepsilon), \\ \qquad 0, (t,x) \in D \backslash D_\varepsilon, \end{cases}$$

where $\varepsilon > 0$ is an arbitrary rather number $v(t) \in U$, $\gamma \in [x_0,x_1)$. Then (A.5) yields that for $(t,x) \in [t_0,t_1] \times [\gamma,x_1]$

$$\Delta y_\varepsilon(t,x_1) = \varepsilon\,[V_{22}(t,x_1;t,\gamma)\,\Delta_v g[t,\gamma]$$

$$+\int_{t_0}^{t}\left[\frac{\partial V_{21}(t,x_1;\tau,s)}{\partial t}\Delta_v f[\tau,s] + \frac{\partial V_{22}(t,x_1;\tau,\gamma)}{\partial t}\Delta_{\overline{u}} g[\tau,\gamma]\right]d\tau + o(\varepsilon).$$

Therefore, taking into account the denotation $K(t,\gamma,v)$ determined by formula (19), we obtain

$$\Delta S_\varepsilon(u) = -E\int_{t_0}^{t_1}\int_{\gamma}^{\gamma+\varepsilon} \Delta_{v(t)} H[t,x]\,dxdt$$

$$+\frac{1}{2}\varepsilon^2\int_{t_0}^{t_1} K(t,\gamma,v)\frac{\partial^2 F_2(t,y(t,x_1))}{\partial y^2}K(t,\gamma,v)dt + o\left(\varepsilon^2\right).$$

The statement of Theorem 3 follows from this expansion. Theorem 3 is proved.

REFERENCES

1. Volin, Yu.M., Ostrovskiy, G.M., and Finkel'shteyn, A.V., Penalty function method and Pontryagin's maximum principle in optimal control problems, *Upravlyaemye sistemy*, 1973, vol. 11, pp. 88–95.

2. Volin, Yu.M., Ostrovskiy, G.M., and Finkel'shteyn, A.V., Penalty functions method and conditions for optimality of generalized solutions with partial derivatives, *Differ. Uravneniya*, 1977, vol. 13, no. 12, pp. 2246–2255.

3. Vasil'yev, O.V., To the study of a singular control in a system with distributed parameters, *Upravlyaemye sistemy*, 1976, vol. 15, pp. 3–15.

4. Vasil'yev, O.V., *Printsip maksimuma L.S. Pontryagina v teorii optimal'nykh sistem s raspredelennymi parametrami* (L.S. Pontryagin's maximum principle in theory of distributed parameters optimal systems), Novosibirsk: Prikl. Matematika, 1978.

5. Mansimov, K.B., To theory of necessary optimality conditions in a distributed parameters system, *J. Comput. Math. Math. Phys.*, 2001, vol. 41, no. 10, pp. 1505–1520.

6. Mansimov, K.B. and Mardanov, M.J., *Kachestvennaya teoriya optimal'nogo upravleniya sistemami Gursa–Darbu* (Qualitative theory of optimal control of Goursat-Darboux systems), Baku: Elm, 2010.

7. Mansimov, K.B. and Mastaliyev, R.O., *Neobkhodimye usloviya optimal'nosti pervogo poryadka v odnoi stokhasticheskoi zadache upravleniya s raspredelennymi parametrami* (First order necessary optimality conditions in a stochastic control problem with distributed parameters), *VSPU-2024*, Moscow: IPU RAN, 17–20 June, 2024, pp. 547–549.

8. Gabasov, R. and Kirillova, F.M., *Printsip maksimuma v teorii optimal'nogo upravleniya* (Maximum principle in optimal control theory), URSS, 2011.

9. Prato, G.D. and Tubaro, L., *Stochastic partial differential equations and applications*, Springer-Verlag, 1985.

10. Oksendal, B., Optimal control of stochastic partial deferential equations, *Stochastic Analysis and Applications*, 2005, vol. 23, pp. 165–179.

11. Bunkin, N.F. and Morozov, A.N., *Stokhasticheskie sistemy v fizike i tekhnike* (Stochastic systems in physics and engineering), Moscow: MGTU, 2011.

12. Qi, Lu. and Xu, Zhang, Control theory for stochastic distributed parameter systems, an engineering perspective, *Annual Reviews in Control*, 2021, vol. 51, pp. 268–330.

13. Qi Lu and Xu Zhang, *Mathematical control theory for stochastic partial differential equations*, Cham, Switzerland: Springer, 2021.

14. Prato, G.D. and Tubaro, L., *Stochastic partial differential equations and applications*, New York: Marcel Dekker, 2002.

15. Mansimov, K.B. and Mastaliyev, R.O., On representation of the solution of Goursat boundary value problem for stochastic hyperbolic partial differential equations of the first order, *Bulletin of Irkutsk State University, series Math.*, 2023, vol. 45, pp. 145–151.

16. Novozhenov, M.M., Sumin, V.I. and Sumin, M.I., *Metody optimal'nogo upravleniya sistemami matematicheskoi fiziki* (Methods of optimal control of systems of mathematical physics), Gorky: Izd. Gor'kovsk. Gos. Univ., 1986.

17. Gabasov, R. and Kirillova, F.M., *Osobye optimal'nye upravleniya* (Singular optimal controls), Moscow: URSS, 2011.

18. Plotnikov, V.I. and Sumin, V.I., Optimization of objects with distributed parameters described by Goursat-Darboux systems, *U.S.S.R. Comput. Math. Math. Phys.*, 1972. vol. 12. no. 1. pp. 73–92.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 1, pp. 77–88.
© The Author(s), 2026 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2026.
Russian Text © The Author(s), 2026, published in Avtomatika i Telemekhanika, 2026, No. 1, pp. 104–119.

Global Stabilization of a Second-Order Integrator by a Discontinuous Feedback

Yu. V. Morozov

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
e-mail: tot1983@ipu.ru

Received October 4, 2024
Revised June 17, 2025
Accepted August 12, 2025

Abstract—The stability of a switching system arising from the use of a discontinuous feedback to stabilize a second-order integrator is analyzed. This is the limiting case of the feedback in the form of nested saturation functions considered previously, where the external saturation function or sigmoid is replaced by the discontinuous function $\operatorname{sgn}(x)$. Such a feedback involves a bounded control resource and ensures constraints on the phase velocity of approaching an equilibrium, which is especially important under large initial deviations. A Lyapunov function for the closed-loop system is proposed, and the global asymptotic stability of the origin is proven using this function and the results of Filippov's theory, provided that the switching curve is a continuously differentiable and monotonic function passing through zero. The global stability of the origin is preserved when relaxing the smoothness requirement for the entire switching function to its continuous differentiability everywhere except for a finite set of points where the derivative does not exist, but continuity holds. In the case of a discontinuous switching curve, the origin is shown to be a sewn center according to Filippov's classification. In this case, the closed-loop system has a semi-stable cycle enclosing a set of discontinuity points on the switching curve. As numerical examples, the level line of the Lyapunov function and phase trajectories on the plane are constructed for different types of switching curves.

Keywords: stabilization of a chain of two integrators, switching system, global asymptotic stability, switching curve, Lyapunov function

DOI: 10.7868/S1608303226010069

1. INTRODUCTION

In 2024, two papers were written in collaboration with A.V. Pesterev, namely, "Global Stabilization of a Chain of Two Integrators by a Feedback in the Form of Nested *Saturators*"[1] and "Global Stabilization of a Chain of Two Integrators by a Feedback in the Form of Nested *Sigmoids*."[2] The slight distinction in titles (*sigmoids* vs. *saturators*) implies a fundamental difference in the systems under consideration. A system with a smooth feedback was studied in the first paper [1] whereas one with a nonsmooth feedback in the second [2]. Although in both cases global stability was proved using Lyapunov functions with a similar structure (a sum of quadratic and integral terms), the proofs differed significantly, as can be easily seen by comparing the two papers. Chronologically, the authors first succeeded in proving stability for the system with a smooth feedback in the form of nested sigmoids. A natural issue arose about the possibility of generalizing the result to the case of a nonsmooth feedback (in the form of nested saturators) and/or a discontinuous feedback

[1] A saturator is a nonsmooth saturation function $\mathrm{sat}_p(\cdot)$, $p > 0$, defined as follows: $\mathrm{sat}_p(w) = w$ for $|w| \leqslant p$, and $\mathrm{sat}_p(w) = p\, w/|w|$ for $|w| > p$.

[2] A sigmoid is a smooth, strictly increasing, odd, and scalar function $\sigma(x)$ satisfying the following conditions: (a) $\sigma(x) \to \pm 1$ as $x \to \pm\infty$, (b) $\max_x \sigma'(x) = \sigma'(0)$, (c) $\sigma'(0) = 1$, and (d) $\sigma(x) = -\sigma(-x)$, $x \in R$.

(obtained, e.g., by replacing the external saturator or sigmoid with $\mathrm{sgn}(x)$). However, the proof of global stability in [1] turned out to be inapplicable to the nonsmooth system because, unlike the smooth case, the derivative of the corresponding Lyapunov function is not negative definite in the entire space, so the presentation in [1] was limited to the smooth case. Later, after the submission of [1], the authors managed to establish global stability for the nonsmooth case as well, showing that the domains where the Lyapunov function is identically zero cannot contain entire trajectories. The new results were published in [2].

To date, the case of a discontinuous system (see various physical examples of such systems in [3]), as well as smooth saturation functions [4–6], remains unexplored. This type of control functions was first described in [4, 5], but in both cases, only local stabilization was proved. Their main difference from the saturation function $\mathrm{sat}(x)$ is the use of a nonlinear function instead of the linear part, such that the derivative exists and is continuous at the extreme points of the linear segment. This difference does not affect the proof of global stability given in [2], and therefore the results of [2] can be used for functions of this type as well.

This paper is devoted to the second, more complex case, i.e., the global stability analysis of a system closed by a discontinuous feedback. To prove global stability, we will employ a candidate Lyapunov function consisting of the sum of quadratic and integral terms [2]:

$$V(x) = \frac{1}{2}x_2^2 + \int\limits_0^{x_1} \mathrm{sat}(k_3\mathrm{sat}(k_1\xi))d\xi.$$

As k_3 tends to infinity, the function $V(x)$ takes the form

$$V(x) = \frac{1}{2}x_2^2 + |x_1|. \tag{1}$$

Function (1) is nonsmooth, so Filippov's theorem [7] cannot be directly applied to it. However, it is possible to utilize his proof of global stability for a particular switching system. The idea of the proof is as follows. First, the non-positivity of the derivative outside the switching surface and outside the discontinuity surface of the derivative of the Lyapunov function is shown. Then, the derivatives of the Lyapunov function on the switching curves are computed. Next, trajectories and point singularities where the derivative of the Lyapunov function is zero are considered, and the Barbashin–Krasovskii theorem [8, 9] is applied. An alternative approach was proposed by Matrosov [10, 11] and further developed by Molchanov [12].[3] It consists in generalizing Lyapunov's stability theorems to the case of convex piecewise linear Lyapunov functions, but its application for planar systems with a known nonlinearity is unreasonable.

For the further presentation, the following definitions from [7] will be needed. Consider an equation (or system) in the vector form

$$\dot{x} = f(t, x) \tag{2}$$

with a piecewise continuous function f in a domain G[4]; $x \in R^n$ and $\dot{x} = \frac{dx}{dt}$; finally, $M \in R^{n+1}$ is a null set (a set of measure zero) of discontinuity points of the function f in x.

For each point (t, x) of the domain G, a set $F(t, x)$ in the n-dimensional space is specified. If at a point (t, x) the function f is continuous, then the set $F(t, x)$ consists of a single point coinciding with the value of f at that point. If (t, x) is a discontinuity point of f, then the set $F(t, x)$ is defined in one way or another.

[3] Matrosov's original works were published earlier than [12], but their availability for interested readers is very limited. Therefore, we refer to the modern collections [10, 11] containing various Matrosov's research, which were released significantly later than the original ones.

[4] G is a finite domain of the $(n+1)$-dimensional space of t, x. The domain G consists of a finite number of domains G_i, $i = 1, \ldots, l$, in each of which the function f is continuous up to the boundary.

Definition 1 [7]. A solution of equation (2) is a solution of the differential inclusion

$$\dot{x} \in F(t, x), \tag{3}$$

i.e., an absolutely continuous vector function $x(t)$ defined on an interval or segment I such that $\dot{x}(t) \in F(t, x(t))$ almost everywhere on I.

Definition 2 [7]. The simplest convex extension (completion). For each point $(t, x) \in G$, let $F(t, x)$ be the smallest convex closed set containing all limit values of the vector-function $f(t, x^*)$ when $(t, x^*) \notin M$, $x^* \to x$, $t = \text{const}$. A solution of equation (2) is a solution of inclusion (3) with the just constructed $F(t, x)$. Since M is a null set, for almost all $t \in I$ the measure of the cross-section of M by the plane $t = \text{const}$ is zero. For such t, the set $F(t, x)$ is defined for all $(t, x) \in G$.

At the continuity points of f, the set $F(t, x)$ consists of a single point $f(t, x)$, and the solution satisfies equation (2) in the usual sense. If a point $(t, x) \in M$ lies on the boundaries of the cross-sections of two or more domains $G_1, \ldots, G_k$ by the plane $t = \text{const}$, then the set $F(t, x)$ is a segment, convex polygon, or polyhedron with vertices $f_i(t, x)$, $i \leqslant k$, where $f_i(t, x) = \lim_{(t, x^*) \in G_i, x^* \to x} f(t, x^*)$, and the limit always exists by construction according to Definition 1.

Definition 3 [7]. A solution $x = \chi(t)$ $(t_0 < t < \infty)$ of the differential inclusion (3) with $I = (t_0, \infty)$ is said to be stable if for each $\varepsilon > 0$ it is possible to find $\delta > 0$ such that for each x_0, $|x_0 - \chi(t_0)| < \delta$, each solution $x(t)$ with the initial condition $x(t_0) = x_0$, $t_0 < t < \infty$, exists and satisfies the inequality $|x(t) - \kappa(t)| < \varepsilon$ $(t_0 < t < \infty)$.

The asymptotic stability of (3) is defined by analogy, but with the additional condition $x(t) - \chi(t) \to 0$ as $t \to \infty$.

Definition 4 [7]. Consider a system with control u :

$$\dot{x} = f(t, x, u), \quad u = u(x) \in U(t, x). \tag{4}$$

A solution is a pair of functions, i.e., an absolutely continuous function $x(t)$ and a measurable function $u(x(t))$, that satisfy system (4) almost everywhere on a given interval.

To consider the set of all solutions of system (4), it can be replaced by the differential inclusion (3), where $F(t, x) = f(t, x, U(t, x))$. Then the stability of a solution $x(t) = \chi(t)$ of inclusion (3) means that for $|x(t) - \chi(t)| < \delta$ the solution $x(t)$ of the equation $\dot{x} = f(t, x, U(t, x))$ satisfies equation (4) for all admissible controls $u(t)$.

Definition 5 [7]. A point c is said to be stationary if $x(t) \equiv c$ is a solution of inclusion (3).

Definition 6 [7]. The distance between points or sets is denoted by r :

$$\begin{aligned} &r(a, b) = \|a - b\| = \sqrt{(a_1 - b_1)^2 + \ldots + (a_n - b_n)^2}, \\ &r(a, B) = \inf_{b \in B} r(a, b), \quad r(A, B) = \inf_{a \in A, b \in B} r(a, b). \end{aligned}$$

A δ-neighborhood M^δ of a set M is the set of points x such that $r(x, M) < \delta$.

Definition 7 [7]. A set $M \subset R^n$ (not necessarily stationary) is said to be stable if for any $\varepsilon > 0$ it is possible to find $\delta > 0$ such that each solution $x(t)$ with an initial condition $x(t_0)$, $t_0 \leqslant t < \infty$, from the δ-neighborhood of the set M exists and satisfies the inequality $r(x(t), M) < \varepsilon$, $t_0 \leqslant t < \infty$, where $r(x, M)$ is the distance from the point x to the set M.

Definition 8 [7]. A stationary set is said to be stable in the large if it is stable and each solution infinitely approaches this set as $t \to \infty$. A stationary set is said to be pointwise stable in the large if it is stable and each solution tends to the stationary point as $t \to \infty$.

This paper addresses only planar systems, i.e., $n = 2$ in the above definitions. A vector on the plane will be written as $x = [x_1; x_2] \in R^2$, where x_1 and x_2 are the projections of the vector onto the coordinate axes, respectively.

The inner product of two vectors x and y will be designated by $x \cdot y = [x_1; x_2] \cdot [y_1; y_2] = x_1 y_1 + x_2 y_2$.

2. SMOOTH SWITCHING CURVE

Consider the problem of stabilizing the second-order integrator

$$\dot{x}_1 = x_2,\ \dot{x}_2 = u(x) \tag{5}$$

using a discontinuous feedback of the form

$$u(x) = -\text{sgn}(S(x_1, x_2)), \tag{6}$$

where $\text{sgn}(w) = 1$ for $w > 0$, $\text{sgn}(w) = -1$ for $w < 0$, and $\text{sgn}(w) = [-1, 1]$ for $w = 0$. The function $S(x_1, x_2)$ defines a switching (discontinuity) curve dividing the phase plane (x_1, x_2) into domains G^- $(S < 0)$ and G^+ $(S > 0)$. According to definitions (3) and (4), in this case, $U(t, x) = [-1, 1]$, and a solution of system (5), (6) is a solution of the differential inclusion (3), where $F(t, x) = [x_2; -\text{sgn}(S(x_1, x_2))]$.

Let us partition the phase space R^2 into sets (see Figs. 1–3) so that $R^2 = D_1 \cup D_2^+ \cup D_2^- \cup D_3^+ \cup D_3^-$, where

- $D_1 = \{(x_1, x_2) : S(x_1, x_2) = 0\}$ is the set of points lying on the discontinuity curve;
- $D_2^+ = \{(x_1, x_2) : S(x_1, x_2) > 0 \,\text{and}\, x_1 < 0\}$;
- $D_2^- = \{(x_1, x_2) : S(x_1, x_2) < 0,\ x_1 > 0\}$;
- $D_3^+ = \{(x_1, x_2) : S(x_1, x_2) > 0 \,\text{and}\, x_1 \geqslant 0\}$;
- $D_3^- = \{(x_1, x_2) : S(x_1, x_2) < 0,\ x_1 \leqslant 0\}$.

Then, in the new designations, $G^- = D_2^- \cup D_3^-$ and $G^+ = D_2^+ \cup D_3^+$.

Theorem 1. *Let*

$$S(x_1, x_2) = x_2 + \phi(x_1), \tag{7}$$

where a continuously differentiable function $\phi(\xi)$, $\xi \in R$, satisfies the following properties:

1) $0 \leqslant \dfrac{d\phi(\xi)}{d\xi} < +\infty$;
2) $\exists\, 0 < \delta \leqslant 1$ *such that* $\dfrac{d\phi(\xi)}{d\xi} > 0,\ |\xi| < \delta$;
3) $\phi(0) = 0$;
4) $\phi(\xi) = -\phi(-\xi)$.

Then system (5)–(7) *with the function $\phi(x_1)$ is globally asymptotically stable.*

Remark 1. The discontinuity surface for the control law is chosen in the form (7) to avoid solvability and uniqueness problems in the system with a discontinuous right-hand side [13]. That is, the switching curve for system (5), (6) has the form $x_2 = -\phi(x_1)$.

Proof. Let us prove that function (1) is a Lyapunov function for system (5)–(7).

1. First, it is necessary to show that $\dot{V}(x) \leqslant 0$ for $S \neq 0$ and $x_1 \neq 0$. Differentiating $V(x)$ along the trajectories of system (5), (6) with $x_1 \neq 0$ yields

$$\dot{V} = -x_2 \text{sgn}(x_2 + \phi(x_1)) + \text{sgn}(x_1) x_2 = -[\text{sgn}(x_2 + \phi(x_1)) - \text{sgn}(\phi(x_1))] x_2. \tag{8}$$

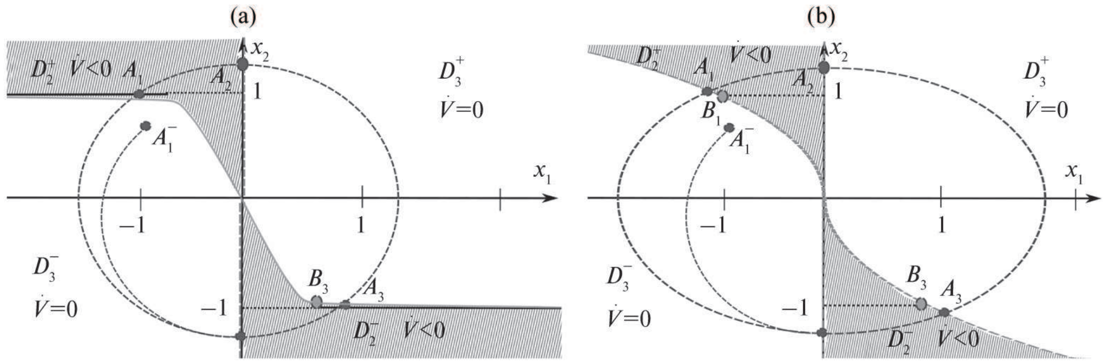

Fig. 1. The partition of the phase plane R^2 into the sets D_1, $D_2^{\pm}$, and $D_3^{\pm}$ for a smooth discontinuity surface: (a) sigmoid and (b) the polynomial function $y = -|\xi|^{1/n-1}\xi$, $n > 1$.

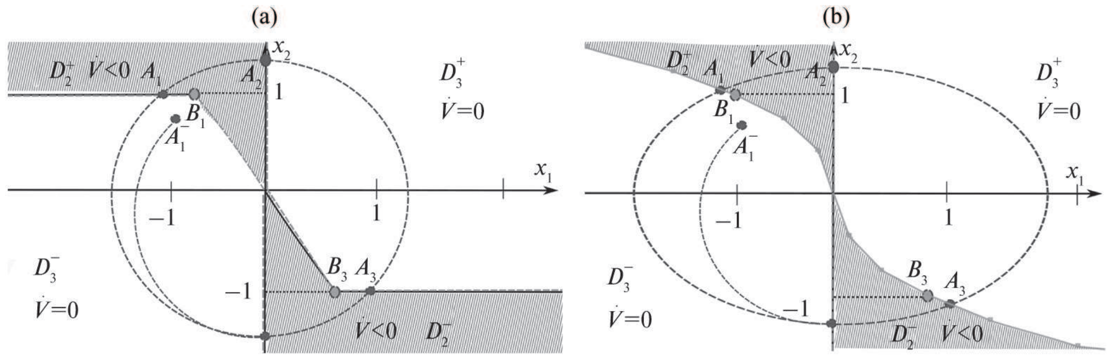

Fig. 2. The partition of the phase plane R^2 into the sets D_1, $D_2^{\pm}$, and $D_3^{\pm}$ for a continuous discontinuity surface: (a) saturator and (b) the piecewise-linear approximation of the polynomial function $y = -|\xi|^{1/n-1}\xi$, $n > 1$.

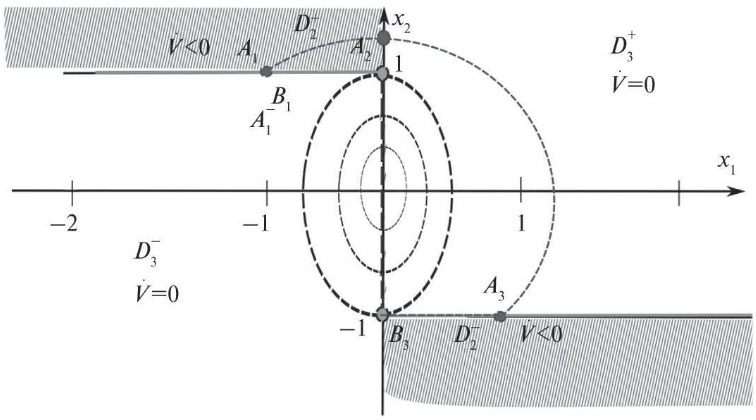

Fig. 3. The partition of the phase plane R^2 into the sets $D_1 = D_1^0 \cup D_1^+ \cup D_1^-$, and $D_2^{\pm}$, $D_3^{\pm}$ for a discontinuous switching function.

Due to the monotonicity of sgn, the inequality $[\mathrm{sgn}(s+s_0) - \mathrm{sgn}(s_0)]s \geqslant 0$ is valid for all $(s+s_0) \neq 0$ and all $s_0 \neq 0$, implying the non-positivity of the derivative: $\dot{V}(x) \leqslant 0$ outside the switching curve for $x_1 \neq 0$. The domains where the strict inequality holds are shaded in Figs. 1 and 2, and they correspond to subsets of the domains $D_2^{\pm}$.

2. Next, we establish that the set $\dot{V}(x) = 0$ does not contain entire trajectories for $S \neq 0$ and $x_1 \neq 0$. The derivative of the function $V(x)$ vanishes for $x_2 = 0$ on the sets D_3^+ and D_3^-, where both bracketed terms are simultaneously equal to $+1$ and -1, respectively. Obviously, these sets cannot contain entire trajectories. Indeed, the trajectories of the system in D_3^- and D_3^+ are the parabolas

$$x_1 = \mp\frac{1}{2}x_2^2 + C. \tag{9}$$

Since no parabola can lie entirely in D_3^- or D_3^+ (see Fig. 1), and the motion occurs with a constant acceleration, the system inevitably reaches the set D_1 (the switching curve) in finite time. Now assume that $x_2 = 0$, then $x_1 = c_1$. Let $c_1 > 0$, then integrating the system gives $x_2 = c_1 t + c_2 = 0$. The latter is true only if $c_1 = 0$ for $t > 0$. This contradiction means that there are no entire trajectories on the set $x_2 = 0$ as well.

3. To proceed, we demonstrate that $\dot{V}(x) < 0$ for $S \neq 0$ and $x_1 = 0$. On this set, $V(x) = x_2^2/2$, and the derivative along the trajectories of the system is $\dot{V}(x) = -x_2 \operatorname{sgn}(x_2 + \phi(x_1))_{x_1=0} = -|x_2| < 0$, $x_2 \neq 0$.

4. Now, we prove that $\dot{V}(x) \leqslant 0$ for $S = 0$ and $x_1 \neq 0$. First, it is necessary to extend the vector field to the discontinuity curve. For this purpose, we construct the convex hull for each pair of vectors on the discontinuity curve. To utilize Filippov's result, the system should be written in new designations to find the projections of the vector field onto the normal at each point of the switching curve. The right-hand side of system (5)–(7) has the vector form $f^+ = [x_2; -1]$ for $S > 0$ and $f^- = [x_2; 1]$ for $S < 0$, with the null discontinuity set in the definition being $M = D_1$. Since S is a smooth surface, let us find the projections of the vector field f_N^+ $(S > 0)$ and f_N^- $(S < 0)$ onto the normal to the discontinuity curve at each point. Using Filippov's formulas [7], we write the equations

$$\begin{aligned} f_N^{\pm}(x_1) &= \frac{\nabla S(x_1) \times f^{\pm}(x_1)}{\|\nabla S(x_1)\|} = -\left(\rho(x_1) \pm 1\right)/\|\nabla S(x_1)\|, \\ \rho(x_1) &= \phi(x_1)\frac{d\phi(x_1)}{dx_1}, \end{aligned} \tag{10}$$

$$\nabla S(x_1) = \left[\frac{d\phi(x_1)}{dx_1}; 1\right], \quad \|\nabla S(x_1)\| = \sqrt{1 + \left(\frac{d\phi(x_1)}{dx_1}\right)^2} \neq 0. \tag{11}$$

By construction of system (5), (6) (only one physical equilibrium and one switching curve), the following point and linear singularities are possible here [7]:

1) $f_N^- f_N^+ > 0$: traversal of the surface (crossing the line), and it is not a linear singularity (case AA_0), e.g., in [14];
2) $f_N^- > 0$, $f_N^+ < 0$, and $f^0 \neq [0; 0]$: a stable sliding mode in which the system moves according to the right-hand side $f^0 = \alpha f^+ + (1 - \alpha) f^-$, $\alpha = f_N^-/(f_N^- - f_N^+) \in (0, 1)$ (case AA_1), e.g., in [15];
3) $f_N^- f_N^+ < 0$ and $f^0 = [0; 0]$: the curve consists of stationary points (case AA_2), e.g., a pendulum with dry friction [16];
4) $f_N^- < 0$, $f_N^+ > 0$, and $f^0 \neq [0; 0]$: an unstable sliding mode, the system constantly leaves the discontinuity surface and cannot pierce it (case AA_3).

First, we show that case AA_2 is impossible for system (5), (6) with the vector field projections (10), (11). In the case under consideration, the vector f_0 has the form

$$f^0 = [-\phi(x_1); 1 - 2\alpha]. \tag{12}$$

For any $0 \leqslant \alpha \leqslant 1$ and $x_1 \neq 0$, $f^0 \neq [0; 0]$, i.e., case AA_2 is impossible since the closed-loop system has no additional stationary points besides the origin.

Let us prove the impossibility of case AA_3 as well. Assume that $f_N^- < 0$; it follows from (10) that $\rho(x_1) > 1$ and, by the monotonicity and oddness of $\phi(x_1)$, we have $f_N^+ < 0$, i.e., case AA_0. Similarly, the assumption $f_N^+ > 0$ leads to $f_N^- > 0$ and, hence, case AA_0. Case AA_3 is impossible.

Now we obtain the conditions of a sliding mode. The above formula for α yields

$$\alpha = \frac{1}{2}(1 - \rho(x_1)). \tag{13}$$

By (13) and the definition of a sliding mode, the latter is possible only for $|\rho(x_1)| < 1$. Clearly, in this case, we have $f_N^- > 0$ and $f_N^+ < 0$.

For $\alpha = 0$ and $\alpha = 1$, $\rho(x_1) = \pm 1$, and for a point $x_1 > 0$ ($x_1 < 0$) we have $f_N^- > 0$ and $f_N^+ = 0$ ($f_N^- = 0$ and $f_N^+ < 0$, respectively). In both cases, this is a point singularity of type *ab* [7], meaning the start or end of a sliding mode. In the case under consideration, this point singularity is not an equilibrium because the magnitude of the vector field at this point is nonzero, and the right-hand side of the system is uniquely defined: either f^+ or f^-.

Thus, in the closed-loop system, only two modes are possible on the switching curve: 1) and 2). A traversal of the discontinuity curve means the continuity of trajectories on it; consequently, at these points, the derivative of the Lyapunov function can be extended by continuity, i.e., in view of items 1–3, it can only be $\dot{V}(x) \leqslant 0$.

It remains to find $\dot{V}$ for $S = 0$. We compute f_0 using (12), (13) and form a new system of differential equations governing the motion of the original system on the discontinuity curve. According to the definition [7], it has the form $\dot{x} = f_0$ and can be written as

$$\dot{x}_1 = -\phi(x_1), \quad \dot{x}_2 = \rho(x_1). \tag{14}$$

After eliminating x_2, the derivative of function (1) along the trajectories of system (14) is given by

$$\dot{V} = -\rho(x_1)\phi(x_1) - \operatorname{sgn}(x_1)\phi(x_1) = -\phi(|x_1|)\left(\phi(|x_1|)\frac{d\phi(x_1)}{dx_1} + 1\right) < 0, \quad x_1 \neq 0. \tag{15}$$

If $\frac{d\phi(x_1)}{dx_1} = 0$, it follows that $x_2 = \text{const}$ and x_1 constantly decreases. This motion will continue until the derivative $\frac{d\phi(x_1)}{dx_1}$ becomes positive, after which the variable x_2 will also start decreasing.

5. The positivity of the function $V(x)$ in the entire space R^2, except for the origin, is obvious, and $V(x)$ tends to infinity as $\|x\| \to \infty$. Thus, the function $V(x)$ satisfies all conditions of the Barbashin–Krasovskii theorem [8, 9] (LaSalle's theorem [17]): the origin is an asymptotically stable equilibrium of system (5)–(7) in the large. The proof of Theorem 1 is complete.

Remark 2. The Lyapunov function (1) for system (5), (6) is not a Lur'e–Postnikov function [18] because system (5), (6) cannot be reduced to the Lur'e form. (A linear part cannot be separated in it.)

Remark 3. In the problem under consideration, the classical approaches to constructing Lyapunov functions for discontinuous systems [19, 20] allow proving only either local stability [21–23] or stability in the large [10, 24].

3. CONTINUOUS SWITCHING CURVE

Theorem 2. *Let us define the set* $Q = R \backslash I_L$, *where* $I_L = \{-\xi^k, -\xi^{k-1}, \ldots, -\xi^1, \xi^1, \ldots, \xi^{k-1}, \xi^k,$ $\xi^k \in R,\ k = 1, \ldots, L,\ L \in Z^+\}$. *Assume that a continuously differentiable function* $\phi_1(\xi)$, $\xi \in Q$, *satisfies the following properties:*

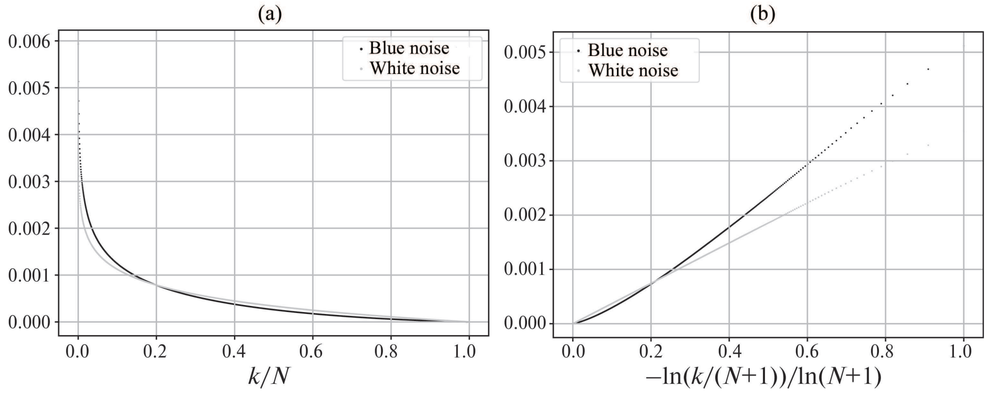

Fig. 4. Examples of functions satisfying the conditions of Theorem 1 or Theorem 2.

1) $0 \leqslant \dfrac{d\phi_1(\xi)}{d\xi} < +\infty$;
2) $\exists\, 0 < \delta \leqslant \xi^1$ *such that* $0 < \dfrac{d\phi_1(\xi)}{d\xi}$, $|\xi| < \delta$;
3) $\phi_1(0) = 0$;
4) $\phi_1(\xi) = -\phi_1(-\xi)$.

In addition, assume the validity of the equality $\phi_1(\xi) = -\phi_1(-\xi)$, $\xi \in I_L$. *Then system* (5)–(7) *with the function* $\phi(x_1) = \phi_1(x_1)$ *is globally asymptotically stable.*

Proof. Since the function $\phi_1(x_1)$ is not smooth, the switching curve $x_2 = -\phi_1(x_1)$ will be non-smooth. Thus, the difference between the functions $\phi(x_1)$ in Theorems 1 and 2 is the absence of derivatives at particular points ξ^k, $k = 1, \ldots, L$, on a null set. Since system (5)–(7) is planar, each such point on the discontinuity curve can be interpreted as the intersection set of only two smooth discontinuity curves, and the vector field can be extended using Fillipov's method in the case of several smooth discontinuity surfaces [7]. We choose any point from the set I_L, e.g., with subscript l; the projection of the vector field $f^{\pm}_{N_l}$ at this point of the phase space, $(x^l_1, -\phi_1(x^l_1))$, is composed of the projections computed for the two adjacent smooth segments of the discontinuity curve. For $\epsilon > 0$, let $f^{\pm}_{N_{l-1}} = f^{\pm}_N(x^l_1 - \epsilon)$ and $f^{\pm}_{N_{l+1}} = f^{\pm}_N(x^l_1 - \epsilon)$ denote the projections of the vector field computed by formulas (10) to the left and right, respectively, of the discontinuity point. Then the projections of the vector field are given by

$$f^{\pm}_{N_l} = \frac{1}{2}(f^{\pm}_{N_{l-1}} + f^{\pm}_{N_{l+1}}). \tag{16}$$

Due to the continuity of the solution, the global stability of the origin fails only if this point is stationary, i.e., under the conditions of case 3) in Theorem 1. More precisely put, we need to check that $f^0_l = \alpha_l f^+ + (1 - \alpha_l) f^- \neq [0; 0]$, $\alpha_l = f^-_{N_l}/(f^-_{N_l} - f^+_{N_l})$. Substituting the corresponding function into (12) yields

$$f^0_l = [-\phi_1(x^l_1); 1 - 2\alpha_l]. \tag{17}$$

Since $x^l_1 \neq 0$ by the conditions of the theorem, equality (17) implies $f^0_l \neq [0; 0]$ for any $0 \leqslant \alpha_l \leqslant 1$. Hence, the point $(x^l_1, -\phi_1(x^l_1))$ is not stationary. Theorem 1 applies to the remaining points. The proof of Theorem 2 is complete.

The characteristic trajectories of systems (5)–(7) with the function $\phi(x_1)$ and $\phi_1(x_1)$ are schematically shown by dashed lines in Figs. 1 and 2, respectively. In both cases, starting the dynamics from the point $A_1 \in D^+_2$ to the point $A_2 \in D^+_3$, the inequality $V(A_1) > V(A_2)$ holds by the negativity of the derivative of the Lyapunov function. Then the system evolves along a parabola in the domain D_3^+, where $\dot{V}=0$, i.e., the inequality $V(A_1) > V(A_3) = V(A_2)$ is true with $A_3 \in D_1$. Next, two options are possible: the trajectory traverses the discontinuity curve and reaches D_2^- or a stable sliding mode and moves along it to the point B_3. At the final point, the strict inequality $V(A_1) > V(B_3)$ holds.

Consider some examples of functions satisfying the conditions of Theorem 1. In Fig. 4, the smooth saturation functions $\Phi_1(\xi) = \sin(\psi(k\xi))$ with $\psi(k\xi) = \mathrm{sat}_{(\frac{\pi}{2k})^{1/3}}(k\xi^3)$ and $\Phi_2(\xi) = \sin(\psi(k\xi))$ with $\psi(k\xi) = \mathrm{sat}_{\xi^*}(\xi^3 + \xi/4)$ are indicated by nos. 1 and 2, respectively. In the latter formula, ξ^* stands for the solution of the polynomial equation $\xi^3 + \xi/4 = \pi/2$.

As an example of a function $\phi_1(x_1)$ satisfying the conditions of Theorem 2, one can take a piecewise-linear function with L bends. For instance, in Fig. 4, no. 3 corresponds to the function $\Phi_3(\xi) = \frac{1}{3}(\mathrm{sat}(k_1\xi) + \mathrm{sat}(k_2\xi) + \mathrm{sat}(k_3\xi))$, $k_1 = 0.3$, $k_2 = 2.65$, $k_3 = 5$, with three bends, and no. 4 to the function $\Phi_4(\xi) = \frac{1}{4}\left(\sum_{i=1}^{4} \mathrm{sat}(k_i\xi)\right)$, $k_i = \frac{i}{3}$, $i = 1, \ldots, 4$, with four bends. The simplest function of this class is sat, the function indicated by no. 5 in Fig. 4.

4. DISCONTINUOUS SWITCHING CURVE

Theorem 3. *Let*

$$\phi_2(\xi) = K\mathrm{sgn}(\xi), \quad \xi \in R, \quad 0 < K < +\infty. \tag{18}$$

System (5)–(7) *with the function* $\phi(x_1) = \phi_2(x_1)$ *has a globally semi-stable cycle passing through the points* $(0, \pm K)$, *and the origin is a sewn center according to Filippov's classification.*

Proof. By analogy with Theorem 2, we first divide the curve $S = 0$ into three smooth segments: $D_1^+ = \{(x_1, x_2) : S(x_1, x_2) = 0, x_1 > 0\}$, $D_1^- = \{(x_1, x_2) : S(x_1, x_2) = 0, x_1 < 0\}$, and $D_1^0 = \{(x_1, x_2) : x_2 \in (-K, K), x_1 = 0\}$.

Since these are straight lines, their normals can be written explicitly: $n^+ = n^- = [0; 1]$ for $D_1^\pm$ and $n^0 = [1; 0]$ for D_1^0. (By definition, the normal direction is chosen towards the set $S > 0$.)

By substituting $\nabla S = n_0$ into (10), we find the projections of the vector field on the set D_1^0 :

$$f_{n_0}^\pm = [1; 0][x_2; \pm 1]_{x_2 \in D_1^0} = x_2, \quad x_2 \in D_1^0. \tag{19}$$

From (19) it follows that $f_{n_0}^+ f_{n_0}^- > 0$, $x_2 \in D_1^0 \backslash 0$. In other words, case AA_0 (see Theorem 1) occurs on the set D_1^0, except for $x_2 = 0$. Thus, the right-hand side of the system is $f = f^+$ for $x_2 > 0$ and $f = f^-$ for $x_2 < 0$ on the set D_1^0, except for $x_2 = 0$. We choose any point on the set D_1^0, e.g., $(0, \delta)$, $0 < \delta < K$. Since system (5)–(7), (18) has an extended right-hand side, it can be integrated from this point. Straightforward transformations bring us to the parabola (9) with $C = \pm\delta^2/2$. The same phase curve is satisfied by the point $(0, -\delta)$ with the opposite sign but the same-magnitude projection of the vector field onto the discontinuity curve. Therefore, in Filippov's terminology [7], the origin is a center. The first part of Theorem 1 is proved.

Now let $\nabla S = [0; 1]$. Using (10), we find the projections of the vector field on the set D_1^+ :

$$f_n^\pm = [0; 1][x_2; \pm 1] = \pm 1 \in D_1^+. \tag{20}$$

On the set D_1^-, we obtain the same result, meaning that a sliding mode occurs on these sets and, moreover, $\alpha_n = \frac{1}{2}$. Let us compute the new vector field f^0 :

$$f_n^0 = \frac{1}{2}[x_2; -1] + \frac{1}{2}[x_2; 1] = [x_2; 0], \quad x_2 \in D_1^\pm. \tag{21}$$

Thus, $f_n^0 = [1; 0]$ on the set D_1^-, and $f_n^0 = [-1; 0]$ on the set D_1^+, and its magnitude does not change. Next, we compute the projection of the vector field at the points $(0, -K)$ and $(0, K)$. For the point $x^l = (0, K)$ in the notation of Theorem 2, we apply formula (16) with the corresponding values of the vector field projections on the normals from the sets D_1^+ and D_1^0 to get

$$f_{N_K}^{\pm} = \frac{1}{2}(x_2 \pm 1)|_{x_2=K} = K \pm 1, \quad \alpha = \frac{1}{2}(1 - K). \tag{22}$$

For the point $(0, K)$,

$$f_K^0 = \frac{1}{2}(1 - K)[x_2; -1] + \frac{1}{2}(1 + K)[x_2; 1] = [K; K]. \tag{23}$$

Obviously, $f_K^0 \neq [0; 0]$, so the point $(0, K)$ is not stationary.ng The above considerations repeated for the point $(0, -K)$ give $f_{-K}^0 = [-K; -K] \neq [0; 0]$, so point $(0, -K)$ is not stationary as well. We construct the solution passing through these points, which has the form (9) with the parameter $C = \pm K^2/2$. Sewing the two parabolas along the x_2 axis yields a closed trajectory representing a limiting cycle [7]. Let us show its stability for all points on the plane outside it. Consider function (1) for the domain outside the cycle. Assume that an arbitrary trajectory starts at the point $A_1 \in D_2^+$ outside the cycle (Fig. 3). By Theorem 1, the derivative of the Lyapunov function at the point $A_1 \in D_2^+$ is negative, vanishing at the point $A_2 \in D_3^+$; hence, the inequality $V(A_1) > V(A_2)$ holds. Since $\dot{V} = 0$ in D_3^+, the system further moves along a parabola to the point $A_3 \in D_1$. In view of the first inequality, we obtain $V(A_3) = V(A_2) < V(A_1)$, i.e., the trajectory has approached the cycle in the sense of the norm of $V(x)$. On the discontinuity curve A_3B_3, a stable sliding mode with the vector field (21) occurs, and hence the trajectory arrives at the point $(0, -K)$, reaching the cycle in finite time without crossing the switching curve. The points from D_3^- are considered similarly. Thus, the cycle of two sewn parabolas is stable for all points on the plane outside it. The proof of Theorem 3 is complete.

5. NUMERICAL EXAMPLES

As an illustration, the level lines of the Lyapunov function (1) were constructed for system (5)–(7) with the functions $\phi = x_1^{1/3}$, $\phi_1 = \mathrm{sat}(x_1)$, and ϕ_2 (18) with $K = 1$. Figure 5 shows one of the level lines (solid line) and several phase trajectories (dashed lines) starting on it. The initial points of the trajectories are marked with circles. According to the figure, none of the trajectories leaves the invariant set bounded by the level line. The trajectory segments evolving along the boundary of the set lie in the subsets D_3^- and D_3^+, where the derivative of the Lyapunov function along the trajectories of the system is zero.

In Fig. 5a, after the trajectory reaches the discontinuity curve, a stable sliding mode occurs until a certain time; and the trajectory subsequently reaches a domain where the discontinuity curve is crossed and a domain where the Lyapunov function's derivative becomes negative, and then the situation repeats. In this case, the origin is a sewn stable focus [7].

In Fig. 5b, after the trajectory reaches the discontinuity curve, a stable sliding mode occurs up to zero. In this case, the origin is a sewn stable node or a linear singularity of type AA_1 [7].

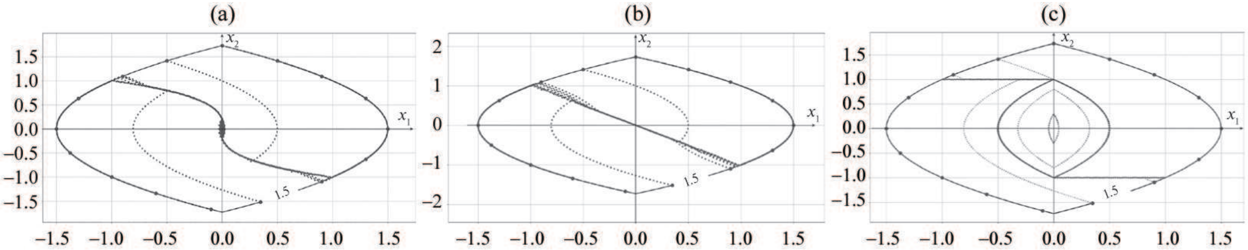

Fig. 5. The level line of the Lyapunov function and phase trajectories on the plane (x_1, x_2). The points indicate the initial conditions corresponding to the phase trajectories plotted.

In Fig. 5c, after the trajectory reaches the discontinuity curve, a stable sliding mode occurs up to the point $(0, 1)$ or $(0, -1)$ (depending on the initial conditions); then the trajectories wind onto the cycle. Unfortunately, due to rounding errors, the trajectory may evolve along a cycle of smaller size. If a point inside the cycle is chosen, another cycle of smaller size is obtained, and so on down to zero. In this case, the origin is a sewn center [7].

6. CONCLUSIONS

This paper has considered the problem of stabilizing a chain of two integrators by a discontinuous feedback. This is a limiting case of nested saturation functions, the feedback studied previously. A Lyapunov function for the closed-loop system has been proposed and then utilized, together with the results of Filippov's theory, to prove the global asymptotic stability of the origin for different types of continuous switching curves. In the case where the switching curve is discontinuous, it has been established that the origin is a sewn center according to Filippov's classification; in addition, the closed-loop system has a semi-stable cycle enclosing this equilibrium. The main result of the work has been formulated as three theorems. As numerical examples, the level line of the Lyapunov function and phase trajectories on the plane have been plotted for different types of switching curves.

REFERENCES

1. Morozov, Y.V. and Pesterev, A.V., Global Stabilization of a Chain of Two Integrators by a Feedback in the Form of Nested Sigmoids, *J. Comput. Syst. Sci. Int.*, 2024, vol. 63, no. 3, pp. 385–389.
2. Pesterev, A.V. and Morozov, Yu.V., Global Stabilization of a Chain of Two Integrators by a Feedback in the Form of Nested Saturators, *Autom. Remote Control*, 2024, vol. 85, no. 4, pp. 391–395.
3. Ponomarev, V.M. and Litvinov, A.P., *Osnovy avtomaticheskogo regulirovaniya i upravleniya* (Fundamentals of Automatic Regulation and Control), Moscow: Vysshaya Shkola, 1974.
4. Olfati-Saber, R., Nonlinear Control of Underactuated Mechanical Systems with Application to Robotics and Aerospace Vehicles, *PhD Dissertation*, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2001.
5. Matyukhin, V.I. and Pyatnitskii, E.S., Controllability of Mechanical Systems in the Class of Controls Bounded Together with Their Derivatives, *Autom. Remote Control*, 2004, vol. 65, no. 8, pp. 1187–1209.
6. Hua, M.-D. and Samson, C., Time Sub-Optimal Nonlinear PI and PID Controllers Applied to Longitudinal Headway Car Control, *Int. J. Control*, 2011, vol. 84, no. 10, pp. 1717–1728.
7. Filippov, A.F., *Differential Equations with Discontinuous Right-Hand Sides*, Kluwer Academic Publishers, 1988.
8. Barbashin, E.A., *Vvedenie v teoriyu ustoichivosti* (Introduction to Stability Theory), Moscow: Nauka, 1967.
9. Krasovskii, N.N., *Nekotorye zadachi teorii ustoichivosti dvizheniya* (Some Problems of the Theory of Stability of Motion), Moscow: Fizmatlit, 1959.
10. Matrosov, V.M., *Metod vektornykh funktsii Lyapunova: analiz dinamicheskikh svoistv nelineinykh sistem* (The Method of Vector Lyapunov Functions: Analysis of Dynamic Properties of Nonlinear Systems), Moscow: Fizmatlit, 2001.
11. Krasovskii, N.N., *Problemy ustoichivosti i upravleniya. Sbornik nauchnykh statei, posvyashchennyi 80-letiyu akademika Vladimira Mefod'evicha Matrosova* (Problems of Stability and Control. Collection of Scientific Articles Dedicated to the 80th Anniversary of Academician Vladimir Mefod'evich Matrosov), Moscow: Fizmatlit, 2013.
12. Molchanov, A.P. and Pyatnitskii, E.S., Lyapunov Functions Determining the Necessary and Sufficient Conditions for Absolute Stability of Nonlinear Nostationary Control Systems. III, *Autom. Remote Control*, 1986, vol. 47, no. 5, pp. 620–630.

13. Matrosov, I.V., On Right-Hand Uniqueness of Solutions to Nondegenerated Algebro-Differential Equations with Discontinuances, *Autom. Remote Control*, 2007, vol. 68, no. 1, pp. 9–17.

14. Utkin, V.I., *Skol'zyashchie rezhimy i ikh primeneniya v sistemakh s peremennoi strukturoi* (Sliding Modes and Their Applications in Variable Structure Systems), Moscow: Nauka, 1974.

15. Utkin, V.I., *Skol'zyashchie rezhimy v zadachakh optimizatsii i upravleniya* (Sliding Modes in Optimization and Control Problems), Moscow: Nauka, 1981.

16. Magnus, K., *Vibrations*, London–Glasgow: Blackie & Son, 1965.

17. La Salle, J. and Lefschetz, S., *Stability by Lyapunov's Direct Method with Applications*, New York: Academic Press, 1961.

18. A. I. Lur'e, A.I. and Postnikov, V.N., On the Theory of Stability of Controlled Systems, *Prikl. Mat. Mekh.*, 1944, no. 3, pp. 246–248.

19. Kim, D.P., *Teoriya avtomaticheskogo upravleniya. Tom 2. Mnogomernye, nelineinye, optimal'nye i adaptivnye sistemy* (Automatic Control Theory. Vol. 2. Multidimensional, Nonlinear, Optimal, and Adaptive Systems), Moscow: Fizmatlit, 2016.

20. Khalil, H.K., *Nonlinear Systems*, 3rd ed., Upper Saddle River: Prentice Hall, 2002.

21. Li, Y. and Lin, Z., *Stability and Performance of Control Systems with Actuator Saturation*, Basel: Birkhauser, 2018.

22. Zubov, V.I., *Ustoichivost' dvizheniya. Metody Lyapunova i ikh primenenie* (Stability of Motion. Lyapunov's Methods and Their Application), Moscow: Vysshaya Shkola, 1984.

23. Rouche, N., Habets, P., and Laloy, M., *Stability Theory by Liapunov's Direct Method*, New York: Springer-Verlag, 1977.

24. Matyukhin, V.I., *Upravlenie mekhanicheskimi sistemami* (Control of Mechanical Systems), Moscow: Fizmatlit, 2009.

This paper was recommended for publication by L.B. Rapoport, a member of the Editorial Board

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 1, pp. 89–96.
© The Author(s), 2026 published by Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 2026.
Russian Text © The Author(s), 2026, published in Avtomatika i Telemekhanika, 2026, No. 1, pp. 120–130.

On Some Properties of Spline Grids in Classification Problems

V. G. Babikov[*,a] **and A. A. Galyaev**[*,b]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a] *babikov@ipu.ru,* [b] *galaev@ipu.ru*

Received August 3, 2025
Revised September 24, 2025
Accepted October 6, 2025

Abstract—This work supplements the authors' series of publications on the topic of detection and classification of weak signals. Continuing this theme, the current study presents results of an applied nature. Here the authors focus on adaptive metric grids applied to information diagrams and the development of new classification features.

Keywords: information entropy, discrete Fourier transform, statistical disequilibrium, statistical complexity, spectral complexity, metric grid

DOI: 10.7868/S1608303226010078

1. INTRODUCTION

Classical methods for solving detection problems use additional information about signal properties. For instance, solving the optimal filtering problem requires knowledge of signal properties: periodicity, frequency band, etc. [1]. Solving the problem of distinguishing between two hypotheses relies on the Neyman–Pearson lemma, determines the fact of exceeding an optimal threshold for a given false alarm probability, and requires estimating the statistical properties of the sample distributions of noise and the signal-noise mixture [2]. Reference [3] presents a variation of the Neyman–Pearson lemma, whose peculiarity is that it does not allow errors when accepting one hypothesis instead of another. Solving the change-point problem requires tuning the algorithm to changes in the unknown statistical characteristics of the distributions of noise and the signal-noise mixture, as does the anomaly search problem [4]. All listed methods demonstrate qualitative and reliable performance when the signal exceeds the noise, but at low signal-to-noise ratios or in heavy noise environments, they often give incorrect answers.

Regarding the use of information characteristics for solving problems in statistical physics, classifying medical tests (e.g., for studying breast histopathology image textures), etc., this approach already has a long history. As early as 1999, [5] introduced a complexity measure based on the average information gain for quantitatively describing the complexity of two-dimensional patterns. [6] raises issues related to developing a general measure of complexity — or regularity, or structure — for two-dimensional systems. Such tasks arise, for example, when studying surfaces in geology.

Article [7] is devoted to studying the possibility of texture classification via two-dimensional ordinal patterns, particularly via the statistical complexity diagram. In [8], the authors confirm that the statistical complexity diagram is a popular tool for distinguishing stochastic signals (noise) from deterministic chaos (in English-language literature, the term for this tool is the complexity-entropy (CE) plane). However, the authors discovered that both high-dimensional deterministic time series and stochastic surrogate data (noises) can be located in the same region of the diagram, and their representations show very similar behavior for different signal-noise mixtures or even signal types. Therefore, classifying this data based on their position in the diagram plane can be difficult or even misleading.

The question of the complexity of distinguishing noise and noise-like deterministic signals is also covered in detail in [9]. That work is devoted to studying statistical and spectral complexity diagrams. The diagrams are constructed as follows: information entropy is plotted on the abscissa axis, statistical (spectral) complexity on the ordinate axis. In the text, statistical complexity diagrams are denoted as (S, C_{TV}) and spectral complexity diagrams as (S, C_S). The work provides relevant definitions, makes significant remarks about diagram sensitivity, and formulates and proves lemmas on estimating upper and lower bounds on statistical and spectral complexity diagrams for various signal-noise mixtures. Some important patterns of behavior of noise-like and weak signals, and possibilities for their detection under white and blue noise conditions, are identified.

In the listed and numerous other works devoted to two-dimensional information analysis methods using information diagrams, the possibility of solving detection and classification problems is demonstrated. While the authors' article [9] presents an alternative interpretation of these methods, which is that statistical complexity diagrams (like some other two-dimensional information diagrams) are in one-to-one correspondence with the number of discrete spectral components of the signal and the signal-to-noise ratio. Quite unexpectedly, it turns out that by calculating the values of information entropy and statistical complexity for the obtained signal-noise mixture, one can simultaneously determine the number of effective spectral discretes and the energy shares of the signal and noise. Nevertheless, this is the case — the article consistently shows, via the construction of analytical grids, that statistical and spectral complexity diagrams can be reduced to a two-dimensional diagram of discreteness and signal-to-noise ratio. Specifically, the mentioned work presents analytical estimates of metric grids for statistical and spectral complexity diagrams. The tool of metric grids (under certain conditions) allows determining a unique correspondence between the tuple of information characteristics (S, C_{TV}) and the tuple of parameters (d – number of discretes in the spectrum, SNR – signal-to-noise ratio). However, when constructing metric grids, it is first necessary to determine what kind of noise is present in the signal-noise mixture (white noise, blue noise, or some other noise). Natural noise, of course, differs from deterministic white, blue, and other noises. In this regard, the construction of so-called adaptive grids is required. This is discussed in the "Adaptive Grids" section of this work.

2. DIAGRAMS OF SPECTRAL AND STATISTICAL COMPLEXITY

In previously published works on information characteristics, the authors focused mainly on new sensitive information criteria, which is important for solving the detection problem [10–12], as well as on constructing metric grids, which play a role in solving the classification problem [3, 9].

The new spectral measure $C_S(p)$ (1), proposed in [10], is especially effective for detecting weak signals with a small number of discretes ($d < 50$ for $N \geqslant 2048$) in a mixture with white noise. [9] provides relevant definitions (essentially, d is the number of harmonic oscillations of equal amplitude in the so-called d-signal). Below, expressions (1)–(4) present a number of notations required for solving the detection/classification problem of weak signals:

$$C_S(p) = -\frac{4}{\log_2 N}\left(\sum_{k=1}^{N} p_k \log_2 p_k\right)\left(\sum_{k=1}^{N} p_k + \frac{K_N}{\ln N} + 1\right), \tag{1}$$

$$C_{TV}(p) = -\frac{4}{\log_2 N}\left(\sum_{i=1}^{N} p_i \log_2 p_i\right)\left(\sum_{i=1}^{N} p_i - \frac{1}{N}\right)^2. \tag{2}$$

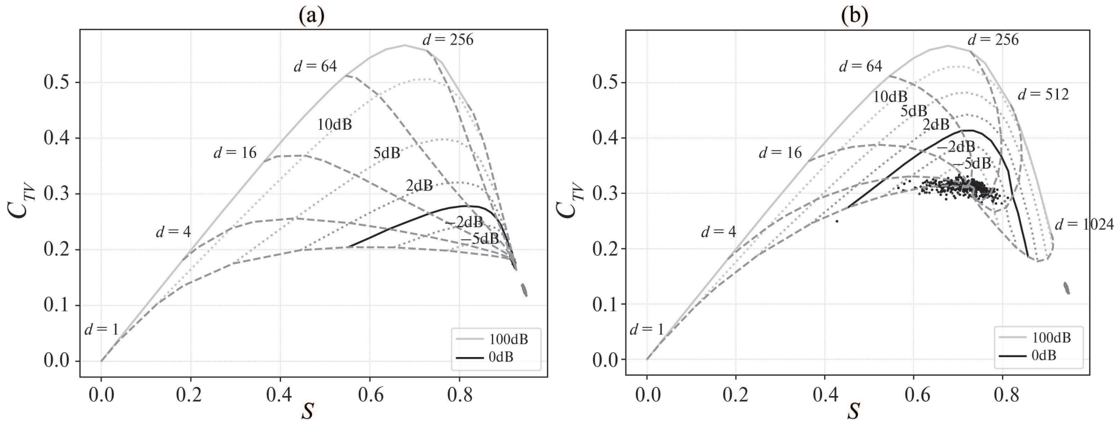

Fig. 1. Comparison of metric grids for blue and pink noises for $N = 2048$ (for white noise: $S = 0.945$, $C_{TV} = 0.125$); (a) – grid for blue noise, (b) – grid for pink noise.

In turn, statistical complexity is the product of information entropy $S(p)$ and statistical disequilibrium D_{TV} ($C_{TV}(p) = S(p) \cdot D_{TV}(p)$):

$$D_{TV}(p) = \frac{N}{4}\left(\sum_{i=1}^{N} p_i - \frac{1}{N}\right)^2, \tag{3}$$

$$S(p) = -\frac{1}{\log_2 N}\left(\sum_{i=1}^{N} p_i \log_2 p_i\right), \tag{4}$$

where N is the length of the spectral data series, p_k is the power spectrum, K_N is a normalization factor close to one and depending on N.

Previously published works have shown that statistical and spectral complexity diagrams are an effective tool for solving detection and classification problems. Moreover, information diagrams based on spectral complexity $C_S(p)$ are less applicable to solving the classification problem than information diagrams based on statistical complexity $C_{TV}(p)$ (2). Thus, an information metric that has an advantage in solving the detection problem is inferior to another metric when solving the classification problem. Therefore, the authors further develop the information diagram method based on statistical complexity, which is more sensitive for determining the number of discretes in the signal spectrum.

Work [9] is devoted to constructing analytical grids for statistical and spectral complexity diagrams (white noise). Note that analytical grids have a great advantage — the possibility of constructing inverse continuous functions and, as a consequence, obtaining a quick result for high-precision estimation of discretes d and the signal-to-noise ratio from the obtained estimates of entropy and statistical complexity. In cases where the noise differs significantly from white, and the signals are weak, it is necessary to build adaptive grids.

Figure 1 shows examples of metric grids on the statistical complexity diagram under blue and pink noise conditions. Note that under pink noise compared to blue or white noise, it becomes possible to “discern” noise-like signals, i.e., classify signals with a large number of discretes, particularly short pulses. Thus, in Fig. 1a (blue noise), signals with the number of discretes $d > 256$ are practically indistinguishable, while in Fig. 1b (pink noise), it becomes possible to distinguish signals with the number of discretes $d > 1024$, i.e., when the number of spectral components of the signal is larger than the size of the entire spectrum (i.e., when the signal is noise-like).

It should also be said that alternative diagrams constructed based on entropy and statistical disequilibrium D_{TV} (3) allow in most cases to more effectively separate discretes and solve the classification problem than diagrams based on statistical complexity C_{TV}. This is because when

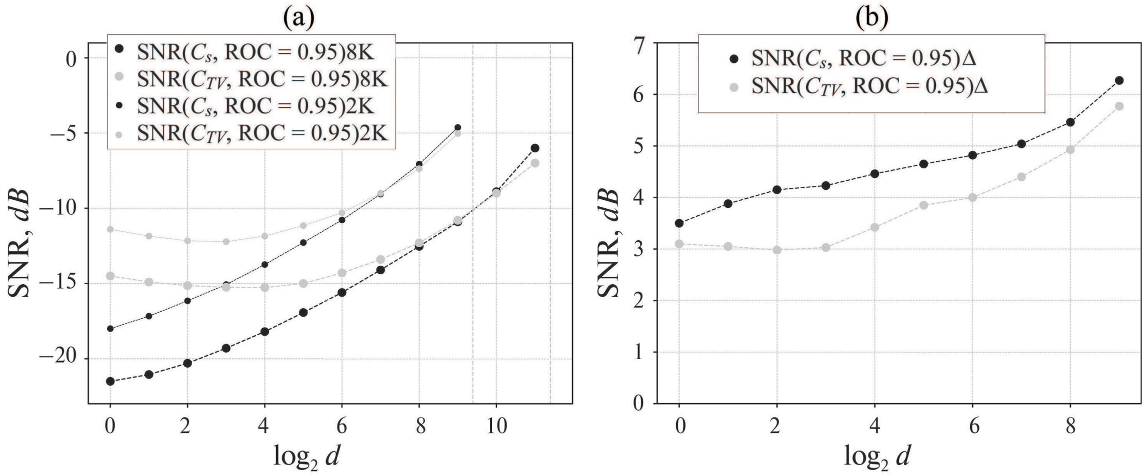

Fig. 2. Comparison of ultimate signal detection capabilities depending on the number of discretes d and spectrum size N; (a) – limiting SNR levels depending on discreteness (AUC = 0.95), (b) – effect of increasing dimensionality (8K vs 2K), change in limiting SNR.

a signal appears in white noise, entropy S (4) and disequilibrium D_{TV} change in opposite directions, and when multiplied ($C_{TV} = S \cdot D_{TV}$), they partially suppress each other. The authors' works consider statistical complexity diagrams because this method is known and often used in practice. Subsequent research will also be devoted to diagrams based on entropy and statistical disequilibrium.

For solving detection and classification problems in practice, the construction of so-called adaptive grids is required. The point is that when solving applied problems, the researcher faces the problem of determining the type of noise, which can differ significantly from white, blue, etc. It is important to have a tool that allows quickly adapting the metric grid to real conditions. This is discussed in more detail later in the article.

Separately, it is worth noting that the spectrum size N determines the capabilities for estimating the number of discretes in the spectrum. Previously, in [3], an estimate was given for determining the maximum number of discretes in white noise: $d < N/4$. Figure 2 presents a comparison of the ultimate capabilities for signal detection via statistical and spectral complexities for $N = 2048$ and for $N = 8192$ depending on the number of signal discretes in white noise. Figure 2b shows a comparison of white metric grids for $N = 2048$ and $N = 8192$. The graphs show that the capabilities for classification and detection of signals using metric grids increase with N [9]. The gray curve demonstrates a decrease in SNR for signals with a small number of discretes by approximately 3 dB when increasing the spectrum size from $N = 2048$ to $N = 8192$ for the statistical complexity diagram and by approximately 4 dB for the spectral complexity diagram. This is an important observation because it becomes clear that with increasing N, the sensitivity of the spectral complexity diagram grows faster than the sensitivity of the statistical complexity diagram. The result of the dependence of the limiting SNR on the number of discretes in the spectrum was obtained via numerical experiment (ROC = 0.95).

3. ADAPTIVE GRIDS

In practice, the construction of so-called adaptive grids is required, which can be obtained as follows: from accumulated observations along the bearings of a phased array where a minimum of energy is observed (conditionally, no signal), the average entropy S and statistical complexity C_{TV} are estimated. Then, noise is generated such that in each window, the mathematical expectation of entropy equals S, and the mathematical expectation of statistical complexity equals C_{TV}. This noise is mixed with simple signals of different numbers of discretes (a combination of sinusoidal

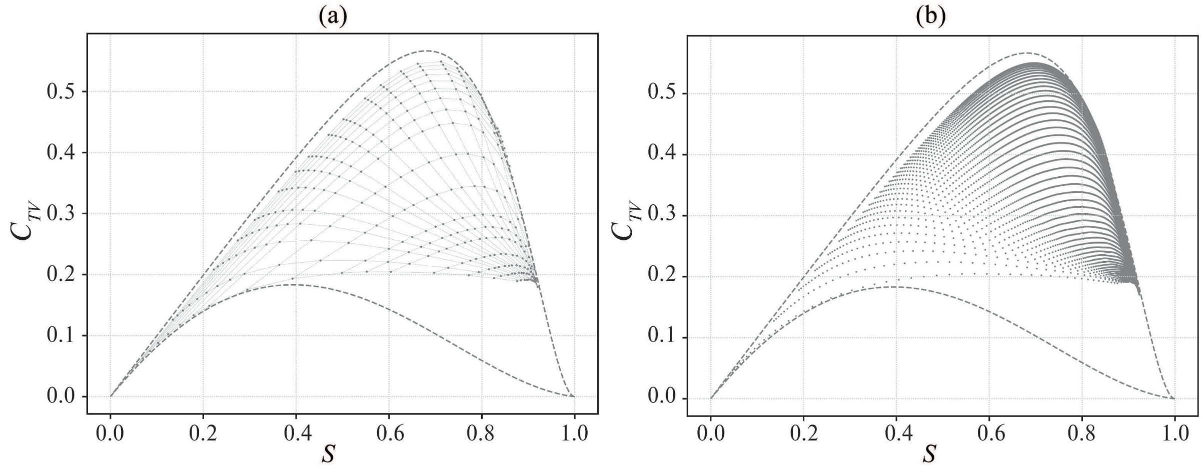

Fig. 3. (a) Estimation of blue grid nodes and (b) calculation of spline grid for the statistical complexity diagram ($N = 2048$).

signals of equal amplitude) to obtain d-signals with given characteristics. Then, by averaging over 200 cases, the entropy and complexity of each node of the metric grid are estimated (each node is characterized by the number of discretes d and the signal-to-noise ratio level SNR). Then, a spline grid is constructed based on the nodes of the coarse grid (obtaining a detailed grid). This method of constructing metric grids is mentioned in [9] for comparing analytical grids with numerical experiments.

Figure 3a presents a metric grid for the statistical complexity diagram ($N = 2048$) for detecting and classifying signals in blue noise, constructed according to the algorithm described above for adaptive grids. The grid nodes are defined by the following discreteness values and signal-to-noise ratios: $d \in \{1, 2, 4, 6, 8, 12, 16, 20, 28, 36, 52, 68, 96, 128, 192, 256, 384, 512\}$, $\mathrm{SNR} \in \{-10, -7, -5, -3, -1, 0, 1, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$. Each node is built by averaging over 200 cases (calculating the grid took about 20 minutes). The Appendix provides an example of constructing an analytical metric grid for blue noise.

Figure 3b presents a metric spline grid, which is built on the basis of the coarse grid 3,a (the figure shows a grid of size 500×50) using two-dimensional interpolation RectBivariateSpline (Python also has a simpler analogue, interp2d). Two-dimensional spline interpolation when constructing a metric grid significantly increases its level of detail (under limited calculation time), which, in turn, allows estimating discreteness and the signal-to-noise ratio with high accuracy [9].

4. FEATURES GENERATED BY DISCRETENESS AND SIGNAL-TO-NOISE RATIO

The classification problem is associated with various information criteria. As shown in [9], through information characteristics such as entropy and statistical complexity, one can obtain estimates of the discreteness level d and the signal-to-noise ratio SNR, which, in turn, generate a whole family of strong classification criteria. Let's focus on some of them.

1) **Correlation level of d and SNR.** A high level of correlation between discreteness and the signal-to-noise ratio within a certain time interval may indicate a change in the distance to the object.
2) **Difference in discreteness** (absolute and relative) between the mean discreteness value d of the signal-noise mixture, measured over k windows (duration of each window τ), and the D estimate of the discreteness of the signal-noise mixture, measured over the entire time window $k\tau$: $D - d$, $D/d - 1$. The level of this difference allows determining how stable the discrete frequencies are. Such a criterion can be useful, for example, in determining heart rate variability.

3) **Discreteness levels and signal-to-noise ratio** for the signal-noise mixture after subtracting the spectral envelope. More on methods of envelope subtraction and line spectrum extraction is discussed in [13].
4) **Energy-weighted frequency** of the d main spectral discretes (having the highest energy). This indicator can be useful for estimating the frequency of the peak of cavitation noise.
5) **Spectrum localization level** and other criteria.

Let us explain the concept of "Spectrum localization level" in more detail. As already mentioned, the discreteness diagram [9] has unique properties that allow extracting additional information for solving the classification problem. In particular, according to the discreteness diagram, the researcher gets the opportunity to formalize the search for the number of "effective" discretes d in the spectrum. This, in turn, allows building a distribution for this random variable. Note also that discreteness d and the signal-to-noise ratio, being unique characteristics of the signal, are also the most important classification features, but this is far from all. One of the most important criteria generated by discreteness is the spectrum localization level. The essence of this criterion is to determine the indices of the d discretes that were identified when forming the discreteness criterion d through the application of the Discreteness Diagram tool. Let Z be the total distance of the d discretes with the maximum energy from the mean value over the sample of these d discretes (the distance is defined as the modulus of the difference between the index of the discrete in the unordered spectrum and the mean value of all indices of these d discretes). So, let's define the minimum and maximum possible total distance of d discretes from their mean:

$$Z_{\min}(N,d) = \begin{cases} \dfrac{d^2-1}{4}, & d = 2k+1, \\ \dfrac{d^2}{4}, & d = 2k. \end{cases}$$

$$Z_{\max}(N,d) = \begin{cases} \dfrac{d(N-d+3)}{4} + \dfrac{1}{2}\left(N - d - \dfrac{N}{d} - 2\right), & d = 2k+1,\ k > 1, \\ \dfrac{dN}{4}, & d = 2k. \end{cases}$$

Let's define the frequency localization criterion based on the above formulas (the criterion value interval should be within $(0, 1)$; for this, we use the estimates $Z_{\min}, Z_{\max}$ for even values of discreteness d). The spectrum localization level is the following expression:

$$K_z = \frac{Z - Z_{\min}}{Z_{\max} - Z_{\min}} = \frac{4Z - d^2}{2(Nd - d^2)}.$$

This criterion allows answering the question of how much the spectrum of the observed signal is localized in a separate spectral region. And the average frequency over the d discretes allows determining the localization region. This criterion allows classifying many types of signals. As a practical application, consider this example: for white noise (as well as for broadband interference or an explosion), the value of this criterion is close to 0.5; for marine vessels, depending on deadweight, this indicator varies within the interval 0.1–0.4; and for the calls of some marine animals, this criterion is close to zero.

5. CONCLUSION

This work is the third in the authors' series of works devoted to metric grids for the complexity diagram. The applied nature of the work is primarily related to the concept of an adaptive grid and the identification of a number of informative classification features. Collectively, all these articles form the method of metric grids and constructing the discreteness diagram, the essence of which is to estimate the number of discretes in the spectrum and the signal-to-noise ratio from a single time window under low signal-to-noise ratio conditions.

FUNDING

The work was partially supported by the Russian Science Foundation grant (project no. 23-19-00134).

APPENDIX

Previously, in [10], an estimate of entropy depending on the size of the spectral series N for white noise was obtained:

$$S(N) \approx 1 + \frac{\gamma - 1}{\ln N} \approx 1 - \frac{0.422}{\ln N}, \tag{A.1}$$

where γ is the Euler–Mascheroni constant.

An approximate estimate of entropy for blue noise, which also gives a good approximation, can be obtained using the formula

$$S(N) \approx 1 - \frac{0.616}{\ln N}. \tag{A.2}$$

It has been experimentally confirmed that this formula for blue noise estimates the true entropy value with an accuracy up to the fourth decimal place ($128 < N < 16\,384$). Figure 4 shows the distribution density of normalized ordered power spectra of white and blue noises for $N = 2048$. Formula (A.2) allows obtaining an approximate metric grid for blue noise:

$$S(q,d,N) = S_0(N)(1-q) - \frac{q}{\ln N}, \tag{A.3}$$

$$D_{TV}(q,d,N) = \frac{1}{4}\left(\alpha(d,N) + 1 + q + \beta(1-q^2) + (1-q)\left(-1 + 2\exp(1-q^{-1})\right)\right)^2, \tag{A.4}$$

$$C_{TV}(q,d,N) = S(q,d,N) \cdot D_{TV}(q,d,N), \tag{A.5}$$

where $N \geqslant 1000$ is the size of the spectral series, and $S_0(N)$ is the value of information entropy for blue noise $S_0(N) \approx 1 - \frac{0.616}{\ln N}$ (for $N = 2048$ $S_0 \approx 0.9192$), $\alpha(d,N) = -\frac{2d}{N}$, $\beta \approx 0.12$ is an empirically obtained coefficient, q is the energy share of the d-signal in the signal-noise mixture.

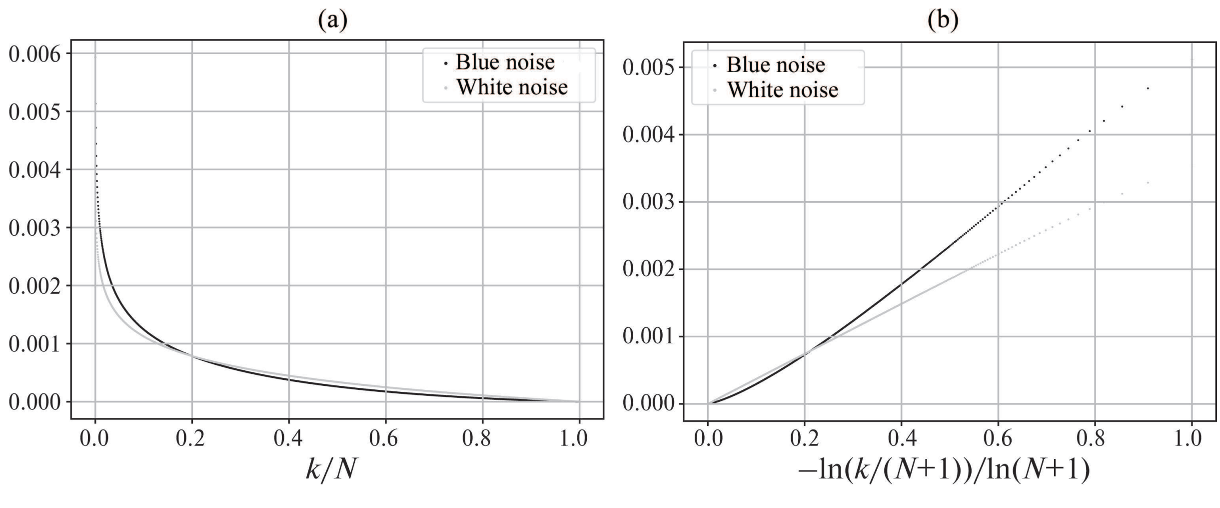

Fig. 4. Explanation for formula (A.1); (a) – ordered spectrum of one realization of a frame of length $N = 2048$ for white and blue noise, (b) – the same realizations for white and blue noises, horizontal axis in logarithmic scale.

REFERENCES

1. Ronald, L.A. and Duncan, W.M., *Signal analysis: time, frequency, scale, and structure*, N.J.: Wiley-IEEE Press, 2004.

2. Shiryaev, A.N., *Probabilistic-statistical methods in decision theory. Modern mathematics*, M.: MTSNMO: NMU, 2020.

3. Babikov, V.G. and Galyaev, A.A., Information diagrams and their capabilities for classifying weak signals, *Probl. Inform. Transmiss.*, 2024, vol. 60, no. 2, pp. 127–140.

4. Kishan, G.M., Chilukuri, K.M., and Huang HuaMing, *Anomaly Detection Principles and Algorithms*, Cham. Springer, 2017.

5. Andrienko, Y.A., Brilliantov, N.V., and Kurths, J., Complexity of two-dimensional patterns, *Eur. Phys. J. B.*, 2000, vol. 15, no. 3, pp. 539–546.

6. Feldman, D.P. and Crutchfield, J.P., Structural information in two-dimensional patterns: Entropy convergence and excess entropy, *Physical Review E.*, 2003, vol. 67, no. 5, pp. 1–12.

7. Bandt, C. and Wittfeld, K., Two new parameters for the ordinal analysis of images, *Chaos*, 2023, vol. 33, no. 7, pp. 1–20.

8. Kottlarz, I. and Parlitz, U., Ordinal pattern-based complexity analysis of high-dimensional chaotic time series, *Chaos*, 2023, vol. 33, no. 7, pp. 1–11.

9. Babikov, V.G. and Galyaev, A.A., Complexity-entropy plane analytical representation, *Probl. Inform. Transmiss.*, 2025, vol. 61, no. 1, pp. 27–40.

10. Galyaev, A.A., Babikov, V.G., Lysenko, P.V., and Berlin, L.M., New spectral measure of complexity and its capabilities for detecting signals in noise, *Dokl. RAN. Matematika, informatika, protsessy upravleniya*, 2024, vol. 518, pp. 80–88.

11. Galyaev, A.A., Babikov, V.G., Lysenko, P.V., and Berlin, L.M., Addition to the article "a new spectral measure of complexity and its capabilities for detecting signals in noise," *Dokl. Math.*, 2024, vol. 518, pp. 89–92.

12. Galyaev, A.A., Berlin, L.M., Lysenko, P.V., and Babikov, V.G., Order Statistics of the Normalized Spectral Distribution for Detecting Weak Signals in White Noise, *Autom. Remote Control*, 2024, no. 12, pp. 49–69.

13. Babikov, O.V. and Babikov, V.G., Estimating the fundamental frequency of a propeller shaft-blade harmonic series using the hilbert transform and autocorrelation, *Mathematical Problems of Control*, 2025, no. 3, pp. 12–23.

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