Volume 87, Number 10,
October 2026

ISSN 0005-1179
CODEN: AURCAT

AUTOMATION AND REMOTE CONTROL

Editor-in-Chief
Andrey A. Galyaev

http://ait.mtas.ru

Automation and Remote Control

Vol. 87, No. 10, October 2026

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, 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. 10, 2026

Linear Systems

Description of Limit Reachable and Null-Controllable Sets for Linear Discrete-Time Systems and Integral Second-Order Constraints

D. N. Ibragimov 863

Optimization of Transients in a Three-Tempo Singularly Perturbed System

A. I. Kalinin and L. I. Lavrinovich 885

Nonlinear Systems

Robust Prescribed Performance Control under Given Constraints for Systems with Relative Degree One

K. A. Lastochkin 901

An Integrated Approach to Fault Accommodation in Fuzzy Discrete-Event Systems

A.E. Shumsky and A.N. Zhirabok 926

Optimization, System Analysis, and Operations Research

Defense Planning for a Moving Target Against Multiple Attackers Using a Defender via the Monte Carlo Tree Search

E. A. Ryabushev and A. A. Galyaev 940

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 10, pp. 863–884.
© 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. 10, pp. 3–30.

LINEAR SYSTEMS

Description of Limit Reachable and Null-Controllable Sets for Linear Discrete-Time Systems and Integral Second-Order Constraints

D. N. Ibragimov

Moscow Aviation Institute (National Research University), Moscow, Russia
e-mail: rikk.dan@gmail.com

Received March 5, 2026
Revised April 13, 2026
Accepted April 15, 2026

Abstract—A discrete-time linear system with second-order integral control constraints is considered. The properties of the limit reachable and null-controllable sets of the system are investigated. Necessary and sufficient conditions for the boundedness of these sets are obtained, including for uncontrollable systems. It is proved that these sets are ellipsoids, when they are bounded, and their support function is constructed explicitly. Numerical simulations are performed.

Keywords: discrete-time systems, controllability, reachability, limit reachable set, limit controllable set, integral constraints on control

DOI: 10.7868/S1608303226100011

1. INTRODUCTION

The study of reachable and null-controllable sets is an actual problem in mathematical control theory. From an applied perspective, these sets can be used to design controllers, test the solvability of boundary value problems, and analyze process quality. For example, this approach is used in the following modern works: [1–5]. When studying controllability and reachability, limit reachable and null-controllable sets are also important. This is because, for systems with constrained control, classical Kalman rank conditions are insufficient to judge the reachability of particular states in an arbitrary finite time.

It should be noted that, based on the mathematical apparatus and methods used to solve analysis and synthesis problems, dynamic systems in control theory are divided into two fundamental classes: discrete-time systems and continuous-time systems. However, due to the development of digital technologies, continuous systems are now often studied by discretization and reduction to discrete analogs [6]. This underscores the relevance of developing effective numerical methods for studying discrete-time systems. For the reasons described above, methods for constructing limit reachable and null-controllable sets are of interest. For discrete time, these methods reduce to calculating the limit of a sequence of compact sets.

For general systems, the only accessible approach to solving this problem is to examine individual states for reachability and controllability, which essentially amounts to a pointwise construction of the desired sets [7–9]. For the class of linear systems, various analytical methods are available, based on the application of convex and functional analysis. A well-established method involves considering the limit reachable and null-controllable sets as fixed points of contractive mappings in the Hausdorff space [10, 11]. This approach allows, on the one hand, to use the simple iteration

method to approach a fixed point, and on the other hand, to construct external estimates of the limit sets with an arbitrary order of accuracy.

The majority of the presented results concern systems with geometric constraints on control. In the case of integral constraints, the situation is more complex: the reachable and null-controllable sets are not polyhedra even under scalar control [12], and their description in recurrent form is impossible using standard mathematical constructions [13]. Currently, in [14], a generalization of the approach based on the principle of contraction mappings [10, 11] is proposed for systems with integral constraints on scalar control. Separately, in [14, Lemma 3], a description of the reachable and null-controllable sets in the form of ellipsoids is obtained for second-order constraints and a diagonalizable matrix of the system.

The goal of this paper is to generalize [14] to the case of linear systems with an arbitrary dynamic matrix and second-order integral constraints on vector control. Specifically, it is necessary to establish necessary and sufficient conditions for the boundedness of the limit reachable and null-controllable sets in terms of the system parameters and to obtain an explicit description of them by a support function or Minkowski functional.

The paper is structured as follows. Section 2 introduces the concepts of limit reachable and null-controllable sets and states the problem. Section 3 formulates necessary and sufficient conditions for the boundedness of these sets. Section 4 describes the structure of limit reachable and null-controllable sets for linear systems with second-order integral constraints. Section 5 demonstrates the theoretical results obtained by numerical simulation.

2. PROBLEM STATEMENT

A linear stationary system with discrete time and integral second-order control constraints (A, B, H, t) is considered:

$$\begin{aligned} &x(k+1) = Ax(k) + Bu(k),\ k \in \mathbb{N} \cup \{0\}, \\ &\sum_{k=0}^{\infty} u(k)^{\mathrm{T}} H u(k) \leqslant t^2, \end{aligned} \tag{1}$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $u(k) \in \mathbb{R}^m$ is the control vector, $H \in \mathbb{R}^{m\times m}$ is a positive definite symmetric matrix, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ are the system matrices, $t > 0$ is the parameter defining the integral constraint.

For system (1), let $\mathcal{Y}_{t,2,\infty}$ denote the limit reachable set, which is the set of those states into which system (1) can be transferred from the origin in a finite number of steps:

$$\mathcal{Y}_{t,2,\infty} = \Bigg\{ x \in \mathbb{R}^n \colon \exists N \in \mathbb{N},\ u(0), \dots, u(N-1) \in \mathbb{R}^m, \sum_{k=0}^{N-1} u(k)^{\mathrm{T}} H u(k) \leqslant t^2,\ x(0) = 0,\ x(N) = x \Bigg\}. \tag{2}$$

Similarly, for system (1) we denote by $\mathcal{X}_{t,2,\infty}$ the limit null-controllable set, which is the set of those states from which system (1) can be transferred to the origin in a finite number of steps:

$$\mathcal{X}_{t,2,\infty} = \Bigg\{ x \in \mathbb{R}^n \colon \exists N \in \mathbb{N},\ u(0), \dots, u(N-1) \in \mathbb{R}^m, \sum_{k=0}^{N-1} u(k)^{\mathrm{T}} H u(k) \leqslant t^2,\ x(0) = x,\ x(N) = 0 \Bigg\}. \tag{3}$$

It is required to obtain an exact description of the sets $\mathcal{Y}_{t,2,\infty}$ and $\mathcal{X}_{t,2,\infty}$ in terms of their Minkowski functionals and support functions, and also to determine necessary and sufficient conditions for boundedness.

Remark 1. The notations for sets (2) and (3) are chosen in accordance with [12, 14]. In these papers, arbitrary ordering of constraints $r \in (1;+\infty)$ is allowed, and similar sets are denoted by $\mathcal{Y}_{t,r,\infty}$ and $\mathcal{X}_{t,r,\infty}$. This paper focuses on the study of an important special case of quadratic constraints defined by a positive-definite quadratic form, which determines the presence of the sign 2 in the indices $\mathcal{Y}_{t,2,\infty}$ and $\mathcal{X}_{t,2,\infty}$.

3. NECESSARY AND SUFFICIENT CONDITIONS FOR THE BOUNDEDNESS OF LIMIT SETS

As demonstrated in [12, Theorems 2, 3] and [15, Theorem 3], the conditions for the boundedness of the limit reachable and 0-controllable sets are determined by the relation between the matrices A and B. Denote by $h_1,\dots,h_n \in \mathbb{R}^n \setminus \{0\}$ the real Jordan basis of the matrix A [16, Section 3.4, Chapter 3], consisting of its eigenvectors and associated vectors or their imaginary and real parts. Consider the following invariant subspaces:

$$\mathbb{L}_{<1} = \mathrm{Lin}\{h\colon h \text{ corresponds to the eigenvalue } \lambda,\ |\lambda|<1\},$$
$$\mathbb{L}_{>1} = \mathrm{Lin}\,\{h\colon h \text{ corresponds to the eigenvalue } \lambda,\ |\lambda|>1\}.$$

Let us group the elements of the real Jordan basis as follows:

$$h_{1,1},\dots,h_{1,n_1},\dots,h_{M,1},\dots,h_{M,n_M}, \tag{4}$$

where $h_{i,j}\in\mathbb{R}^n$ is the adjoint vector of the matrix A of rank j corresponding to the eigenvalue $\lambda_i\in\mathbb{R}$, or $h_{i,j} = (h_{i,j}^{re}\ h_{i,j}^{im})\in\mathbb{R}^{n\times 2}$ are the real and imaginary parts of the adjoint vector of the matrix A of rank j corresponding to the eigenvalue $\lambda_i\in\mathbb{C}$. Then the real Jordan form of the matrix A consists of M cells, and n_i or $2n_i$ is the dimension of the ith cell, depending on the type of the eigenvalue. Moreover, by [16, Theorem 3.4.5], the following representation is true:

$$\begin{aligned} &A = S\Lambda S^{-1},\\ &S = \big(h_{1,1},\dots,h_{1,n_1},\dots,h_{M,1},\dots,h_{M,n_M}\big),\\ &\Lambda = \mathrm{diag}(J_1,\dots,J_M). \end{aligned} \tag{5}$$

For all $i=\overline{1,M}$, the Jordan cell J_i has one of the following two forms:

$$J_i = \begin{pmatrix} \lambda_i & 1 & \cdots & 0\\ 0 & \lambda_i & \ddots & 0\\ \vdots & \vdots & \ddots & 1\\ 0 & 0 & \cdots & \lambda_i \end{pmatrix} \in \mathbb{R}^{n_i\times n_i},\ \lambda_i\in\mathbb{R}, \tag{6}$$

$$J_i = \begin{pmatrix} r_iA_{\varphi_i} & I & \cdots & O\\ O & r_iA_{\varphi_i} & \ddots & O\\ \vdots & \vdots & \ddots & I\\ O & O & \cdots & r_iA_{\varphi_i} \end{pmatrix} \in \mathbb{R}^{2n_i\times 2n_i},\ \lambda_i = r_i\mathrm{e}^{\mathbf{i}\varphi_i}\in\mathbb{C}. \tag{7}$$

Here and below, A_φ and $I\in\mathbb{R}^{2\times 2}$ denote the rotation matrix and the identity matrix, respectively:

$$A_\varphi = \begin{pmatrix} \cos\varphi & \sin\varphi\\ -\sin\varphi & \cos\varphi \end{pmatrix},\ I = \begin{pmatrix} 1 & 0\\ 0 & 1 \end{pmatrix}.$$

The following criterion for the boundedness of the sets $\mathcal{Y}_{t,2,\infty}$ and $\mathcal{X}_{t,2,\infty}$ is valid.

Theorem 1. *Let the sets $\mathcal{Y}_{t,2,\infty}$ and $\mathcal{X}_{t,2,\infty}$ in system* (1) *be defined by relations* (2) *and* (3), *respectively, $B = (b_1, \dots, b_m)$. Then*

1) *$\mathcal{Y}_{t,2,\infty}$ is bounded if and only if $b_1, \dots, b_m \in \mathbb{L}_{<1}$;*

2) *$\mathcal{X}_{t,2,\infty}$ is bounded if and only if $\det A \neq 0$ and $b_1, \dots, b_m \in \mathbb{L}_{>1}$.*

The proof of Theorem 1 and all subsequent statements is given in the Appendix.

Example 1. Let us consider for $n = 3$ and $m = 2$ matrices

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0.9 & 1 \\ 0 & 0 & 0.9 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}^{-1}, \; B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 2 & 0 \end{pmatrix}, \; H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Generally speaking, in the classical sense, for discrete-time system (1), the matrix A is not stable, since it has an eigenvalue $\lambda_1 = 2$, which is greater than 1 in absolute value. At the same time, the following representation is valid:

$$\mathbb{L}_{<1} = \operatorname{Lin}\left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}, \; b_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{L}_{<1}, \; b_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{L}_{<1}.$$

It follows that the condition of item 1 of Theorem 1 is satisfied and the limit reachable set $\mathcal{Y}_{t,2,\infty}$ of the system (A, B, H, t) is bounded.

On the other hand, if for $m = 1$ we choose

$$B = (2, 2, 0)^{\mathrm{T}}, \; H = 1,$$

then the following formulas will be correct:

$$\mathbb{L}_{>1} = \operatorname{Lin}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}, \; b_1 = 2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \in \mathbb{L}_{>1}.$$

Then the system (A, B, H, t) will satisfy the conditions of item 2 of Theorem 1, and the limit null-controllable set $\mathcal{X}_{t,2,\infty}$ will be bounded.

4. DESCRIPTION OF LIMIT REACHABLE AND NULL-CONTROLLABLE SETS IN THE FORM OF ELLIPSOIDS

Theorem 1 defines necessary and sufficient conditions for the boundedness of sets (2) and (3), but does not specify their structure. In the case of boundedness, the sets $\mathcal{Y}_{t,2,\infty}$ and $\mathcal{X}_{t,2,\infty}$ are ellipsoids. A similar result was obtained in [14, Lemma 3], but only in the case of scalar control and a diagonalizable matrix A. For system (1), the problem of constructing matrices of quadratic forms of the ellipsoids $\mathcal{Y}_{t,2,\infty}$ and $\mathcal{X}_{t,2,\infty}$ reduces to the summation of matrix series. Let us present the theoretical information, which is necessary for this.

Lemma 1. *Let in system* (1) $B = (b_1, \dots, b_m)$.

1) *If $b_1, \dots, b_m \in \mathbb{L}_{<1}$, then the following series converges:*

$$H_{\mathcal{Y}} \triangleq \sum_{k=0}^{\infty} A^k B H^{-1} B^{\mathrm{T}} (A^k)^{\mathrm{T}};$$

2) *if* $\det A \neq 0$ *and* $b_1, \ldots, b_m \in \mathbb{L}_{>1}$*, then the following series converges:*

$$H_{\mathcal{X}} \triangleq \sum_{k=1}^{\infty} A^{-k} B H^{-1} B^{\mathrm{T}} (A^{-k})^{\mathrm{T}}.$$

To describe the limit reachable and null-controllable sets of system (1), we generalize the concept of an ellipsoid. The classical definition assumes that for some positive definite symmetric $H \in \mathbb{R}^{m\times m}$ and $t > 0$, an ellipsoid is a set

$$\mathcal{U} = \left\{ u \in \mathbb{R}^m \colon u^{\mathrm{T}} H u \leqslant t^2 \right\}.$$

In this case $\mathcal{U}$ is determined by the Minkowski functional [17, §15, Chapter III].

However, this definition is not suitable for describing ellipsoidal sets whose dimension is less than the dimension of the space. In this case, it turns out to be more convenient to define the set $\mathcal{U}$ in terms of the support function [17, §4, Chapter I]. For a non-negative definite symmetric $\tilde{H} \in \mathbb{R}^{m\times m}$ and $t > 0$, an ellipsoid is defined as a set $\mathcal{U} \subset \mathbb{R}^m$ whose support function is given by the equality

$$\mathcal{S}(p,\mathcal{U}) \triangleq \sup_{u\in\mathcal{U}} p^{\mathrm{T}} u = t\sqrt{p^{\mathrm{T}} \tilde{H} p}. \tag{8}$$

If the matrix $\tilde{H}$ is positive definite and, as a consequence, non-singular, then the representation for the Minkowski functional is also true:

$$\mu(u,\mathcal{U}) \triangleq \inf\{\gamma > 0 \colon u \in \gamma\mathcal{U}\} = \frac{1}{t}\sqrt{u^{\mathrm{T}} \tilde{H}^{-1} u}. \tag{9}$$

Example 2. Consider a two-dimensional circle in three-dimensional space:

$$\mathcal{U} = \left\{ u \in \mathbb{R}^3 \colon u_1^2 + u_2^2 \leqslant 1,\ u_3 = 0 \right\}.$$

The set $\mathcal{U}$ can be considered as an ellipsoid in the general sense, although it is not described by a single inequality for a quadratic form. Nevertheless, the support function $\mathcal{U}$ satisfies the representation

$$\mathcal{S}(p,\mathcal{U}) = \sqrt{p_1^2 + p_2^2} = \sqrt{p^{\mathrm{T}} \tilde{H} p},\ \tilde{H} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Here $H \in \mathbb{R}^{3\times 3}$ is non-negative definite and degenerate, which does not allow us to use it to go to formula (9) for the Minkowski functional.

Theorem 2. *Let* $H_{\mathcal{Y}}, H_{\mathcal{X}} \in \mathbb{R}^{n\times n}$ *be taken from Lemma 1. Then*

1) *if* $\mathcal{Y}_{t,2,\infty}$ *is bounded, then the following representation for the support function is valid:*

$$\mathcal{S}(p, \mathcal{Y}_{t,2,\infty}) = t\sqrt{p^{\mathrm{T}} H_{\mathcal{Y}} p};$$

2) *if* $\mathcal{X}_{t,2,\infty}$ *is bounded, then the following representation for the support function is valid:*

$$\mathcal{S}(p, \mathcal{X}_{t,2,\infty}) = t\sqrt{p^{\mathrm{T}} H_{\mathcal{X}} p}.$$

Theorem 2 states that, in the case of boundedness, $\mathcal{Y}_{t,2,\infty}$ and $\mathcal{X}_{t,2,\infty}$ are ellipsoids. Transition to a description using the Minkowski functional is possible only if the matrices $H_{\mathcal{Y}}$ and $H_{\mathcal{X}}$ are invertible. This property can be guaranteed, for example, in the special case of control systems.

Lemma 2. *Let the controllability condition be satisfied for the system* (1):

$$\operatorname{rank}(B\ AB\ \dots\ A^{n-1}B) = n.$$

Then

1) *if* $b_1, \dots, b_m \in \mathbb{L}_{<1}$, *then* $\det H_{\mathcal{Y}} \neq 0$;
2) *if* $\det A \neq 0$ *and* $b_1, \dots, b_m \in \mathbb{L}_{>1}$, *then* $\det H_{\mathcal{X}} \neq 0$.

Due to Lemma 2, Theorem 2 and expressions (8)–(9), in the case of controlled system (1), the following representation for the Minkowski functional is valid:

$$\mu(x, \mathcal{Y}_{t,2,\infty}) = \frac{1}{t}\sqrt{x^{\mathrm{T}} H_{\mathcal{Y}}^{-1} x}, \quad \mu(x, \mathcal{X}_{t,2,\infty}) = \frac{1}{t}\sqrt{x^{\mathrm{T}} H_{\mathcal{X}}^{-1} x}.$$

Remark 2. The matrices $H_{\mathcal{Y}}$ and $H_{\mathcal{X}}$ are usually called the Gramians of a discrete-time linear system. If the matrix A is Schur, i.e., all its eigenvalues are less than 1 in absolute value, then $H_{\mathcal{Y}}$ is a solution to the discrete Lyapunov equation [18, Theorem K.2]:

$$P = APA^{\mathrm{T}} + BH^{-1}B^{\mathrm{T}}.$$

A large number of numerical methods are known for solving this equation [19, 20]. Also, one common analytical approach is to use the canonical Jordan decomposition of the matrix A [21]. Further in the article, we will refine this approach for a more general case, which is defined by Theorem 1 and allows for matrices A other than Schur matrices, and construct an explicit form for the Gramians $H_{\mathcal{Y}}$ and $H_{\mathcal{X}}$.

Although Theorem 2 defines the structure of sets (2) and (3) of system (1), their construction is reduced to calculating the matrix series from Lemma 1 for $H_{\mathcal{Y}}$ and $H_{\mathcal{X}}$. To calculate these series, we use the canonical Jordan form of the matrix A [16, Theorem 3.1.11]:

$$\begin{aligned}
&A = S\Lambda S^{-1},\\
&S \in \mathbb{C}^{n\times n},\ \det S \neq 0,\ \Lambda = \operatorname{diag}(J_1, \dots, J_M) \in \mathbb{C}^{n\times n},\\
&J_i = \lambda_i I + P_{n_i} \in \mathbb{C}^{n_i\times n_i},\ P_{n_i} = \begin{pmatrix} 0 & 1 & \cdots & 0\\ 0 & 0 & \ddots & 0\\ \vdots & \vdots & \ddots & 1\\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{n_i\times n_i},\ i = \overline{1, M},\\
&n_1 + \dots + n_M = n.
\end{aligned} \tag{10}$$

Here $\lambda_i \in \mathbb{C}$ is an eigenvalue of the matrix A of multiplicity n_i.

In notation (10) for all $i, j = \overline{1, M}$ we define the matrices $D_{ij} \in \mathbb{C}^{n_i\times n_j}$ from the following relations:

$$\begin{pmatrix} D_{11} & \dots & D_{1M}\\ \vdots & \ddots & \vdots\\ D_{M1} & \dots & D_{MM} \end{pmatrix} \triangleq S^{-1}BH^{-1}B^{\mathrm{T}}(S^{-1})^{\mathrm{T}},$$

for $\lambda_i\lambda_j \neq 1$ we introduce matrices

$$\Gamma_{ij} \triangleq \sum_{r=0}^{n_i-1}\sum_{q=0}^{n_j-1}\sum_{l=0}^{\min\{r,q\}} \frac{(r+q-l)!}{(r-l)!(q-l)!l!}\frac{\lambda_i^{r-l}\lambda_j^{q-l}}{(1-\lambda_i\lambda_j)^{r+q-l+1}} P_{n_i}^r D_{ij} (P_{n_j}^q)^{\mathrm{T}}.$$

Theorem 3. *Let $H_{\mathcal{Y}}$ and $H_{\mathcal{X}}$ be taken from Lemma 1. Then*
1) *if $b_1, \ldots, b_m \in \mathbb{L}_{<1}$, then*

$$H_{\mathcal{Y}} = S \begin{pmatrix} \Gamma^{\mathcal{Y}}_{11} & \ldots & \Gamma^{\mathcal{Y}}_{1M} \\ \vdots & \ddots & \vdots \\ \Gamma^{\mathcal{Y}}_{M1} & \ldots & \Gamma^{\mathcal{Y}}_{MM} \end{pmatrix} S^{\mathrm{T}},$$

where for all $i, j = \overline{1, M}$ in the case $|\lambda_i| \geqslant 1$ or $|\lambda_j| \geqslant 1$ it is true that $\Gamma^{\mathcal{Y}}_{ij} = O$, or in the case $|\lambda_i|, |\lambda_j| \in [0; 1)$ it is true that $\Gamma^{\mathcal{Y}}_{ij} = \Gamma_{ij}$;
2) *if $\det A \neq 0$ and $b_1, \ldots, b_m \in \mathbb{L}_{>1}$, then*

$$H_{\mathcal{X}} = S \begin{pmatrix} \Gamma^{\mathcal{X}}_{11} & \ldots & \Gamma^{\mathcal{X}}_{1M} \\ \vdots & \ddots & \vdots \\ \Gamma^{\mathcal{X}}_{M1} & \ldots & \Gamma^{\mathcal{X}}_{MM} \end{pmatrix} S^{\mathrm{T}},$$

where for all $i, j = \overline{1, M}$, in the case $|\lambda_i| \leqslant 1$ or $|\lambda_j| \leqslant 1$, $\Gamma^{\mathcal{X}}_{ij} = O$, or in the case $|\lambda_i|, |\lambda_j| \in (1; +\infty)$, $\Gamma^{\mathcal{X}}_{ij} = -\Gamma_{ij}$.

In general, the matrices in decomposition (10), and therefore the matrices Γ_{ij}, may be complex. However, as a result of the transformations, $H_{\mathcal{Y}}$ and $H_{\mathcal{X}}$ will turn out to be real.

Example 3. Consider the matrices

$$A = \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; H = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

The matrix A has complex conjugate eigenvalues $\lambda_1 = \sqrt{2} + \mathbf{i}\sqrt{2}$, $\lambda_2 = \sqrt{2} - \mathbf{i}\sqrt{2}$ and the following expansion of form (10):

$$A = \begin{pmatrix} -\mathbf{i} & \mathbf{i} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} + \mathbf{i}\sqrt{2} & 0 \\ 0 & \sqrt{2} - \mathbf{i}\sqrt{2} \end{pmatrix} \begin{pmatrix} -\mathbf{i} & \mathbf{i} \\ 1 & 1 \end{pmatrix}^{-1}.$$

Due to Theorem 1, the limit null-controllable set $\mathcal{X}_{t,2,\infty}$ is bounded, and in the notation of Theorem 3 it is true that $M = 2$, $n_1 = n_2 = 1$ and

$$S^{-1} B H^{-1} B^{\mathrm{T}} \left(S^{-1}\right)^{\mathrm{T}} = \begin{pmatrix} \frac{\mathbf{i}}{2} & 1 \\ 1 & -\frac{\mathbf{i}}{2} \end{pmatrix}.$$

This leads to the following values:

$$-\Gamma^{\mathcal{X}}_{11} = \sum_{r=0}^{1-1} \sum_{q=0}^{1-1} \sum_{l=0}^{\min\{r,q\}} \frac{(r+q-l)!}{(r-l)!(q-l)!l!} \frac{\lambda_1^{r-l} \lambda_1^{q-l}}{(1-\lambda_1\lambda_1)^{r+q-l+1}} P_1^r D_{11} (P_1^q)^{\mathrm{T}} = \frac{1}{1-(\sqrt{2}+\mathbf{i}\sqrt{2})^2} \frac{\mathbf{i}}{2},$$

$$-\Gamma^{\mathcal{X}}_{22} = \frac{1}{1-(\sqrt{2}-\mathbf{i}\sqrt{2})^2} \left(-\frac{\mathbf{i}}{2}\right), \; -\Gamma^{\mathcal{X}}_{12} = -\Gamma^{\mathcal{X}}_{21} = \frac{1}{1-(\sqrt{2}+\mathbf{i}\sqrt{2})(\sqrt{2}-\mathbf{i}\sqrt{2})},$$

$$H_{\mathcal{X}} = \begin{pmatrix} \frac{22}{51} & -\frac{1}{17} \\ -\frac{1}{17} & \frac{46}{51} \end{pmatrix}, \; \mathcal{S}(p, \mathcal{X}_{t,2,\infty}) = \sqrt{\frac{22}{51} p_1^2 - \frac{2}{17} p_1 p_2 + \frac{46}{51} p_2^2}.$$

5. NUMERICAL EXPERIMENTS

Let us demonstrate the results of Theorems 1–3 by numerical simulation. Consider various discrete-time systems of form (1) with bounded reachable sets $\mathcal{Y}_{t,2,\infty}$. First, we construct these sets explicitly using Theorem 3. Then, we generate a set of admissible processes and plot their terminal states as elements of the limit reachable set. All numerical values will be rounded to four decimal places.

Example 4. Let us consider system (1) for $n = m = 2$ and the following parameter values:

$$A = \begin{pmatrix} 0.8791 & -1.1321 \\ 0.6429 & 0.0934 \end{pmatrix} = S \begin{pmatrix} 0.4863 + 0.7573\mathbf{i} & 0 \\ 0 & 0.4863 - 0.7573\mathbf{i} \end{pmatrix} S^{-1},$$

$$S = \begin{pmatrix} 0.7986 & 0.7986 \\ 0.2771 - 0.5342\mathbf{i} & 0.2771 + 0.5342\mathbf{i} \end{pmatrix},$$

$$B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \ H = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \ t = 1.$$

The matrix A has a pair of complex conjugate eigenvalues $\lambda_{1,2} = 0.4863 \pm 0.7573\mathbf{i}$, whose absolute values are strictly less than 1. Therefore, by Theorem 1, the set $\mathcal{Y}_{t,2,\infty}$ is bounded, and by Theorem 2, it is an ellipsoid. The matrix $H_{\mathcal{Y}}$ can be calculated explicitly by Theorem 3. The following values are valid:

$$M = 2, \ n_1 = n_2 = 1,$$
$$D_{11} = -1.5609 - 0.2711\mathbf{i}, \ D_{12} = D_{21} = 2.0836, \ D_{22} = -1.5609 + 0.2711\mathbf{i}.$$

This leads to the equality

$$H_{\mathcal{Y}} = S \begin{pmatrix} \dfrac{D_{11}}{1 - \lambda_1\lambda_1} & \dfrac{D_{12}}{1 - \lambda_1\lambda_2} \\ \dfrac{D_{21}}{1 - \lambda_2\lambda_1} & \dfrac{D_{22}}{1 - \lambda_2\lambda_2} \end{pmatrix} S^{\mathrm{T}} = \begin{pmatrix} 12.9553 & 3.9416 \\ 3.9416 & 7.8973 \end{pmatrix}.$$

Limit reachable set (2) and terminal states of various processes are shown in Fig. 1.

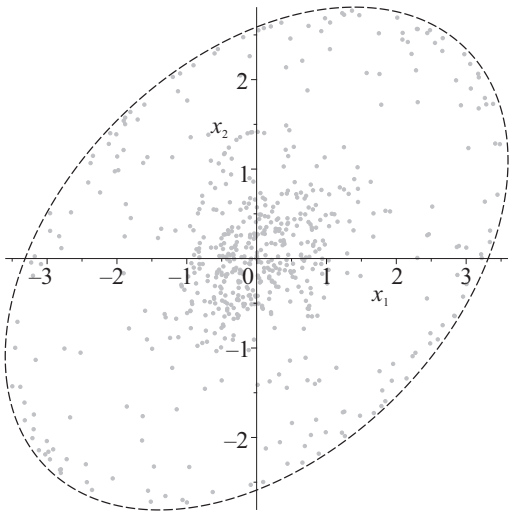


Fig. 1. The set $\mathcal{Y}_{t,2,\infty}$ and various terminal states for the system (A, B, H, t) from Example 4.

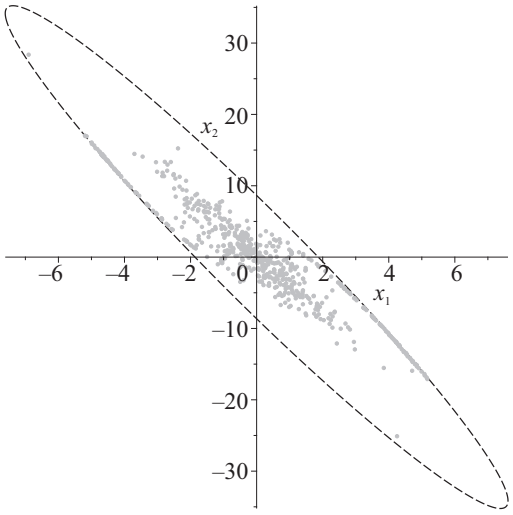


Fig. 2. The set $\mathcal{Y}_{t,2,\infty}$ and various terminal states for the system (A, B, H, t) from Example 5.

Example 5. Let us consider system (1) for $n = m = 2$ and the following values:

$$A = \begin{pmatrix} -1.3315 & -0.0769 \\ 2.7692 & -0.4085 \end{pmatrix} = S \begin{pmatrix} -0.87 & 1 \\ 0 & -0.87 \end{pmatrix} S^{-1},$$
$$S = \begin{pmatrix} -1 & 1 \\ 6 & 7 \end{pmatrix},\ B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},\ H = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix},\ t = 1.$$

The matrix A has one real eigenvalue $\lambda_1 = -0.87$, whose absolute value is strictly less than 1. Therefore, by Theorem 1, the set $\mathcal{Y}_{t,2,\infty}$ is bounded, and by Theorem 2, it is an ellipsoid. The matrix $H_{\mathcal{Y}}$ can be calculated explicitly using Theorem 3. The following values are valid:

$$M = 1,\ n_1 = 2,\ D_{11} = \begin{pmatrix} 0.2975 & -0.2739 \\ -0.2739 & 0.2975 \end{pmatrix}.$$

This leads to the equality

$$H_{\mathcal{Y}} = S \left(\frac{D_{11}}{1-\lambda_1^2} + \frac{\lambda_1 D_{11} P_2^{\mathrm{T}}}{(1-\lambda_1^2)^2} + \frac{\lambda_1 P_2 D_{11}}{(1-\lambda_1^2)^2} + \frac{P_2 D_{11} P_2^{\mathrm{T}}}{(1-\lambda_1^2)^2} + 2\frac{\lambda_1^2 P_2 D_{11} P_2^{\mathrm{T}}}{(1-\lambda_1^2)^3} \right) S^{\mathrm{T}}$$
$$= \begin{pmatrix} 57.9130 & -259.9758 \\ -259.9758 & 1241.6945 \end{pmatrix}.$$

Limit reachable set (2) and terminal states of various processes are shown in Fig. 2.

Example 6. Let us consider system (1) for $n = 3$, $m = 2$, and the following values:

$$A = \begin{pmatrix} -0.8421 & -0.0196 & 0.0976 \\ 0.3013 & -0.9058 & 0.1081 \\ -0.1239 & -0.0093 & -0.9540 \end{pmatrix} = S \begin{pmatrix} -0.8910 + 0.1270\mathbf{i} & 0 & 0 \\ 0 & -0.8910 - 0.1270\mathbf{i} & 0 \\ 0 & 0 & -0.92 \end{pmatrix} S^{-1},$$

$$S = \begin{pmatrix} 0.1703 + 0.3405\mathbf{i} & 0.1703 - 0.3405\mathbf{i} & -0.1562 \\ 0.8513 & 0.8513 & 0.9370 \\ -0.3575 + 0.0511\mathbf{i} & -0.3575 - 0.0511\mathbf{i} & 0.3123 \end{pmatrix},\ B = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 1 & 1 \end{pmatrix},\ H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

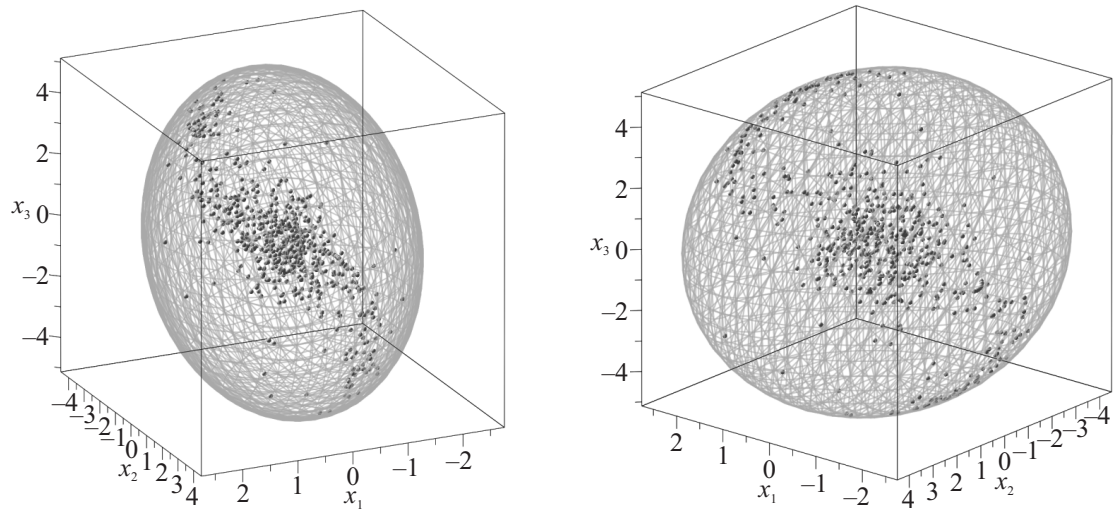


Fig. 3. The set $\mathcal{Y}_{t,2,\infty}$ and various terminal states for the system (A, B, H, t) from Example 6.

We also assume that $t = 1$. The matrix A has a pair of complex conjugate eigenvalues $\lambda_{1,2} = -0.8910 \pm 0.1270\mathbf{i}$ and one real eigenvalue $\lambda_3 = -0.92$, whose absolute values are strictly less than 1. Therefore, by Theorem 1, the set $\mathcal{Y}_{t,2,\infty}$ is bounded, and by Theorem 2, it is an ellipsoid. The matrix $H_{\mathcal{Y}}$ can be calculated explicitly using Theorem 3, similarly to Example 4:

$$M = 3,\ n_1 = n_2 = n_3 = 1,$$

$$H_{\mathcal{Y}} = \begin{pmatrix} 7.5669 & 4.6171 & 1.4047 \\ 4.6171 & 20.2330 & -5.4371 \\ 1.4047 & -5.4371 & 26.3126 \end{pmatrix}.$$

Limit reachable set (2) and terminal states of various processes are shown in Fig. 3.

Example 7. Let us consider system (1) for $n = 3$, $m = 2$, and the following values:

$$A = \begin{pmatrix} -0.4838 & -0.1586 & -0.2412 \\ 1.9715 & 0.2153 & 3.1897 \\ -1.1023 & 0.3068 & -0.5215 \end{pmatrix} = S \begin{pmatrix} 0.95 & 0 & 0 \\ 0 & -0.87 & 1 \\ 0 & 0 & -0.87 \end{pmatrix} S^{-1},$$

$$S = \begin{pmatrix} -1 & 1 & 3 \\ 6 & 7 & 1 \\ 2 & -3 & 0 \end{pmatrix},\ B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 2 \end{pmatrix},\ H = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},\ t = 1.$$

The matrix A has a real non-multiple eigenvalue $\lambda_1 = 0.95$ and a real multiple eigenvalue $\lambda_2 = -0.87$, whose absolute values are strictly less than 1. Therefore, by Theorem 1, the set $\mathcal{Y}_{t,2,\infty}$ is bounded, and by Theorem 2, it is an ellipsoid. The matrix $H_{\mathcal{Y}}$ can be calculated explicitly using Theorem 3, similarly to Example 5:

$$M = 2,\ n_1 = 1,\ n_2 = 2,$$

$$H_{\mathcal{Y}} = \begin{pmatrix} 11.1345 & 52.5427 & -32.4125 \\ 52.5427 & 735.8157 & -266.4128 \\ -32.4125 & -266.4128 & 136.2212 \end{pmatrix}.$$

Limit reachable set (2) and terminal states of various processes are shown in Fig. 4.

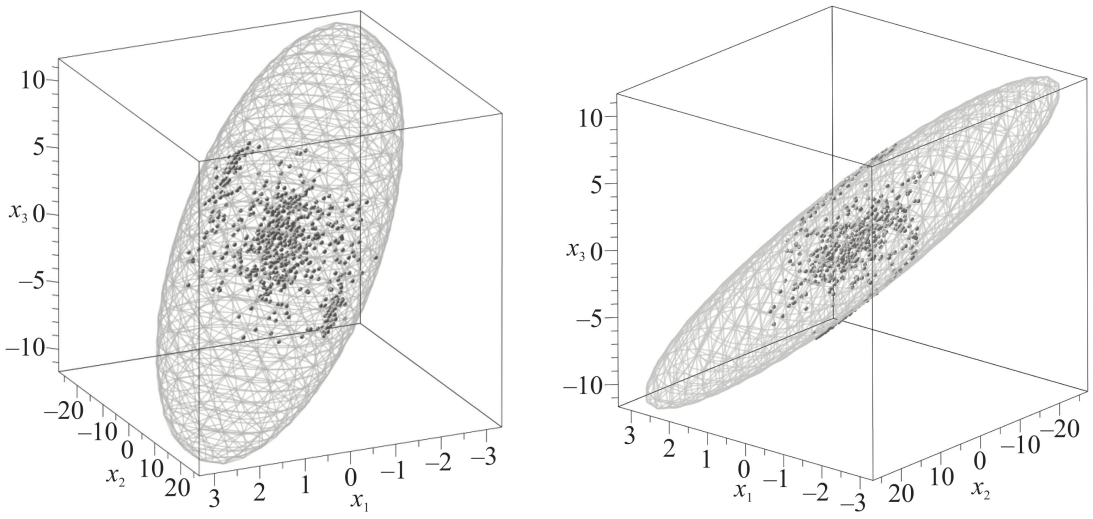


Fig. 4. The set $\mathcal{Y}_{t,2,\infty}$ and various terminal states for the system (A, B, H, t) from Example 7.

In Examples 4–7, only the construction of limit reachable sets (2) is considered, since the numerical modeling of its elements is simpler compared to the null-controllable sets (3). At the same time, according to Theorems 1–3, the construction of the set $\mathcal{X}_{t,2,\infty}$ is similar. It should also be noted that the case $m = 1$ is considered in detail in [14] and is therefore not presented here.

6. CONCLUSION

This paper solves the problem of constructing limit reachable and null-controllable sets for discrete-time linear systems with second-order integral constraints. Necessary and sufficient conditions for the boundedness of these sets are formulated in the form of Theorem 1, which reduces to the relationship between the matrix B and the eigensubspaces $\mathbb{L}_{<1}$ and $\mathbb{L}_{>1}$. For controllable systems, these conditions are equivalent to a constraint on the spectral radius of the matrix A for reachable sets and the matrix A^{-1} for null-controllable sets, respectively. It is also proved that, in the case of boundedness, the sets $\mathcal{Y}_{t,2,\infty}$ and $\mathcal{X}_{t,2,\infty}$ are ellipsoids. Their support function is constructed explicitly by the Jordan decomposition of the matrix A, which is formulated as Theorem 3. To confirm the obtained theoretical results, numerical experiments were conducted for systems of different dimensions, with different types of eigenvalues of the matrix A.

FUNDING

The work was carried out with the financial support of the Russian Science Foundation (project no. 25-21-00487).

APPENDIX

Let us introduce into $\mathbb{R}^n$ the dot product and the vector norm by the formulas

$$\langle x, y\rangle \triangleq x^{\mathrm{T}} y = \sum_{i=1}^{n} x_i y_i,\ \ \|x\| \triangleq \sqrt{\langle x, y\rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Proof of Theorem 1. Without loss of generality, we assume that the elements of basis (4) are grouped in such a way that for the class $M_0 \leqslant M$ the following representation is valid:

$$\mathbb{L}_{<1} = \operatorname{Lin}\left\{h_{i,j} \colon i = \overline{1, M_0},\ j = \overline{1, n_i}\right\}.$$

For each $u(k) \in \mathbb{R}^m$ we define the expansion in basis (4):

$$Bu(k) = \sum_{i=1}^{M} \sum_{j=1}^{n_i} h_{i,j} \alpha_{i,j}(k),$$

where $\alpha_{i,j}(k) \in \mathbb{R}$ if $\lambda_i \in \mathbb{R}$, and $\alpha_{i,j}(k) \in \mathbb{R}^2$ if $\lambda_i \in \mathbb{C}$. Taking into account representation (5), the following equalities hold:

$$S^{-1}Bu(k) = \operatorname{vec}\left(\alpha_{1,1}(k), \ldots, \alpha_{1,n_1}(k), \ldots, \alpha_{M,1}(k), \ldots, \alpha_{M,n_M}(k)\right),$$

$$\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_M \end{pmatrix} \triangleq \sum_{k=0}^{N-1} \Lambda^k S^{-1} Bu(N-k-1)$$

$$= \begin{pmatrix} \sum\limits_{k=0}^{N-1} J_1^k \operatorname{vec}\left(\alpha_{1,1}(N-k-1), \ldots, \alpha_{1,n_1}(N-k-1)\right) \\ \vdots \\ \sum\limits_{k=0}^{N-1} J_M^k \operatorname{vec}\left(\alpha_{M,1}(N-k-1), \ldots, \alpha_{M,n_M}(N-k-1)\right) \end{pmatrix}, \ N \in \mathbb{N}.$$

Let $b_1, \ldots, b_m \in \mathbb{L}_{<1}$. Then for any $k \in \mathbb{N} \cup \{0\}$, $i = \overline{M_0+1, M}$ and all $j = \overline{1, n_i}$, $\alpha_{i,j}(k) = 0$, and therefore $\beta_i = 0$.

Let $i = \overline{1, M_0}$. Let representation (6) holds for the Jordan cell J_i. Let us introduce the matrix

$$P = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{n_i \times n_i}.$$

Note the following property of the matrix P:

$$(P^l x)_j = \begin{cases} x_{j+l}, & j+l \leqslant n_i, \\ 0, & j+l > n_i, \end{cases} = \begin{cases} x_{j+l}, & l \leqslant n_i - j, \\ 0, & l > n_i - j. \end{cases}$$

Let C_k^l denotes the binomial coefficients. According to Newton's binomial theorem

$$J_i^k = (\lambda_i I + P)^k = \sum_{l=0}^{k} C_k^l \lambda_i^{k-l} P^l,$$

$$\begin{pmatrix} \beta_{i,1} \\ \vdots \\ \beta_{i,n_i} \end{pmatrix} \triangleq \beta_i = \sum_{k=0}^{N-1} J_i^k \begin{pmatrix} \alpha_{i,1}(N-k-1) \\ \vdots \\ \alpha_{i,n_i}(N-k-1) \end{pmatrix} = \sum_{k=0}^{N-1} \sum_{l=0}^{k} C_k^l \lambda_i^{k-l} P^l \begin{pmatrix} \alpha_{i,1}(N-k-1) \\ \vdots \\ \alpha_{i,n_i}(N-k-1) \end{pmatrix}.$$

Taking into account the properties of the matrix P, for all $j = \overline{1, n_i}$ the following representation is valid:

$$\beta_{i,j} = \sum_{k=0}^{N-1} \sum_{l=0}^{\min\{k, n_i - j\}} C_k^l \lambda_i^{k-l} \alpha_{i,j+l}(N-k-1). \tag{A.1}$$

Assuming $N - 1 \geqslant n_i - j$, we change the order of summation:

$$\beta_{i,j} = \sum_{l=0}^{n_i-j} \sum_{k=l}^{N-1} C_k^l \lambda_i^{k-l} \alpha_{i,j+l}(N-k-1) = \sum_{l=0}^{n_i-j} \sum_{k=0}^{N-l-1} C_{k+l}^l \lambda_i^k \alpha_{i,j+l}(N-k-l-1). \tag{A.2}$$

We take into account that $u(k)$, by (1), belongs to the ellipsoidal set $\mathcal{U} = \{u \in \mathbb{R}^m \colon u^{\mathrm{T}} H u \leqslant t^2\}$, which is compact. Then, by the Weierstrass theorem [22, Chapter VII, §2, Section 2] and the continuous dependence of the numbers $\alpha_{i,j}(k)$ on $u(k)$, there exists a number $\alpha_{\max} > 0$ such that for all $k \in \mathbb{N} \cup \{0\}$, $i = \overline{1, M}$, $j = \overline{1, n_i}$, the constraint $|\alpha_{i,j}(k)| \leqslant \alpha_{\max}$ holds. The following estimate is valid:

$$|\beta_{i,j}| \leqslant \sum_{l=0}^{n_i - j} \sum_{k=0}^{N-l-1} C_{k+l}^l |\lambda_i|^k |\alpha_{i,j+l}(N-k-l-1)| \leqslant \alpha_{\max} \sum_{l=0}^{n_i-j} \sum_{k=0}^{\infty} C_{k+l}^l |\lambda_i|^k$$

$$= \sum_{l=0}^{n_i-j} \frac{\alpha_{\max}}{(1-|\lambda_i|)^{l+1}} = \frac{(n_i - j + 1)\alpha_{\max}}{(1-|\lambda_i|)^{n_i - j + 1}} \leqslant \frac{n_i \alpha_{\max}}{(1-|\lambda_i|)^{n_i}}.$$

We will finally get

$$\|\beta_i\|^2 = \sum_{j=1}^{n_i} |\beta_{i,j}|^2 \leqslant \frac{n_i^3 \alpha_{\max}^2}{(1-|\lambda_i|)^{2n_i}}. \tag{A.3}$$

Let representation (7) be valid for the Jordan cell J_i. Setting $P \in \mathbb{R}^{2n_i \times 2n_i}$, we obtain from Newton's binomial theorem

$$J_i^k = \left(\operatorname{diag}(r_i A_{\varphi_i}, \ldots, r_i A_{\varphi_i}) + P^2\right)^k = \sum_{l=0}^{k} C_k^l \operatorname{diag}\left(r_i^{k-l} A_{(k-l)\varphi_i}, \ldots, r_i^{k-l} A_{(k-l)\varphi_i}\right) P^{2l},$$

$$\begin{pmatrix} \beta_{i,1} \\ \vdots \\ \beta_{i,n_i} \end{pmatrix} \triangleq \beta_i = \sum_{k=0}^{N-1} J_i^k \begin{pmatrix} \alpha_{i,1}(N-k-1) \\ \vdots \\ \alpha_{i,n_i}(N-k-1) \end{pmatrix}$$

$$= \sum_{k=0}^{N-1} \sum_{l=0}^{k} \operatorname{diag}\left(r_i^{k-l} A_{(k-l)\varphi_i}, \ldots, r_i^{k-l} A_{(k-l)\varphi_i}\right) P^{2l} \begin{pmatrix} \alpha_{i,1}(N-k-1) \\ \vdots \\ \alpha_{i,n_i}(N-k-1) \end{pmatrix}.$$

Taking into account the properties of the matrix P, for all $j = \overline{1, n_i}$ the following representation is valid:

$$\beta_{i,j} = \sum_{k=0}^{N-1} \sum_{l=0}^{\min\{k, n_i - j\}} C_k^l r_i^{k-l} A_{(k-l)\varphi_i} \alpha_{i,j+l}(N-k-1). \tag{A.4}$$

Assuming $N - 1 \geqslant n_i - j$, we change the order of summation:

$$\beta_{i,j} = \sum_{l=0}^{n_i-j} \sum_{k=l}^{N-1} C_k^l r_i^{k-l} A_{(k-l)\varphi_i} \alpha_{i,j+l}(N-k-1) \tag{A.5}$$

$$= \sum_{l=0}^{n_i-j} \sum_{k=0}^{N-l-1} C_{k+l}^l r_i^{k-l} A_{(k-l)\varphi_i} \alpha_{i,j+l}(N-k-l-1).$$

We take into account that $u(k)$, by (1), belongs to the ellipsoidal set $\mathcal{U} = \{u \in \mathbb{R}^m \colon u^{\mathrm{T}} H u \leqslant t^2\}$, which is compact. Then, by Weierstrass's theorem [22, Chapter VII, §2, Section 2] and the continuous dependence of the two-dimensional vectors $\alpha_{i,j}(k)$ on $u(k)$, there exists a number $\alpha_{\max} > 0$

such that for all $k \in \mathbb{N} \cup \{0\}$, $i = \overline{1, M}$, $j = \overline{1, n_i}$, the constraint $\|\alpha_{i,j}(k)\| \leqslant \alpha_{\max}$ holds. The following estimate is valid:

$$\|\beta_{i,j}\| \leqslant \sum_{l=0}^{n_i-j} \sum_{k=0}^{N-l-1} C_{k+l}^{l} r_i^k \|A_{k\varphi_i} \alpha_{i,j+l}(N-k-l-1)\|$$

$$= \sum_{l=0}^{n_i-j} \sum_{k=0}^{N-l-1} C_{k+l}^{l} r_i^k \|\alpha_{i,j+l}(N-k-l-1)\| \leqslant \alpha_{\max} \sum_{l=0}^{n_i-j} \sum_{k=0}^{\infty} C_{k+l}^{l} r_i^k$$

$$= \sum_{l=0}^{n_i-j} \frac{\alpha_{\max}}{(1-r_i)^{l+1}} = \frac{(n_i-j+1)\alpha_{\max}}{(1-r_i)^{n_i-j+1}} \leqslant \frac{n_i \alpha_{\max}}{(1-r_i)^{n_i}}.$$

We will finally get

$$\|\beta_i\|^2 = \sum_{j=1}^{n_i} \|\beta_{i,j}\|^2 \leqslant \frac{n_i^3 \alpha_{\max}^2}{(1-r_i)^{2n_i}}. \tag{A.6}$$

We take into account the inequalities (A.3) and (A.6) and obtain for all $N \geqslant n$ and the condition $x(0) = 0$ the estimate

$$\|x(N)\| \overset{(1)}{=} \left\| S \sum_{k=0}^{N-1} \Lambda^k S^{-1} B u(k) \right\| \leqslant \|S\| \left(\sum_{i=1}^{M_0} \|\beta_i\|^2 \right)^{\frac{1}{2}} \leqslant \frac{\|S\| \left(M_0 n^3 \alpha_{\max}^2\right)^{\frac{1}{2}}}{\left(1 - \max\limits_{i=\overline{1,M_0}} |\lambda_i|\right)^n},$$

i.e., by definition $\mathcal{Y}_{t,2,\infty}$ is bounded.

Now suppose that for some $\tilde{m} = \overline{1, m}$ the condition $b_{\tilde{m}} \notin \mathbb{L}_{<1}$ is satisfied. I.e., there exists $u \in \mathbb{R}^m$ such that in the expansion over basis (4)

$$Bu = \sum_{i=1}^{M} \sum_{j=1}^{n_i} h_{i,j} \alpha_{i,j}$$

For some $i = \overline{M_0 + 1, M}$ and $j = \overline{1, n_i}$, it is true that $\alpha_{i,j} \neq 0$. Fix these i, j, and $N \geqslant n$. Without loss of generality, we assume that either $j = n_i$ or $\alpha_{i,j+1} = \ldots = \alpha_{i,n_i} = 0$. Let J_i satisfy representation (6). Choose

$$u(N-k-1) = \frac{\sqrt{6} t \operatorname{sgn}(\lambda_i^k)}{(k+1)\pi} \frac{u}{\sqrt{u^{\mathrm{T}} H u}}, \ k = \overline{0, N-1}.$$

Then, on the one hand,

$$\sum_{k=0}^{N-1} u(k)^{\mathrm{T}} H u(k) = \sum_{k=0}^{N-1} \left(\frac{\sqrt{6} t \operatorname{sgn}(\lambda_i^k)}{(k+1)\pi} \right)^2 \frac{u^{\mathrm{T}} H u}{\sqrt{u^{\mathrm{T}} H u} \sqrt{u^{\mathrm{T}} H u}} < \frac{6t^2}{\pi^2} \sum_{k=0}^{\infty} \frac{1}{(k+1)^2} = t^2,$$

i.e., the selected control actions are admissible for system (1). On the other hand, for this control, according to (A.2), $\beta_{i,j}$ satisfies the relation

$$\beta_{i,j} = \sum_{l=0}^{n_i-j} \sum_{k=0}^{N-l-1} C_{k+l}^{l} \lambda_i^k \alpha_{i,j+l}(N-k-l-1) = \sum_{k=0}^{N-1} \lambda_i^k \alpha_{i,j}(N-k-1)$$

$$= \sum_{k=0}^{N-1} |\lambda_i|^k \frac{\sqrt{6} t}{(k+1)\pi} \frac{\alpha_{i,j}}{\sqrt{u^{\mathrm{T}} H u}} \geqslant \frac{\sqrt{6} t \alpha_{i,j}}{\pi \sqrt{u^{\mathrm{T}} H u}} \sum_{k=0}^{N-1} \frac{1}{k+1} \xrightarrow{N \to \infty} +\infty.$$

Let J_i satisfy representation (7). Choose an increasing sequence $k_1, k_2, \ldots \in \mathbb{N} \cup \{0\}$ such that the condition $|\varphi_i k_p| \mod 2\pi < \frac{\pi}{6}$ holds for all $p \in \mathbb{N}$. Such a sequence exists. For the case $\varphi_i = 2\pi\frac{K'}{K''}$, it suffices to set $k_p = pK''$. For the opposite case, by Weyl's theorem, the set $\{k\varphi_i \mod 2\pi\colon k \in \mathbb{N} \cup \{0\}\}$ is dense in $[0; 2\pi)$. Set for all $k = \overline{0, k_N}$

$$u(k_N - k) = \begin{cases} 0, & k \notin \{k_1, \ldots, k_N\}, \\ \dfrac{\sqrt{6}t}{p\pi}\dfrac{u}{\sqrt{u^{\mathrm{T}}Hu}}, & k = k_p \in \{k_1, \ldots, k_N\}. \end{cases}$$

Then

$$\sum_{k=0}^{k_N} u(k)^{\mathrm{T}} H u(k) = \sum_{p=1}^{N} \left(\frac{\sqrt{6}t}{p\pi}\right)^2 \frac{u^{\mathrm{T}}Hu}{\sqrt{u^{\mathrm{T}}Hu}\sqrt{u^{\mathrm{T}}Hu}} < \frac{6t^2}{\pi^2}\sum_{p=1}^{\infty}\frac{1}{p^2} = t^2,$$

i.e., the selected control actions are admissible for system (1). On the other hand, for this control, according to (A.5), $\beta_{i,j}$ satisfies the relation

$$\begin{aligned}
\|\beta_{i,j}\|^2 &= \left\langle \sum_{l=0}^{n_i-j}\sum_{k=0}^{k_N-l} C_{k+l}^{l} r_i^k A_{k\varphi_i}\alpha_{i,j+l}(k_N-k-l), \sum_{l=0}^{n_i-j}\sum_{k=0}^{k_N-l} C_{k+l}^{l} r_i^k A_{k\varphi_i}\alpha_{i,j+l}(k_N-k-l)\right\rangle \\
&= \left\langle \sum_{k=0}^{k_N} r_i^k A_{k\varphi_i}\alpha_{i,j}(k_N-k), \sum_{k=0}^{k_N} r_i^k A_{k\varphi_i}\alpha_{i,j}(k_N-k)\right\rangle \\
&= \left\langle \sum_{p=1}^{N} r_i^{k_p} A_{k_p\varphi_i}\alpha_{i,j}(k_N-k_p), \sum_{q=1}^{N} r_i^{k_q} A_{k_q\varphi_i}\alpha_{i,j}(k_N-k_q)\right\rangle \\
&= \sum_{p=1}^{N}\sum_{q=1}^{N} r_i^{k_p+k_q}\left\langle A_{(k_p-k_q)\varphi_i}\alpha_{i,j}(k_N-k_p), \alpha_{i,j}(k_N-k_q)\right\rangle \\
&= \sum_{p=1}^{N}\sum_{q=1}^{N} r_i^{k_p+k_q}\frac{6t^2}{\pi^2 pq}\frac{\|\alpha_{i,j}\|^2}{u^{\mathrm{T}}Hu}\cos(\varphi_i(k_p-k_q)) \geqslant \sum_{p=1}^{N}\sum_{q=1}^{N}\frac{6t^2}{\pi^2 pq}\frac{\|\alpha_{i,j}\|^2}{2u^{\mathrm{T}}Hu} \\
&= \frac{\|\alpha_{i,j}\|^2}{u^{\mathrm{T}}Hu}\frac{3t^2}{\pi^2}\left(\sum_{p=1}^{N}\frac{1}{p}\right)\left(\sum_{q=1}^{N}\frac{1}{q}\right) \xrightarrow{N\to\infty} +\infty.
\end{aligned}$$

Thus, since

$$\left\|\sum_{k=0}^{k_N}\Lambda^k S^{-1}Bu(k_N-k)\right\| = \|S^{-1}x(k_N+1)\| \leqslant \|S^{-1}\|\|x(k_N+1)\|,$$

then

$$\|x(k_N+1)\| \geqslant \frac{1}{\|S^{-1}\|}\left\|\sum_{k=0}^{k_N}\Lambda^k S^{-1}Bu(k_N-k)\right\| = \frac{\left(\sum\limits_{l=1}^{M}\|\beta_l\|^2\right)^{\frac{1}{2}}}{\|S^{-1}\|} \geqslant \frac{\|\beta_{i,j}\|}{\|S^{-1}\|} \xrightarrow{N\to\infty} +\infty.$$

Therefore, $\mathcal{Y}_{t,2,\infty}$ is unbounded by definition.

Now we prove item 2. If $\det A = 0$, then there exists an element of the real Jordan basis $h_{i,j} \in \mathbb{R}^n \setminus \{0\}$ corresponding to the eigenvalue $\lambda_i = 0$. Then, by (1), the inclusion $\mathrm{Lin}\{h_{i,j}\} \subset \mathcal{X}_{t,2,\infty}$ holds, which means that set (3) is unbounded. If $\det A \neq 0$, then by the definitions of (2) and (3),

the set $\mathcal{Y}_{t,2,\infty}$ for the system (A, B, H, t) and the set $\mathcal{X}_{t,2,\infty}$ for the system $(A^{-1}, A^{-1}B, H, t)$ coincide. This implies item 2.

Theorem 1 is proved.

Proof of Lemma 1. Without loss of generality, we will assume that the elements of basis (4) are grouped in such a way that for some $M_0 \leqslant M$ the following representation is true:

$$\mathbb{L}_{<1} = \operatorname{Lin}\left\{h_{i,j}\colon i = \overline{1, M_0},\ j = \overline{1, n_i}\right\}.$$

Since H^{-1} is positive definite and symmetric, it satisfies the Cholesky decomposition [16, Corollary 7.2.9], i.e., there exists a lower triangular non-singular matrix $L \in \mathbb{R}^{m\times m}$ such that $H^{-1} = LL^{\mathrm{T}}$. In terms of (5), we introduce the notation

$$\begin{pmatrix} \gamma_{1,1}(k) & \dots & \gamma_{1,m}(k) \\ \vdots & \ddots & \vdots \\ \gamma_{M,1}(k) & \dots & \gamma_{M,m}(k) \end{pmatrix} \stackrel{\triangle}{=} \Lambda^k S^{-1} BL,$$

where $\gamma_{i,1}(k), \dots, \gamma_{i,m}(k)$ are vectors whose dimension coincides with the dimension of the Jordan block J_i. Then, from the inclusion $b_1, \dots, b_m \in \mathbb{L}_{<1}$, it follows that $\gamma_{i,l}(k) = 0$ if $i = \overline{M_0 + 1, M}$, $l = \overline{1, m}$. In the case $i = \overline{1, M_0}$, the vectors $\gamma_{i,l}(k)$ are the product of the powers of the Jordan blocks J_i^k and the corresponding blocks of the matrix $S^{-1}BL$. Consequently, the order of smallness of the elements of the vectors $\gamma_{i,l}(k)$ coincides with the order of smallness of the elements J_i^k, which, according to the reasoning from [16, Section 3.2.5], can be estimated from above by the quantity

$$k^{n_i}|\lambda_i|^k \leqslant k^n \lambda_{\max}^k, \text{ where } \lambda_{\max} \stackrel{\triangle}{=} \max\{|\lambda|\colon \lambda \text{ is eigenvalue of } A,\ |\lambda| < 1\}.$$

From here, by the sufficient condition for convergence [22, Chapter III, §1, Section 4, Corollary 2], it follows that the matrix series $\sum\limits_{k=0}^{\infty} \gamma_{i,l}(k)\gamma_{j,l}(k)^{\mathrm{T}}$ converges elementwise for all $i, j = \overline{1, M}$.

Consider the chain of equalities

$$H_{\mathcal{Y}} = \sum_{k=0}^{\infty} A^k B H^{-1} B^{\mathrm{T}} (A^k)^{\mathrm{T}} = S \left(\sum_{k=0}^{\infty} \Lambda^k S^{-1} B H^{-1} B^{\mathrm{T}} (S^{-1})^{\mathrm{T}} (\Lambda^k)^{\mathrm{T}} \right) S^{\mathrm{T}}$$

$$= S \sum_{k=0}^{\infty} \sum_{l=1}^{m} \begin{pmatrix} \gamma_{1,l}(k)\gamma_{1,l}(k)^{\mathrm{T}} & \dots & \gamma_{1,l}(k)\gamma_{M,l}(k)^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \gamma_{M,l}(k)\gamma_{1,l}(k)^{\mathrm{T}} & \dots & \gamma_{M,l}(k)\gamma_{M,l}(k)^{\mathrm{T}} \end{pmatrix} S^{-1}$$

$$= \sum_{l=1}^{m} S \begin{pmatrix} \sum\limits_{k=0}^{\infty} \gamma_{1,l}(k)\gamma_{1,l}(k)^{\mathrm{T}} & \dots & \sum\limits_{k=0}^{\infty} \gamma_{1,l}(k)\gamma_{M,l}(k)^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \sum\limits_{k=0}^{\infty} \gamma_{M,l}(k)\gamma_{1,l}(k)^{\mathrm{T}} & \dots & \sum\limits_{k=0}^{\infty} \gamma_{M,l}(k)\gamma_{M,l}(k)^{\mathrm{T}} \end{pmatrix} S^{-1}.$$

From the convergence of the matrix series in the extreme right-hand side of these relations follows clause 1 of the Lemma 1.

Item 2 of the lemma follows from item 1 due to the fact that the eigenvalues of A are mutually inverse to the eigenvalues of A^{-1}.

Lemma 1 is proved.

Proof of Theorem 2. Let us prove item 1. For arbitrary $N \in \mathbb{N}$, $u(0), \dots, u(N-1) \in \mathbb{R}^m$, $p \in \mathbb{R}^n \setminus \{0\}$, we introduce the notations

$$p_0 = \begin{pmatrix} \left(A^{N-1}B\right)^{\mathrm{T}} p \\ \vdots \\ B^{\mathrm{T}}p \end{pmatrix} \in \mathbb{R}^{mN},\ u_0 = \begin{pmatrix} u(0) \\ \vdots \\ u(N-1) \end{pmatrix} \in \mathbb{R}^{mN},\ H_0 = \begin{pmatrix} H & \dots & O \\ \vdots & \ddots & \vdots \\ O & \dots & H \end{pmatrix} \in \mathbb{R}^{mN\times mN}.$$

Then by definition (2) we get

$$\mathcal{S}(p, \mathcal{Y}_{t,2,\infty}) = \sup_{x\in\mathcal{Y}_{t,2,\infty}} \langle p, x\rangle = \sup_{N\in\mathbb{N}} \max_{\sum\limits_{k=0}^{N-1} u(k)^{\mathrm{T}}Hu(k)\leqslant t^2} \left\langle p, \sum_{k=0}^{N-1} A^k B u(N-k-1)\right\rangle$$

$$= \sup_{N\in\mathbb{N}} \max_{\sum\limits_{k=0}^{N-1} u(k)^{\mathrm{T}}Hu(k)\leqslant t^2} \sum_{k=0}^{N-1} \left\langle \left(A^kB\right)^{\mathrm{T}} p, u(N-k-1)\right\rangle = \sup_{N\in\mathbb{N}} \max_{u_0^{\mathrm{T}}H_0u_0\leqslant t^2} \langle p_0, u_0\rangle .$$

Taking into account the invertibility of the matrix H_0 and definitions (8), (9), it is true that

$$\mathcal{S}(p, \mathcal{Y}_{t,2,\infty}) = \sup_{N\in\mathbb{N}} t\sqrt{p_0^{\mathrm{T}}H_0^{-1}p_0} = t \sup_{N\in\mathbb{N}} \sqrt{p^{\mathrm{T}} \sum_{k=0}^{N-1} A^kBH^{-1}B^{\mathrm{T}}(A^k)^{\mathrm{T}}p}$$

$$= t\sqrt{p^{\mathrm{T}} \sum_{k=0}^{\infty} A^kBH^{-1}B^{\mathrm{T}}(A^k)^{\mathrm{T}}p} = t\sqrt{p^{\mathrm{T}}H_{\mathcal{Y}}p}.$$

The last transition is correct, since by Lemma 1 the corresponding series converges to $H_{\mathcal{Y}}$.

Item 2 of the theorem follows from the fact that if $\det A \neq 0$, then due to definitions (2) and (3) the set $\mathcal{Y}_{t,2,\infty}$ for the system (A, B, H, t) and the set $\mathcal{X}_{t,2,\infty}$ for the system $(A^{-1}, A^{-1}B, H, t)$ coincide.

Theorem 2 is proved.

Proof of Lemma 2. Let us prove item 1. The series $\sum\limits_{k=0}^{\infty} A^kBH^{-1}B^{\mathrm{T}}(A^k)^{\mathrm{T}}$, due to Lemma 1, converges to a non-negative definite symmetric matrix $H_{\mathcal{Y}}$. The following representation holds:

$$\sum_{k=0}^{n-1} A^kBH^{-1}B^{\mathrm{T}}(A^k)^{\mathrm{T}} = (B\ \dots\ A^{n-1}B) \begin{pmatrix} H^{-1} & \dots & O \\ \vdots & \ddots & \vdots \\ O & \dots & H^{-1} \end{pmatrix} \begin{pmatrix} B^{\mathrm{T}} \\ \vdots \\ (A^{n-1}B)^{\mathrm{T}} \end{pmatrix}.$$

Moreover, according to the controllability condition, each of the factors is a matrix of full rank. Consequently, their product is also a matrix of full rank [16, Section 0.4.6], which means that $\det \sum\limits_{k=0}^{n-1} A^kBH^{-1}B^{\mathrm{T}}(A^k)^{\mathrm{T}} \neq 0$. Since $\sum\limits_{k=0}^{n-1} A^kBH^{-1}B^{\mathrm{T}}(A^k)^{\mathrm{T}}$ is also non-negative definite and symmetric by construction, this matrix turns out to be positive definite. Then $H_{\mathcal{Y}}$ is also positive definite, since it can be represented as the sum of positive definite and non-negative definite matrices:

$$H_{\mathcal{Y}} = \sum_{k=n}^{\infty} A^kBH^{-1}B^{\mathrm{T}}(A^k)^{\mathrm{T}} + \sum_{k=0}^{n-1} A^kBH^{-1}B^{\mathrm{T}}(A^k)^{\mathrm{T}}.$$

Item 2 of the lemma follows from the fact that the eigenvalues of A are mutually inverse to the eigenvalues of A^{-1}.

Lemma 2 is proved.

Lemma 3. *Let $r, q \in \mathbb{N} \cup \{0\}$, and let $\lambda, \mu \in \mathbb{C}$ be numbers such that $|\lambda| < 1$, $|\mu| < 1$. Then*

$$\sum_{k=\max\{r,q\}}^{\infty} C_k^r C_k^q \lambda^{k-r} \mu^{k-q} = \sum_{l=0}^{\min\{r,q\}} \frac{(r+q-l)!}{(r-l)!(q-l)!l!} \frac{\lambda^{q-l}\mu^{r-l}}{(1-\lambda\mu)^{r+q-l+1}}.$$

Proof of Lemma 3. To prove the lemma, we use the well-known formula for $|\alpha| < 1$:

$$\sum_{k=0}^{\infty} C_{k+r}^r \alpha^k = \frac{1}{(1-\alpha)^{r+1}}. \tag{A.7}$$

We choose neighborhoods of zero in the complex plane $\mathcal{O}_{\delta_1}(0), \mathcal{O}_{\delta_2}(0) \subset \mathbb{C}$ so that for any $x \in \mathcal{O}_{\delta_1}(0)$ and $y \in \mathcal{O}_{\delta_2}(0)$ the following inequalities hold:

$$|(x+\lambda)(y+\mu)| < 1,$$
$$\left|\frac{x(y+\mu)}{1-\lambda(y+\mu)}\right| < 1,$$
$$\left|\frac{y\lambda}{1-\lambda\mu}\right| < 1.$$

Such neighborhoods exist because $|\lambda| < 1$, $|\mu| < 1$. Consider the function $G\colon \mathcal{O}_{\delta_1}(0) \times \mathcal{O}_{\delta_2}(0) \to \mathbb{C}$ of the following form:

$$G(x,y) = \frac{1}{1-(x+\lambda)(y+\mu)}.$$

On the one side,

$$G(x,y) \overset{\text{(A.7)}}{=} \sum_{k=0}^{\infty} (x+\lambda)^k (y+\mu)^k = \sum_{k=0}^{\infty} \sum_{r=0}^{k} \sum_{q=0}^{k} C_k^q C_k^r \lambda^{k-r} \mu^{k-q} x^r y^q$$
$$= \sum_{r=0}^{\infty} \sum_{q=0}^{\infty} \left(\sum_{k=\max\{r,q\}}^{\infty} C_k^q C_k^r \lambda^{k-r} \mu^{k-q} \right) x^r y^q. \tag{A.8}$$

On the other hand, from the definitions of δ_1 and δ_2 we obtain

$$G(x,y) = \frac{1}{1-\lambda(y+\mu)-x(y+\mu)} = \frac{1}{1-\lambda(y+\mu)} \frac{1}{1-\frac{x(y+\mu)}{1-\lambda(y+\mu)}}$$
$$\overset{\text{(A.7)}}{=} \frac{1}{1-\lambda(y+\mu)} \sum_{r=0}^{\infty} \left(\frac{x(y+\mu)}{1-\lambda(y+\mu)} \right)^r = \sum_{r=0}^{\infty} x^r (y+\mu)^r \frac{1}{(1-\lambda(y+\mu))^{r+1}}$$
$$= \sum_{r=0}^{\infty} \sum_{l=0}^{r} x^r C_r^l y^l \mu^{r-l} \frac{1}{(1-\lambda\mu)^{r+1}} \frac{1}{\left(1-\frac{y\lambda}{1-\lambda\mu}\right)^{r+1}}$$
$$\overset{\text{(A.7)}}{=} \sum_{r=0}^{\infty} \sum_{l=0}^{r} \sum_{q=0}^{\infty} x^r C_r^l y^l \mu^{r-l} \frac{1}{(1-\lambda\mu)^{r+1}} C_{r+q}^r \left(\frac{y\lambda}{1-\mu\lambda} \right)^q$$
$$= \sum_{r=0}^{\infty} \sum_{l=0}^{r} \sum_{q=l}^{\infty} x^r C_r^l y^l \mu^{r-l} \frac{1}{(1-\lambda\mu)^{r+1}} C_{r+q-l}^r \left(\frac{y\lambda}{1-\mu\lambda} \right)^{q-l}$$
$$= \sum_{r=0}^{\infty} \sum_{l=0}^{r} \sum_{q=l}^{\infty} \frac{C_r^l C_{r+q-l}^r \mu^{r-l} \lambda^{q-l}}{(1-\lambda\mu)^{q+r-l+1}} x^r y^q = \sum_{r=0}^{\infty} \sum_{q=0}^{\infty} \left(\sum_{l=0}^{\min\{r,q\}} \frac{C_r^l C_{r+q-l}^r \mu^{r-l} \lambda^{q-l}}{(1-\lambda\mu)^{q+r-l+1}} \right) x^r y^q. \tag{A.9}$$

Since $G(x,y)$ is analytic on its domain by construction, it can be represented as a power series in a unique way [23, Section A, Chapter I]. Therefore, from the representations (A.8) and (A.9) it follows that Lemma 3 is true.

Lemma 4. *Let $r,q \in \mathbb{N}\cup\{0\}$, and let $\lambda,\mu\in\mathbb{C}$ be numbers such that $|\lambda|>1$, $|\mu|>1$. Then*

$$\sum_{k=1}^{\infty} C_{k+r-1}^{r} C_{k+q-1}^{q} \lambda^{-k-r}\mu^{-k-q} = \sum_{l=0}^{\min\{r,q\}} \frac{(r+q-l)!}{(r-l)!(q-l)!l!}\frac{(-1)^l\lambda^{q-l}\mu^{r-l}}{(\lambda\mu-1)^{r+q-l+1}}.$$

Proof of Lemma 4. We choose neighborhoods of zero in the complex plane $\mathcal{O}_{\delta_1}(0),\mathcal{O}_{\delta_2}(0)\subset\mathbb{C}$ so that for any $x\in\mathcal{O}_{\delta_1}(0)$ and $y\in\mathcal{O}_{\delta_2}(0)$ the following inequalities hold:

$$|(\lambda-x)|>1,\ |(\mu-y)|>1,$$

$$\left|\frac{x(\mu-y)}{\lambda(\mu-y)-1}\right|<1,$$

$$\left|\frac{y\lambda}{\lambda\mu-1}\right|<1.$$

Such neighborhoods exist because $|\lambda|>1$, $|\mu|>1$. Consider the function $F\colon \mathcal{O}_{\delta_1}(0)\times\mathcal{O}_{\delta_2}(0)\to\mathbb{C}$ of the following form:

$$F(x,y)=\frac{1}{(\lambda-x)(\mu-y)-1}.$$

On the one side,

$$F(x,y)=\frac{1}{(\lambda-x)(\mu-y)}\frac{1}{1-(\lambda-x)^{-1}(\mu-y)^{-1}}$$

$$\overset{(A.7)}{=}\frac{1}{(\lambda-x)(\mu-y)}\sum_{k=0}^{\infty}(\lambda-x)^{-k}(\mu-y)^{-k}=\sum_{k=0}^{\infty}\frac{1}{(\lambda-x)^{k+1}}\frac{1}{(\mu-y)^{k+1}}$$

$$=\sum_{k=0}^{\infty}\frac{1}{(\lambda\mu)^{k+1}}\frac{1}{\left(1-\frac{x}{\lambda}\right)^{k+1}}\frac{1}{\left(1-\frac{y}{\mu}\right)^{k+1}}$$

$$\overset{(A.7)}{=}\sum_{k=0}^{\infty}\frac{1}{(\lambda\mu)^{k+1}}\sum_{r=0}^{\infty}C_{k+r}^{r}\left(\frac{x}{\lambda}\right)^r\sum_{q=0}^{\infty}C_{k+q}^{q}\left(\frac{y}{\mu}\right)^q$$

$$=\sum_{r=0}^{\infty}\sum_{q=0}^{\infty}\sum_{k=0}^{\infty}C_{k+r}^{r}C_{k+q}^{q}\lambda^{-r-k-1}\mu^{-q-k-1}x^ry^q$$

$$=\sum_{r=0}^{\infty}\sum_{q=0}^{\infty}\left(\sum_{k=1}^{\infty}C_{k+r-1}^{r}C_{k+q-1}^{q}\lambda^{-r-k}\mu^{-q-k}\right)x^ry^q. \tag{A.10}$$

On the other hand,

$$F(x,y)=\frac{1}{\lambda(\mu-y)-1-x(\mu-y)}=\frac{1}{\lambda(\mu-y)-1}\frac{1}{1-\frac{x(\mu-y)}{\lambda(\mu-y)-1}}$$

$$\overset{(A.7)}{=}\frac{1}{\lambda(\mu-y)-1}\sum_{r=0}^{\infty}\left(\frac{x(\mu-y)}{\lambda(\mu-y)-1}\right)^r=\sum_{r=0}^{\infty}\frac{x^r(\mu-y)^r}{(\lambda\mu-1-\lambda y)^{r+1}}$$

$$=\sum_{r=0}^{\infty}\sum_{l=0}^{r}x^rC_r^l(-y)^l\mu^{r-l}\frac{1}{(\lambda\mu-1)^{r+1}}\frac{1}{\left(1-\frac{y\lambda}{\lambda\mu-1}\right)^{r+1}}$$

$$\overset{(A.7)}{=}\sum_{r=0}^{\infty}\sum_{l=0}^{r}\sum_{q=0}^{\infty}x^r C_r^l(-y)^l\mu^{r-l}\frac{1}{(\lambda\mu-1)^{r+1}}C_{r+q}^r\left(\frac{y\lambda}{\lambda\mu-1}\right)^q$$

$$=\sum_{r=0}^{\infty}\sum_{l=0}^{r}\sum_{q=l}^{\infty}x^r C_r^l(-y)^l\mu^{r-l}\frac{1}{(\lambda\mu-1)^{r+1}}C_{r+q-l}^r\left(\frac{y\lambda}{\lambda\mu-1}\right)^{q-l}$$

$$=\sum_{r=0}^{\infty}\sum_{l=0}^{r}\sum_{q=l}^{\infty}\frac{(-1)^l C_r^l C_{r+q-l}^r\mu^{r-l}\lambda^{q-l}}{(\lambda\mu-1)^{q+r-l+1}}x^r y^q$$

$$=\sum_{r=0}^{\infty}\sum_{q=0}^{\infty}\left(\sum_{l=0}^{\min\{r,q\}}\frac{(-1)^l C_r^l C_{r+q-l}^r\mu^{r-l}\lambda^{q-l}}{(1-\lambda\mu)^{q+r-l+1}}\right)x^r y^q. \tag{A.11}$$

Since $F(x,y)$ is analytic on its domain by construction, it can be represented as a power series in a unique way [23, Section A, Chapter I]. Therefore, from the representations (A.10) and (A.11) it follows that Lemma 4 is true.

Proof of Theorem 3. Let us prove item 1. Due to (10) and the definition of $H_{\mathcal{Y}}$

$$H_{\mathcal{Y}}=S\left(\sum_{k=0}^{\infty}\Lambda^k S^{-1}BH^{-1}B^{\mathrm{T}}(S^{-1})^{\mathrm{T}}(\Lambda^k)^{\mathrm{T}}\right)S^{\mathrm{T}}$$

$$=S\left(\sum_{k=0}^{\infty}\begin{pmatrix}J_1^k D_{11}(J_1^k)^{\mathrm{T}} & \dots & J_1^k D_{1M}(J_M^k)^{\mathrm{T}}\\ \vdots & \ddots & \vdots\\ (J_M^k)^{\mathrm{T}}D_{M1}(J_1^k)^{\mathrm{T}} & \dots & J_M^k D_{MM}(J_M^k)^{\mathrm{T}}\end{pmatrix}\right)S^{\mathrm{T}},$$

i.e., it is true that

$$H_{ij}^{\mathcal{Y}}=\sum_{k=0}^{\infty}J_i^k D_{ij}(J_j^k)^{\mathrm{T}}.$$

From the inclusion $b_1,\dots,b_m\in\mathbb{L}_{<1}$ and (10) it follows that $D_{ij}=O$ if $|\lambda_i|\geqslant 1$ or $|\lambda_j|\geqslant 1$, i.e., $H_{ij}^{\mathcal{Y}}=O$. Otherwise, we take into account that for $k\geqslant n_0$ the equality $P_{n_0}^k=O$ holds, and consider the series

$$\sum_{k=0}^{\infty}J_i^k D_{ij}(J_j^k)^{\mathrm{T}}=\sum_{k=0}^{\infty}\left(\lambda_i I+P_{n_i}\right)^k D_{ij}\left(\lambda_j I+P_{n_j}^{\mathrm{T}}\right)^k$$

$$=\sum_{k=0}^{\infty}\sum_{r=0}^{\min\{k,n_i-1\}}\sum_{q=0}^{\min\{k,n_j-1\}}C_k^r\lambda_i^{k-r}P_{n_i}^r D_{ij}C_k^q\lambda_j^{k-q}(P_{n_j}^q)^{\mathrm{T}}$$

$$=\sum_{r=0}^{n_i-1}\sum_{q=0}^{n_j-1}\sum_{k=\max\{r,q\}}^{\infty}C_k^r C_k^q\lambda_i^{k-r}\lambda_j^{k-q}P_{n_i}^r D_{ij}(P_{n_j}^q)^{\mathrm{T}}.$$

Then, taking into account Lemma 3, item 1 is proved.

Now we prove item 2. Due to (10) and the definition of $H_{\mathcal{X}}$

$$H_{\mathcal{X}}=S\left(\sum_{k=1}^{\infty}\Lambda^{-k}S^{-1}BH^{-1}B^{\mathrm{T}}(S^{-1})^{\mathrm{T}}(\Lambda^{-k})^{\mathrm{T}}\right)S^{\mathrm{T}}$$

$$=S\left(\sum_{k=1}^{\infty}\begin{pmatrix}J_1^{-k} D_{11}(J_1^{-k})^{\mathrm{T}} & \dots & J_1^{-k} D_{1M}(J_M^{-k})^{\mathrm{T}}\\ \vdots & \ddots & \vdots\\ (J_M^{-k})^{\mathrm{T}}D_{M1}(J_1^{-k})^{\mathrm{T}} & \dots & J_M^{-k} D_{MM}(J_M^{-k})^{\mathrm{T}}\end{pmatrix}\right)S^{\mathrm{T}},$$

i.e., it is true that

$$H_{ij}^{\mathcal{X}} = \sum_{k=1}^{\infty} J_i^{-k} D_{ij} (J_j^{-k})^{\mathrm{T}}.$$

According to [16, Section 3.2.5] the following representation is valid for the Jordan cell:

$$J_i^{-k} = \sum_{r=0}^{n_i-1} (-1)^r C_{k+r-1}^r \frac{1}{\lambda_i^{k+r}} P_{n_i}^r. \tag{A.12}$$

From the inclusion $b_1, \ldots, b_m \in \mathbb{L}_{>1}$ and (10) it follows that $D_{ij} = O$ if $|\lambda_i| \leqslant 1$ or $|\lambda_j| \leqslant 1$, i.e., $H_{ij}^{\mathcal{X}} = O$. Otherwise, we take into account that for $k \geqslant n_0$ the equality $P_{n_0}^k = O$ holds, and consider the series

$$\sum_{k=1}^{\infty} J_i^k D_{ij} (J_j^k)^{\mathrm{T}} = \sum_{k=1}^{\infty} (\lambda_i I + P_{n_i})^{-k} D_{ij} \left(\lambda_j I + P_{n_j}^{\mathrm{T}}\right)^{-k}$$

$$\overset{(A.12)}{=} \sum_{k=0}^{\infty} \sum_{r=0}^{n_i-1} \sum_{q=0}^{n_j-1} (-1)^{r+q} C_{k+r-1}^r C_{k+q-1}^q \frac{1}{\lambda_i^{k+r}} \frac{1}{\lambda_j^{k+q}} P_{n_i}^r D_{ij} (P_{n_j}^q)^{\mathrm{T}}$$

$$= \sum_{r=0}^{n_i-1} \sum_{q=0}^{n_j-1} \sum_{k=1}^{\infty} C_{k+r-1}^r C_{k+q-1}^q (-\lambda_i)^{-k-r} (-\lambda_j)^{-k-q} P_{n_i}^r D_{ij} (P_{n_j}^q)^{\mathrm{T}}.$$

Then, taking into account Lemma 4, item 2 is proved.

Theorem 3 is proved.

REFERENCES

1. Foss, E., Rosengren, A.J., and Biria, A.D., Reachability Analysis of Low-Thrust, Cislunar Spacecraft Using Set-Based Computing, *J. Astronaut. Sci.*, 2026, vol. 73, no. 2. https://doi.org/10.1007/s40295-025-00552-0

2. Liu, E.I. and Althoff, M., Specification-Compliant Motion Planning of Cooperative Vehicles Using Reachable Sets, *Cooperatively Interacting Vehicles*, 2024, pp. 321–345. https://doi.org/10.1007/978-3-031-60494-2_12

3. Zhang, Y., Zhang, H., and Xu, X., Backward Reachability Analysis of Neural Feedback Systems Using Hybrid Zonotopes, *IEEE Control Systems Letters*, 2023, vol. 7, pp. 2779–2784. https://doi.org/10.1109/LCSYS.2023.3289572

4. Zhang, Y., Zhang, H., and Xu, X., Reachability Analysis of Neural Network Control Systems With Tunable Accuracy and Efficiency, *IEEE Control Systems Letters*, 2024, vol. 8, pp. 1697–1702. https://doi.org/10.1109/LCSYS.2024.3415471

5. Serry, M., Yang, L., Ozay, N. and Liu, J., Safe Tracking Control of Discrete-Time Nonlinear Systems Using Backward Reachable Sets, *American Control Conference*, 2024, pp. 1172–1179. https://doi.org/10.23919/ACC60939.2024.10644670

6. Zykov, I.V., Approximate Calculation of Reachable Sets for Linear Control Systems with Different Control Constraints, *Izv. IMI UdGU*, 2022, vol. 60, pp. 16–33. https://doi.org/10.35634/2226-3594-2022-60-02

7. Natherson, R.A. and Scheeres, D.J., Reachable Set Computation with Terminal Velocity Constraints, *J. Guidance, Control. Dynam.*, 2024, vol. 47, no. 12, pp. 2560–2572. https://doi.org/10.2514/1.G008118

8. Guseinov, K.G., Ozer, O., Akyar, E., and Ushakov, V.N., The Approximation of Reachable Sets of Control Systems with Integral Constraint on Controls, *Nonlinear Differ. Equat. Appl.*, 2007, vol. 14, pp. 57–73. https://doi.org/10.1007/s00030-006-4036-6

9. Polyak, B.T., Convexity of the Reachable Set of Nonlinear Systems under L2-Bounded Controls, *Dynam. Contin. Discrete Impuls. Syst. Ser. A Math. Anal.*, 2004, vol. 11, pp. 255–267.

10. Simkina, A., On the External Estimation of the Limit Reachable Set for the Linear Discrete-Time System Based on Support Hyperplanes, *Advances in Systems Science and Applications*, 2024, vol. 24, no. 4, pp. 66–81. https://doi.org/10.25728/assa.2024.2024.4.1970

11. Berendakova, A.V. and Ibragimov, D.N., About the Method for Constructing External Estimates of the Limit 0-Controllability Set for the Linear Discrete-Time System with Bounded Control, *Autom. Remote Control*, 2023, vol. 84, no. 2, pp. 83–104. https://doi.org/10.1134/S0005117923020030

12. Ibragimov, D.N. and Sirotin, A.N., On Some Properties of Sets of Bounded Controllability for Stationary Linear Discrete Systems with Total Control Constraints, *J. Comput Syst. Sci. Int.*, 2023, vol. 62, no. 6, pp. 727–756. https://doi.org/10.1134/S1064230723050088

13. Ibragimov, D.N., On the Method for Constructing Null-Controllable Sets for Linear Discrete-Time Systems with Summary Control Constraints, *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2024, vol. 25, no. 10, pp. 503–512. https://doi.org/10.17587/mau.25.503-512

14. Ibragimov, D.N., On the External Estimation of Reachable and Null-Controllable Limit Sets for Linear Discrete-Time Systems with a Summary Constraint on the Scalar Control, *Autom. Remote Control*, 2024, vol. 85, no. 4, pp. 321–340. https://doi.org/10.1134/S0005117924040027

15. Ibragimov, D.N. and Samonov S.S., ON the Conditions of Limited Sets of Reachability and Controllability for Linear Systems with Discrete Time and Total First-Order Constraints on Scalar Control, *Modeling and Data Analysis*, 2025, vol. 15, no. 1, pp. 51–80. https://doi.org/10.17759/mda.2025150104

16. Horn, R. and Johnson, C., *Matrichnyi analiz* (Matrix Analysis), Moscow: Mir, 1989.

17. Rockafellar, R.T., *Convex Analysis*, Princeton: Princeton Univ. Press, 1970. Translated under the title *Vypuklyi analiz*, Moscow: Mir, 1973.

18. Polyak, B.T., Hlebnikov, M.V. and Rapoport, L.B., *Matematicheskaya teoriya avtomaticheskogo upravleniya* (Mathematical Theory of Automatic Control), Moscow: LENAND, 2019.

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

20. Yadykin, I.B. and Galyaev, A.A., On the Methods for Calculation of Grammians and Their Use in Analysis of Linear Dynamic Systems, *Autom. Remote Control*, 2013, vol. 74, no. 2, pp. 207–224. https://doi.org/10.1134/S0005117913020045

21. Zubov, N.E., Zybin, E.Yu., Mikrin, E.A., Misrikhanov, M.Sh., and Ryabchenko, V.N., General Analytical Forms for the Solution of the Sylvester and Lyapunov Equations for Continuous and Discrete Dynamic Systems, *J. Comput Syst. Sci. Int.*, 2017, vol. 56, no. 1, pp. 3–20. https://doi.org/10.1134/S1064230717010130

22. Zorich, V.A., *Mathematical Analysis I*, Berlin: Springer, 2015. https://doi.org/10.1007/978-3-662-48792-1

23. Gunning, R. and Rossi, H., *Analytic Functions of Several Complex Variables*, USA: Prentice-Hall, 1965. Translated under the title *Analiticheskie funktsii mnogikh kompleksnykh peremennykh*, Moscow: Mir, 1969.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 10, pp. 885–900.
© 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. 10, pp. 31–51.

LINEAR SYSTEMS

Optimization of Transients in a Three-Tempo Singularly Perturbed System

A. I. Kalinin[*,a] **and L. I. Lavrinovich**[*,b]

[*] *Belarusian State University, Minsk, Belarus*
e-mail: [a] *kalininai@bsu.by,* [b] *lavrinovich@bsu.by*

Received September 15, 2025
Revised April 20, 2026
Accepted April 28, 2026

Abstract—This paper considers the problem of minimum-energy control of a linear singularly perturbed system with three groups of variables changing at significantly different rates. Asymptotic approximations to the solution of this problem are constructed in the form of open-loop control and feedback control. The main advantage is that the proposed computational procedures decompose the original problem into three unperturbed optimal control problems of lower dimension.

Keywords: small parameter, singularly perturbed system, quadratic functional, optimal control, feedback, asymptotic approximations

DOI: 10.7868/S1608303226100022

1. INTRODUCTION

In the mathematical theory of optimal processes, considerable attention is given to optimization problems for singularly perturbed systems containing small parameters at some derivatives. This is due to the efficiency of asymptotic solution methods for such problems, which decompose the original problem into lower-dimensional optimal control ones. Moreover, the asymptotic approach avoids integrating singularly perturbed systems, which are stiff. Integration of such systems entails serious challenges manifested in high computational cost and the inevitable accumulation of computational errors. The first works on optimal control problems with singular perturbations addressed dynamic systems containing only two groups of variables with significantly different rates of change, the so-called slow and fast variables; see the reviews [1–4]. Subsequently, following the needs of applications, researchers considered optimization of systems with several groups of fast variables containing parameters of different orders of smallness at the derivatives. Such works were surveyed in [5]. This paper is devoted to constructing asymptotic approximations, in the form of open-loop control and feedback control, to the solution of the minimum-energy control problem for a singularly perturbed system with two groups of fast variables. The approach to constructing such approximations was first presented in [6]. Note that a similar problem for a two-tempo singularly perturbed system was studied in [7].

2. PROBLEM STATEMENT

In the class of r-dimensional control actions $u(t) = (u_1(t), \dots, u_r(t))$, $t \in T = [t_*, t^*]$, with piecewise continuous components, we consider a time-invariant optimal control problem of the form

$$\begin{aligned} \dot{x} &= A_1 x + B_1 y + C_1 z + D_1 u, \quad x(t_*) = x_*, \\ \mu \dot{y} &= A_2 x + B_2 y + C_2 z + D_2 u, \quad y(t_*) = y_*, \\ \mu^2 \dot{z} &= A_3 x + B_3 y + C_3 z + D_3 u, \quad z(t_*) = z_*, \end{aligned} \tag{1}$$

$$x(t^*) = 0, \quad y(t^*) = 0, \quad z(t^*) = 0, \tag{2}$$

$$J(u) = \frac{1}{2}\int\limits_{t_*}^{t^*} u^{\mathrm{T}} P u dt \to \min, \tag{3}$$

where μ is a small positive parameter; t_* and t^* are given time instants ($t_* < t^*$); x is the n_1-dimensional vector of slow variables; finally, y and z are the vectors of fast variables of dimensions n_2 and n_3, respectively. The other elements of the problem have compatible dimensions. In the performance criterion (3), P is a positive definite symmetric matrix.

Assumption 1. *The matrices C_3 and $B = B_2 - C_2 C_3^{-1} B_3$ are stable, i.e., the real parts of all their eigenvalues are negative.*

Let us introduce concepts to clarify the meaning of asymptotic approximations to the solution of this problem.

Definition 1. A control $u^{(N)}(t, \mu)$, $t \in T$, with piecewise continuous components is called an asymptotically suboptimal (open-loop) control of order N ($N = 0, 1, 2, \dots$) if it transfers the dynamic system (1) to a state of order $O(\mu^{N+1})$ and differs from the optimal control, in terms of the performance criterion (3), by a quantity of the same order of smallness.

Definition 2. A vector function $u^{(N)}(x, y, z, t, \mu)$ is called an asymptotically suboptimal feedback control of order N if, for any initial state (x_*, y_*, z_*, t_*), $t_* < t^*$, we have $u^{(N)}(x_*, y_*, z_*, t_*, \mu) = u^{(N)}(t_*, \mu)$, where $u^{(N)}(t, \mu)$, $t \in T$, is an asymptotically suboptimal control of order N in problem (1)–(3).

This paper is devoted to designing asymptotically suboptimal controls in problem (1)–(3). In addition, an asymptotically suboptimal feedback control of zero order is designed.

To shorten the notation, we will adopt the following designations:

$$A(\mu) = \begin{pmatrix} A_1 & B_1 & C_1 \\ A_2/\mu & B_2/\mu & C_2/\mu \\ A_3/\mu^2 & B_3/\mu^2 & C_3/\mu^2 \end{pmatrix}, \quad D(\mu) = \begin{pmatrix} D_1 \\ D_2/\mu \\ D_3/\mu^2 \end{pmatrix}.$$

3. FIRST BASIC PROBLEM

The essence of the asymptotically suboptimal control design approach proposed below lies in the asymptotic expansion of the initial values (at the time instant t^*) of the adjoint variables in integer powers of the small parameter. The leading coefficients of these expansions are determined by the initial values of the adjoint variables in three unperturbed optimal control problems of lower dimension. The first of these is the degenerate problem

$$\begin{gathered} \dot{x} = A_0 x + D_0 u, \ x(t_*) = x_*, \quad x(t^*) = 0, \\ J_1(u) = \frac{1}{2}\int\limits_{t_*}^{t^*} u^{\mathrm{T}} P u dt \to \min, \end{gathered} \tag{4}$$

where

$$\begin{gathered} A_0 = A_1 - C_1 C_3^{-1} A_3 - (B_1 - C_1 C_3^{-1} B_3) B^{-1} (A_2 - C_2 C_3^{-1} A_3), \\ D_0 = D_1 - C_1 C_3^{-1} D_3 - (B_1 - C_1 C_3^{-1} B_3) B^{-1} D_4, \quad D_4 = D_2 - C_2 C_3^{-1} D_3. \end{gathered} \tag{5}$$

It will be called the *first basic problem*.

Assumption 2. *The dynamic system in problem* (4) *is completely controllable* [8], *i.e.,*

$$\operatorname{rank}\left(D_0, A_0 D_0, A_0^2 D_0, \dots, A_0^{n_1 - 1} D_0\right) = n_1.$$

Under this assumption, admissible controls exist in the first basic problem, and hence it has a unique solution representing a normal extremal [9]. In other words, Pontryagin's maximum principle [10] in this case can be stated as follows: let $\overline{u}(t)$ and $x^0(t)$, $t \in T$, be an optimal control and an optimal trajectory, respectively, in problem (4); then there exists a solution $\psi^0(t)$, $t \in T$, of the adjoint system $\dot{\psi} = -A_0^{\mathrm{T}}\psi$ such that

$$\psi^{0\mathrm{T}}(t)D_0\overline{u}(t) - \frac{1}{2}\overline{u}^{\mathrm{T}}(t)P\overline{u}(t) = \max_{u\in R^r}\left(\psi^{0\mathrm{T}}(t)D_0u - \frac{1}{2}u^{\mathrm{T}}Pu\right), \quad t \in T.$$

This condition directly implies

$$\overline{u}(t) = P^{-1}D_0^{\mathrm{T}}\psi^0(t), \quad t \in T. \tag{6}$$

With $\lambda_0 = \psi^0(t^*)$, we have

$$\psi^{0\mathrm{T}}(t) = \lambda_0^{\mathrm{T}}F_0(t), \quad \psi^{0\mathrm{T}}(t)D_0 = \lambda_0^{\mathrm{T}}\Phi_0(t), \quad t \in T, \tag{7}$$

where

$$\Phi_0(t) = F_0(t)D_0, \quad t \in T, \tag{8}$$

and $F_0(t)$, $t \in T$, is a matrix function of dimensions $(n_1 \times n_1)$ representing the solution of the initial-value problem

$$\dot{F}_0 = -F_0A_0, \quad F_0(t^*) = E_{n_1}. \tag{9}$$

Due to (6), (7), and the Cauchy formula,

$$x^0(t^*) = F_0(t_*)x_* + S_1\lambda_0, \tag{10}$$

where

$$S_1 = \int_{t_*}^{t^*}\Phi_0(t)P^{-1}\Phi_0^{\mathrm{T}}(t)dt. \tag{11}$$

Under Assumption 2, this matrix is nonsingular, which can easily be verified using the implicit controllability criterion [11]. Since λ_0 is uniquely defined, according to the stated maximum principle, the optimal control in the first basic problem is associated with a unique vector of adjoint variables.

4. SECOND BASIC PROBLEM

At the second stage of computations, we solve an infinite-horizon optimal control problem of the form

$$\begin{gathered}\frac{dy}{ds} = By + D_4u, \quad y(0) = B^{-1}D_4\overline{u}(t^*),\\ y(-\infty) = 0, \quad J_2(u) = \frac{1}{2}\int_{-\infty}^{0}u^{\mathrm{T}}Puds \to \min,\end{gathered} \tag{12}$$

further called the *second basic problem.*

Assumption 3. *The controllability condition*

$$\operatorname{rank}\left(D_4, BD_4, B^2D_4, \ldots, B^{n_2-1}D_4\right) = n_2$$

holds.

This assumption ensures the existence of admissible controls in the second basic problem, implying in turn that problem (12) has a unique solution [9] and is normal. In this case, Pontryagin's maximum principle [10] for the problem can be stated as follows: let $u^*(s)$ and $y^*(s)$, $s \leqslant 0$, be

an optimal control and an optimal trajectory, respectively, in problem (12); then there exists a solution $\Pi\psi(s)$, $s \leqslant 0$, of the adjoint system

$$\frac{d}{ds}\Pi\psi = -B^{\mathrm{T}}\Pi\psi$$

such that

$$\Pi\psi^{\mathrm{T}}(s)D_4u^*(s) - \frac{1}{2}u^{*\mathrm{T}}(s)Pu^*(s) = \max_{u\in R^r}\left(\Pi\psi^{\mathrm{T}}(s)D_4u - \frac{1}{2}u^{\mathrm{T}}Pu\right), \quad s \leqslant 0.$$

This condition directly yields

$$u^*(s) = P^{-1}D_4^{\mathrm{T}}\Pi\psi(s), \quad s \leqslant 0. \tag{13}$$

With $\sigma_0 = \Pi\psi(0)$, we have

$$\Pi\psi^{\mathrm{T}}(s) = \sigma_0^{\mathrm{T}}G(s), \quad \Pi\psi^{\mathrm{T}}(s)D_4 = \sigma_0^{\mathrm{T}}\Pi\Phi(s), \quad s \leqslant 0, \tag{14}$$

where

$$\Pi\Phi(s) = G(s)D_4, \quad s \leqslant 0, \tag{15}$$

and $G(s)$, $s \leqslant 0$, is a matrix function of dimensions $(n_2 \times n_2)$ satisfying the differential equation

$$\frac{dG}{ds} = -GB, \quad G(0) = E_{n_2}. \tag{16}$$

By the implicit controllability criterion [11], under Assumption 3, the matrix

$$S_2 = \int\limits_{-\infty}^{0} \Pi\Phi(s)P^{-1}\Pi\Phi^{\mathrm{T}}(s)ds \tag{17}$$

is nonsingular. Based on (13), (14), and the Cauchy formula, we obtain

$$y^*(0) = S_2\sigma_0, \tag{18}$$

and, according to the above maximum principle, the optimal control in the second basic problem is associated with a unique vector of adjoint variables.

5. THIRD BASIC PROBLEM

The third basic problem has the form

$$\begin{gathered}\frac{dz}{d\tau} = C_3z + D_3u, \quad z(0) = C_3^{-1}D_3((\overline{u}(t^*) + u^*(0)),\\ z(-\infty) = 0, \quad J_3(u) = \frac{1}{2}\int\limits_{-\infty}^{0} u^{\mathrm{T}}Pud\tau \to \min.\end{gathered} \tag{19}$$

Assumption 4. *The controllability condition*

$$\operatorname{rank}\left(D_3, C_3D_3, C_3^2D_3, \ldots, C_3^{n_3-1}D_3\right) = n_3$$

holds.

This assumption ensures the existence of admissible controls in the third basic problem. Therefore, problem (19) has a unique solution [9] and is normal. In this case, Pontryagin's maximum principle [10] for the problem can be stated as follows: let $u_*(\tau)$ and $z_*(\tau)$, $\tau \leqslant 0$, be an optimal control and an optimal trajectory, respectively, in problem (19); then there exists a solution $Q\psi(\tau)$, $\tau \leqslant 0$, of the adjoint system

$$\frac{d}{d\tau}Q\psi = -C_3^{\mathrm{T}}Q\psi$$

such that

$$Q\psi^{\mathrm{T}}(\tau)D_3u_*(\tau)-\frac{1}{2}u_*^{\mathrm{T}}(\tau)Pu_*(\tau)=\max_{u\in R^r}\left(Q\psi^{\mathrm{T}}(\tau)D_3u-\frac{1}{2}u^{\mathrm{T}}Pu\right),\quad \tau\leqslant 0.$$

This condition directly implies

$$u_*(\tau)=P^{-1}D_3^{\mathrm{T}}Q\psi(\tau),\quad \tau\leqslant 0. \tag{20}$$

With $\chi_0=Q\psi(0)$, we have

$$Q\psi^{\mathrm{T}}(\tau)=\chi_0^{\mathrm{T}}G_1(\tau),\quad Q\psi^{\mathrm{T}}(\tau)D_3=\chi_0^{\mathrm{T}}Q\Phi(\tau),\quad \tau\leqslant 0, \tag{21}$$

where

$$Q\Phi(\tau)=G_1(\tau)D_3,\quad \tau\leqslant 0, \tag{22}$$

and $G_1(\tau)$, $\tau\leqslant 0$, is a matrix function of dimensions $(n_3\times n_3)$ satisfying the differential equation

$$\frac{dG_1}{d\tau}=-G_1C_3,\quad G_1(0)=E_{n_3}. \tag{23}$$

By the implicit controllability criterion [11], under Assumption 4, the matrix

$$S_3=\int\limits_{-\infty}^{0}Q\Phi(\tau)P^{-1}Q\Phi^{\mathrm{T}}(\tau)d\tau \tag{24}$$

is nonsingular. Based on (20), (21), and the Cauchy formula, we obtain

$$z_*(0)=S_3\chi_0, \tag{25}$$

and, according to the above maximum principle, the optimal control in the third basic problem is associated with a unique vector of adjoint variables.

Let us emphasize: the initial values σ_0 and χ_0 of the adjoint variable vectors are the only information about the solutions of the second and third basic problems used below to design asymptotically suboptimal controls. There is no need to construct the optimal controls $u^*(s)$, $s\leqslant 0$, and $u_*(\tau)$, $\tau\leqslant 0$, which (incidentally) is impossible during the numerical solution of the problem. In the second and third basic problems, the Hamiltonian along the optimal control is zero. By utilizing this property at $s=0$ and $\tau=0$, we establish the conditions for the vectors σ_0 and χ_0 :

$$\frac{1}{2}\sigma_0^{\mathrm{T}}D_4P^{-1}D_4^{\mathrm{T}}\sigma_0+\sigma_0D_4\overline{u}(t^*)=0,$$
$$\frac{1}{2}\chi_0^{\mathrm{T}}D_3P^{-1}D_3^{\mathrm{T}}\chi_0+\chi_0D_3(\overline{u}(t^*)+P^{-1}D_4^{\mathrm{T}}\sigma_0)=0.$$

After solving the basic problems, we find the vectors

$$\begin{aligned}\nu_0&=\sigma_0-\left((B_1-C_1C_3^{-1}B_3)B^{-1}\right)^{\mathrm{T}}\lambda_0,\\ \eta_0&=\chi_0-\left(C_1C_3^{-1}\right)^{\mathrm{T}}\lambda_0+\left(C_2C_3^{-1}\right)^{\mathrm{T}}\nu_0\end{aligned} \tag{26}$$

and compile the matrix

$$I_0=\begin{pmatrix}S_1 & 0_{n_1\times n_2} & 0_{n_1\times n_3}\\ I_{21} & S_2 & 0_{n_2\times n_3}\\ I_{31} & I_{32} & S_3\end{pmatrix} \tag{27}$$

of dimensions $(n_1+n_2+n_3)\times(n_1+n_2+n_3)$. The blocks S_1, S_2, and S_3 are given by formulas (11), (17), and (24), and

$$I_{21}=-B^{-1}\left(A_2-C_2C_3^{-1}A_3\right)S_1+\left(D_2-C_2C_3^{-1}D_3\right)+S_2\left(B^{-1}D_4\right)^{\mathrm{T}},$$
$$I_{31}=-C_3^{-1}\left(A_3S_1+B_2I_{21}+D_3P^{-1}D_0\right)+S_3\left(C_3^{-1}D_3\right)^{\mathrm{T}},$$
$$I_{32}=-C_3^{-1}\left(B_2S_2+D_3P^{-1}D_0\right)+S_3\left(C_3^{-1}D_3\right)^{T}.$$

As noted above, S_1, S_2, and S_3 are nonsingular matrices, and hence $\det I_0\neq 0$.

6. ASYMPTOTIC ANALYSIS OF THE SOLUTION OF THE ORIGINAL PROBLEM

One can speak of asymptotically suboptimal controls only if problem (1)–(3) is solvable. We will show that under Assumptions 1–4, problem (1)–(3) with a sufficiently small number μ has a unique optimal control representing a normal extremal. In other words, Pontryagin's maximum principle [10] can be stated as follows: let $u^0(t,\mu)$, $t\in T$, be an optimal control in problem (1)–(3); then there exists a solution $\psi_1^0(t,\mu)$, $\psi_2^0(t,\mu)$, $\psi_3^0(t,\mu)$, $t\in T$, of the adjoint system

$$\begin{aligned}\dot\psi_1&=-A_1^{\mathrm{T}}\psi_1-A_2^{\mathrm{T}}\psi_2-A_3^{\mathrm{T}}\psi_3,\\ \mu\dot\psi_2&=-B_1^{\mathrm{T}}\psi_1-B_2^{\mathrm{T}}\psi_2-B_3^{\mathrm{T}}\psi_3,\\ \mu^2\dot\psi_3&=-C_1^{\mathrm{T}}\psi_1-C_2^{\mathrm{T}}\psi_2-C_3^{\mathrm{T}}\psi_3\end{aligned}\tag{28}$$

such that

$$u^0(t,\mu)=P^{-1}\left(D_1^{\mathrm{T}}\psi_1(t,\mu)+D_2^{\mathrm{T}}\psi_2(t,\mu)+D_3^{\mathrm{T}}\psi_3(t,\mu)\right),\quad t\in T.\tag{29}$$

Theorem 1. *Under Assumptions 1–4, problem* (1)–(3) *has a unique solution* $u^0(t,\mu)$, $t\in T$, *representing a normal extremal of the form* (29), *where* $(\psi_1^0(t,\mu),\psi_2^0(t,\mu),\psi_3^0(t,\mu))$, $t\in T$, *is the vector of adjoint variables corresponding, by the maximum principle, to the optimal control. The values* $\lambda(\mu)=\psi_1^0(t^*,\mu)$, $\nu(\mu)=\psi_2^0(t^*,\mu)$, *and* $\eta(\mu)=\psi_3^0(t^*,\mu)$ *admit the asymptotic expansions*

$$\lambda(\mu)\sim\lambda_0+\sum_{k=1}^{\infty}\mu^k\lambda_k,\ \ \nu(\mu)\sim\nu_0+\sum_{k=1}^{\infty}\mu^k\nu_k,\ \ \eta(\mu)\sim\eta_0+\sum_{k=1}^{\infty}\mu^k\eta_k,\tag{30}$$

in which ν_0 *and* η_0 *are given by formulas* (26) *while* λ_0, σ_0, *and* χ_0 *are the initial values of the adjoint variables in the basic problems.*

Proof. We introduce the vectors $\theta=(\lambda,\nu,\eta)$ and $\theta_0=(\lambda_0,\nu_0,\eta_0)$ and denote by $\psi_1(t,\theta,\mu)$, $\psi_2(t,\theta,\mu)$, $\psi_3(t,\theta,\mu)$, $t\in T$, $\theta\in R^{n_1+n_2+n_3}$, the solution of the adjoint system (28) with the initial conditions

$$\psi_1(t^*)=\lambda,\quad \psi_2(t^*)=\nu,\quad \psi_3(t^*)=\eta.\tag{31}$$

Let $x(t,\theta,\mu)$, $y(t,\theta,\mu)$, $z(t,\theta,\mu)$, $t\in T$, $\theta\in R^{n_1+n_2+n_3}$, be the trajectory of system (1) generated by the control

$$u(t,\theta,\mu)=P^{-1}\left(D_1^{\mathrm{T}}\psi_1(t,\theta,\mu)+D_2^{\mathrm{T}}\psi_2(t,\theta,\mu)+D_3^{\mathrm{T}}\psi_3(t,\theta,\mu)\right),\quad t\in T.\tag{32}$$

We will show that the left-hand sides of the equations

$$x(t^*,\theta,\mu)=0,\quad y(t^*,\theta,\mu)=0,\quad z(t^*,\theta,\mu)=0\tag{33}$$

can be asymptotically expanded in integer powers of the small parameter, and obtain formulas for the coefficients of these expansions.

Consider the fundamental matrix $F(t,\mu)$, $t\in T$, of dimensions $(n_1+n_2+n_3)\times(n_1+n_2+n_3)$ as the solution of the initial-value problem

$$\dot F=-FA(\mu),\quad F(t^*)=E_{n_1+n_2+n_3}. \tag{34}$$

It is convenient to represent the solution of this singularly perturbed equation in the block form

$$F(t,\mu)=\begin{pmatrix} F_1(t,\mu) & F_2(t,\mu) & F_3(t,\mu)\\ F_4(t,\mu) & F_5(t,\mu) & F_6(t,\mu)\\ F_7(t,\mu) & F_8(t,\mu) & F_9(t,\mu)\end{pmatrix},$$

where F_i, $i=\overline{1,9}$, are matrices of dimensions $n_1\times n_1$, $n_1\times n_2$, $n_1\times n_3$, $n_2\times n_1$, $n_2\times n_2$, $n_2\times n_3$, $n_3\times n_1$, $n_3\times n_2$, and $n_3\times n_3$, respectively. Using the method of boundary functions [12], they can be expanded into asymptotic series:

$$\begin{gathered} F_i(t,\mu)\sim\sum_{k=0}^{\infty}\mu^k(F_{ik}(t)+\Pi_kF_i(s)+Q_kF_i(\tau)),\\ s=(t-t^*)/\mu,\quad \tau=(t-t^*)/\mu^2,\quad t\in T,\quad i=\overline{1,9}.\end{gathered} \tag{35}$$

We emphasize that these are asymptotic expansions uniform with respect to $t\in T$. Note also that the matrix functions $\Pi_kF_i(s)$, $s\leqslant 0$, and $Q_kF_i(\tau)$, $\tau\leqslant 0$, called the boundary terms, have the upper bounds

$$\|\Pi_kF_i(s)\|\leqslant\beta_{1k}\exp(\gamma_{1k}s),\quad \|Q_kF_i(s)\|\leqslant\beta_{2k}\exp(\gamma_{2k}s) \tag{36}$$

for all $i=\overline{1,9}$, $k=0,1,\dots$, where β_{1k}, β_{2k}, γ_{1k}, and γ_{2k} are positive constants. Let us provide several leading coefficients of the expansions (35), needed to prove the theorem:

$$\begin{gathered} F_{10}=F_0,\quad F_{40}=KF_0,\quad F_{70}=MF_0,\quad F_{i0}=0,\quad i=2,3,5,6,8,9,\\ F_{21}=F_0L,\quad F_{51}=KF_0L,\quad F_{81}=MF_0L,\quad F_{i1}=0,\quad i=3,6,9,\\ F_{32}=-F_0(C_1+LC_2)C_3^{-1},\quad F_{62}=-KF_0(C_1+LC_2)C_3^{-1},\quad F_{92}=-MF_0(C_1+LC_2)C_3^{-1},\\ \Pi_0F_4=-GK,\quad \Pi_0F_5=G,\quad \Pi_0F_7=C_3^{-1}B_3GK,\\ \Pi_0F_8=-C_3^{-1}B_3G,\quad \Pi_0F_i=0,\quad i=1,2,3,6,9,\\ \Pi_1F_2=-LG,\quad \Pi_1F_3=0,\quad \Pi_1F_6=-GC_2C_3^{-1},\\ \Pi_1F_9=C_3^{-1}B_3GC_2C_3^{-1},\quad \Pi_2F_3=LGC_2C_3^{-1},\\ Q_0F_i=0,\quad i=\overline{1,6},\quad Q_0F_7=G_1C_3^{-1}A_3,\\ Q_0F_8=G_1C_3^{-1}B_3,\quad Q_9F_0=G_1,\\ Q_1F_2=0,Q_1F_3=0,\quad Q_1F_5=C_2C_3^{-1}G_1C_3^{-1}B_3,\\ Q_1F_6=C_2C_3^{-1}G_1,\quad Q_2F_3=C_1C_3^{-1}G_1.\end{gathered} \tag{37}$$

Here, $K=-B^{-1}(A_2-C_2C_3^{-1}A_3)$, $M=-C_3^{-1}(A_3-B_3K)$, and $L=-(B_1-C_1C_3^{-1}B_3)B^{-1}$, while F_0, G, and G_1 are the solutions of equations (9), (16), and (23), respectively.

Note that

$$\begin{aligned}\psi_1^{\mathrm T}(t,\theta,\mu)&=\lambda^{\mathrm T}F_1(t,\mu)+\mu\nu^{\mathrm T}F_4(t,\mu)+\mu^2\eta^{\mathrm T}F_7(t,\mu),\\ \mu\psi_2^{\mathrm T}(t,\theta,\mu)&=\lambda^{\mathrm T}F_2(t,\mu)+\mu\nu^{\mathrm T}F_5(t,\mu)+\mu^2\eta^{\mathrm T}F_8(t,\mu),\\ \mu^2\psi_3^{\mathrm T}(t,\theta,\mu)&=\lambda^{\mathrm T}F_3(t,\mu)+\mu\nu^{\mathrm T}F_6(t,\mu)+\mu^2\eta^{\mathrm T}F_9(t,\mu).\end{aligned}$$

Therefore, due to (35) and (37), we arrive at the asymptotic expansions

$$\begin{gathered}\psi_i(t,\theta,\mu)\sim\sum_{k=0}^{\infty}\mu^k\left(\psi_{ik}(t,\theta)+\Pi_k\psi_i(s,\theta)+Q_k\psi_i(\tau,\theta)\right),\\ s=(t-t^*)/\mu,\quad \tau=(t-t^*)/\mu^2,\quad t\in T,\ i=1,2,3,\end{gathered}\tag{38}$$

where

$$\begin{gathered}\psi_{10}^{\mathrm{T}}(t,\theta)=\lambda^{\mathrm{T}}F_0(t),\quad \psi_{20}^{\mathrm{T}}(t,\theta)=\lambda^{\mathrm{T}}F_0(t)L,\\ \psi_{30}^{\mathrm{T}}(t,\theta)=-\lambda^{\mathrm{T}}F_0(t)(C_1-LC_2)C_3^{-1},\\ \Pi_0^{\mathrm{T}}\psi_1(s,\theta)=0,\quad \Pi_0^{\mathrm{T}}\psi_2(s,\theta)=-\left(\lambda^{\mathrm{T}}L-\nu^{\mathrm{T}}\right)G(s),\\ \Pi_0^{\mathrm{T}}\psi_3(s,\theta)=\left(\lambda^{\mathrm{T}}L-\nu^{\mathrm{T}}\right)G(s)C_2C_3^{-1},\\ Q_0^{\mathrm{T}}\psi_1(\tau,\theta)=0,\quad Q_0^{\mathrm{T}}\psi_2(\tau,\theta)=0,\\ Q_0^{\mathrm{T}}\psi_3(\tau,\theta)=\left(\lambda^{\mathrm{T}}C_1C_3^{-1}+\nu^{\mathrm{T}}C_2C_3^{-1}+\eta^{\mathrm{T}}\right)G_1(\tau),\\ \psi_{1k}^{\mathrm{T}}(t,\theta)=\lambda^{\mathrm{T}}F_{1k}(t)+\nu^{\mathrm{T}}F_{4,k-1}(t)+\eta^{\mathrm{T}}F_{7,k-2}(t),\\ \psi_{2k}^{\mathrm{T}}(t,\theta)=\lambda^{\mathrm{T}}F_{2,k+1}(t)+\nu^{\mathrm{T}}F_{5,k}(t)+\eta^{T}F_{8,k-1}(t),\\ \psi_{3k}^{\mathrm{T}}(t,\theta)=\lambda^{\mathrm{T}}F_{3,k+2}(t)+\nu^{\mathrm{T}}F_{6,k+1}(t)+\eta^{T}F_{9,k}(t),\\ \Pi_k^{\mathrm{T}}\psi_1(s,\theta)=\lambda^{\mathrm{T}}\Pi_kF_1(s)+\nu\Pi_{k-1}F_4(s)+\eta^{\mathrm{T}}\Pi_{k-2}F_7(s),\\ \Pi_k^{\mathrm{T}}\psi_2(s,\theta)=\lambda^{\mathrm{T}}\Pi_{k+1}F_2(s)+\nu\Pi_kF_5(s)+\eta^{\mathrm{T}}\Pi_{k-1}F_8(s),\\ \Pi_k^{\mathrm{T}}\psi_3(s,\theta)=\lambda^{\mathrm{T}}\Pi_{k+2}F_3(s)+\nu\Pi_{k+1}F_6(s)+\eta^{\mathrm{T}}\Pi_kF_9(s),\\ Q_k^{\mathrm{T}}\psi_1(\tau,\theta)=\lambda^{\mathrm{T}}Q_kF_1(\tau)+\nu^{\mathrm{T}}Q_{k-1}F_4(\tau)+\eta^{\mathrm{T}}Q_{k-2}F_7(\tau),\\ Q_k^{\mathrm{T}}\psi_2(\tau,\theta)=\lambda^{\mathrm{T}}Q_{k+1}F_2(\tau)+\nu^{\mathrm{T}}Q_kF_5(\tau)+\eta^{\mathrm{T}}Q_{k-1}F_8(\tau),\\ Q_k^{\mathrm{T}}\psi_3(\tau,\theta)=\lambda^{\mathrm{T}}Q_{k+2}F_3(\tau)+\nu^{\mathrm{T}}Q_{k+1}F_6(\tau)+\eta^{\mathrm{T}}Q_kF_9(\tau),\\ F_{7,-1}(t)=0,\quad \Pi_{-1}F_7(s)=0,Q_{-1}F_7(\tau)=0,\\ k\geqslant 1,\quad t\in T,\quad s\leqslant 0,\quad \tau\leqslant 0.\end{gathered}\tag{39}$$

These expansions are uniform in the domain $\|\theta-\theta_0\|\leqslant\epsilon_0$, $t\in T$, with a positive number ϵ_0. Note that from (7), (14), (21), (26), and (39) it follows that

$$\begin{gathered}\psi_{10}(t,\theta_0)=\psi^0(t),\quad \psi_{20}(t,\theta_0)=L^{\mathrm{T}}\psi^0(t),\\ \psi_{30}(t,\theta_0)=-\left((C_1-LC_2)C_3^{-1}\right)^{\mathrm{T}}\psi^0(t),\quad t\in T,\\ \Pi_0\psi_2(s,\theta_0)=\Pi\psi(s),\quad \Pi_0\psi_3(s,\theta_0)=-\left(C_2C_3^{-1}\right)^{\mathrm{T}}\Pi\psi(s),\quad s\leqslant 0,\\ Q_0\psi_3(\tau,\theta_0)=Q\psi(\tau),\quad \tau\leqslant 0.\end{gathered}\tag{40}$$

In view of (38), the control (29) also admits the asymptotic expansion

$$\begin{gathered}u(t,\theta,\mu)\sim\sum_{k=0}^{\infty}\mu^k\left(u_k(t,\theta)+\Pi_ku(s,\theta)+Q_ku(\tau,\theta)\right),\\ s=(t-t^*)/\mu,\quad \tau=(t-t^*)/\mu^2,\quad t\in T,\end{gathered}\tag{41}$$

where

$$\begin{gathered}u_k(t,\theta)=P^{-1}(D_1^{\mathrm{T}}\psi_{1k}(t,\theta)+D_2^{\mathrm{T}}\psi_{2k}(t,\theta)+D_3^{\mathrm{T}}\psi_{3k}(t,\theta)),\\ \Pi_k u(s,\theta)=P^{-1}(D_1^{\mathrm{T}}\Pi_k\psi_1(s,\theta)+D_2^{\mathrm{T}}\Pi_k\psi_2(s,\theta)+D_3^{\mathrm{T}}\Pi_k\psi_3(s,\theta)),\\ Q_k u(\tau,\theta)=P^{-1}(D_1^{\mathrm{T}}Q_k\psi_1(\tau,\theta)+D_2^{\mathrm{T}}Q_k\psi_2(\tau,\theta)+D_3^{\mathrm{T}}Q_k\psi_3(\tau,\theta)),\\ t\in T,\quad s\leqslant 0,\quad \tau\leqslant 0,\quad k=0,1,2,\dots.\end{gathered}\tag{42}$$

Along with formulas (5), (6), (13), (20), and (40), the conditions $\Pi_0\psi_1(s,\theta)=0$, $Q_0\psi_1(\tau,\theta)=0$, and $Q_0\psi_2(\tau,\theta)=0$ hold. Consequently,

$$\begin{gathered}u_0(t,\theta_0)=\overline{u}(t),\quad t\in T,\quad \Pi_0 u(s,\theta_0)=u^*(s),\quad s\leqslant 0,\\ Q_0u(\tau,\theta_0)=u_*(\tau),\quad \tau\leqslant 0.\end{gathered}\tag{43}$$

Based on the Cauchy formula, we have

$$\begin{pmatrix}x(t^*,\theta,\mu)\\ y(t^*,\theta,\mu)\\ z(t^*,\theta,\mu)\end{pmatrix}=F(t_*,\mu)\begin{pmatrix}x_*\\ y_*\\ z_*\end{pmatrix}+\int\limits_{t_*}^{t^*}F(t,\mu)D(\mu)u(t,\theta,\mu)dt.$$

According to (5), (35)–(39), (41), and (42), this vector function admits the asymptotic expansion

$$\begin{gathered}x(t^*,\theta,\mu)\sim\sum_{k=0}^{\infty}\mu^k x_k(\theta),\quad y(t^*,\theta,\mu)\sim\sum_{k=0}^{\infty}\mu^k y_k(\theta),\\ z(t^*,\theta,\mu)\sim\sum_{k=0}^{\infty}\mu^k z_k(\theta),\end{gathered}\tag{44}$$

where

$$\begin{gathered}x_0(\theta)=F_0(t_*)x_*+\int\limits_{t_*}^{t^*}F_0(t)D_0u_0(t,\theta)dt,\\ y_0(\theta)=Kx_0(\theta)+\int\limits_{-\infty}^{0}(G(s)D_4(u_0(t^*,\theta)+\Pi_0u(s,\theta))ds,\\ z_0(\theta)=Mx_0(\theta)+\int\limits_{-\infty}^{0}-C_3^{-1}B_3G(s)D_4(u_0(t^*,\theta)+\Pi_0u(s,\theta))ds\\ +\int\limits_{-\infty}^{0}G_1(\tau)D_3(u_0(t^*,\theta)+\Pi_0u(0,\theta)+Q_0u(\tau,\theta))d\tau,\\ x_k(\theta)=F_{1k}(t^*)x_*+F_{2k}(t^*)y_*+F_{3k}(t^*)z_*+\int\limits_{t_*}^{t^*}\sum_{p=0}^{k}(F_{1,k-p}(t)D_1u_p(t,\theta))dt\\ +\int\limits_{t_*}^{t^*}\sum_{p=0}^{k}(F_{2,k-p}(t)D_2u_p(t,\theta))dt+\int\limits_{t_*}^{t^*}\sum_{p=2}^{k+2}(F_{3,k+2-p}(t)D_3u_p(t,\theta))dt\\ +\int\limits_{-\infty}^{0}\sum_{p=0}^{k-1}((F_{1,k-1-p}(t^*)+\Pi_{k-1-p}F_1(s))D_1(u_p(t^*,\theta)+\Pi_pu(s,\theta)))ds\\ +\int\limits_{-\infty}^{0}\sum_{p=0}^{k}((F_{2,k-p}(t^*)+\Pi_{k-p}F_2(s))D_2(u_p(t^*,\theta)+\Pi_pu(s,\theta)))ds\end{gathered}\tag{45}$$

$$+\int_{-\infty}^{0}\sum_{p=0}^{k+1}((F_{3,k-p+1}(t^*)+\Pi_{k+1-p}F_3s))D_3(u_p(t^*,\theta)+\Pi_pu(s,\theta)))ds$$
$$+\int_{-\infty}^{0}\sum_{p=0}^{k-2}((F_{1,k-2-p}(t^*)+\Pi_{k-2-p}F_1(0)+Q_{k-2-p}F_1(\tau))D_1(u_p(t^*,\theta)$$
$$+\Pi_pu(0,\theta))+Q_pu(\tau,\theta))d\tau$$
$$+\int_{-\infty}^{0}\sum_{p=0}^{k-1}((F_{2,k-1-p}(t^*)+\Pi_{k-1-p}F_2(0)+Q_{k-1-p}F_2(\tau))D_2u_p(t^*,\theta)$$
$$+\Pi_pu(0,\theta))+Q_pu(\tau,\theta))d\tau$$
$$\int_{-\infty}^{0}\sum_{p=0}^{k}((F_{3,k-p}(t^*)+\Pi_{k-p}F_3(0)+Q_{k-p}F_3(\tau))D_3(u_p(t^*,\theta)$$
$$+\Pi_pu(0,\theta))+Q_pu(\tau,\theta))d\tau,\quad k\geqslant 1.$$

For $y_k(\theta)$, $z_k(\theta)$, $k\geqslant 1$, analogous formulas hold with the only difference being that the matrices F_{1i}, F_{2i}, F_{3i}, Π_iF_1, Π_iF_2, Π_iF_3, Q_iF_1, Q_iF_2, and Q_iF_3 are replaced by F_{4i}, F_{5i}, F_{6i}, Π_iF_4, Π_iF_5, Π_iF_6, Q_iF_4, Q_iF_5, and Q_iF_6 (in the first case) and by F_{7i}, F_{8i}, F_{9i}, Π_iF_7, Π_iF_8, Π_iF_9, Q_iF_7, Q_iF_8, and Q_iF_9 (in the second case).

Using the implicit function theorem, we verify that system (33) is uniquely solvable for θ under sufficiently small μ. It can be written in the form

$$R(\theta,\mu)=0. \tag{46}$$

According to (44), the uniform asymptotic expansion

$$R(\theta,\mu)\sim\sum_{k=0}^{\infty}\mu^kR_k(\theta) \tag{47}$$

holds in the domain $\|\theta-\theta_0\|\leqslant\epsilon_0$, where $R_k(\theta)=(x_k(\theta),y_k(\theta),z_k(\theta))$.

With $R(\theta,0)=R_0(\theta)$, the vector function $R(\theta,\mu)$ becomes continuous together with its partial derivatives with respect to the components of the vector θ in the domain $\|\theta-\theta_0\|<\epsilon_0$, $0\leqslant\mu<\mu_0$, where μ_0 is a sufficiently small positive number.

Due to (43), (45), and the Cauchy formula, we have $x_0(\theta_0)=x^0(t^*)=0$. The matrix function $G(s)$, $s\leqslant 0$, is a solution of the initial-value problem (16), and since $G(s)\to 0$ as $s\to-\infty$,

$$\int_{-\infty}^{0}G(s)D_4u_0(t^*,\theta_0)ds=\int_{-\infty}^{0}\left(-\frac{dG(s)}{ds}B^{-1}D_4(u_0(t^*,\theta_0)\right)ds=-B^{-1}D_4(u_0(t^*,\theta_0).$$

At the same time, conditions (43) and the Cauchy formula give

$$\int_{-\infty}^{0}G(s)D_4\Pi_0u(s,\theta_0)ds=-y^*(0)=B^{-1}D_4u_0(t^*,\theta_0).$$

Hence, from the second formula of (45), we obtain $y_0(\theta_0)=0$.

Similarly, since the matrix function $G_1(\tau)$, $\tau \leqslant 0$, is a solution of the initial-value problem (23) and $G_1(\tau) \to 0$ as $\tau \to -\infty$,

$$\int_{-\infty}^{0} G_1(\tau) D_3 \left(u_0(t^*, \theta_0) + \Pi_0 u(0, \theta_0)\right) d\tau$$
$$= \int_{-\infty}^{0} \left(-\frac{dG_1(\tau)}{d\tau} C_3^{-1} D_3 (u_0(t^*, \theta_0) + \Pi_0 u(0, \theta_0))\right) d\tau$$
$$= -C_3^{-1} D_3 (u_0(t^*, \theta_0) + \Pi_0 u(0, \theta_0)).$$

At the same time, from (43) and the Cauchy formula it follows that

$$\int_{-\infty}^{0} G_1(\tau) D_3 Q_0 u(\tau, \theta_0) d\tau = -z^*(0) = C_3^{-1} D_3 (u_0(t^*, \theta_0) + \Pi_0 u(0, \theta_0)).$$

Consequently, in view of the third formula of (45) and

$$\int_{-\infty}^{0} -C_3^{-1} B_3 G(s) D_4 (u_0(t^*, \theta) + \Pi_0(s, \theta)) ds = C_3^{-1} B_3 (K x_0(\theta_0) - y_0(\theta_0)) = 0,$$

we obtain $z_0(\theta_0) = 0$. Thus, $R(\theta_0, 0) = R_0(\theta_0) = 0$.

Taking into account (8), (15), (22), (37), (39), and (42), direct differentiation of the vector function (45) yields $\partial R(\theta_0, 0)/\partial\theta = \partial R_0(\theta_0)/\partial\theta = I_0$ (see (27)). Since this Jacobian matrix is nonsingular, the conditions of the implicit function theorem hold for system (46) or, equivalently, for system (33). According to this theorem, a vector function $\theta(\mu) = (\lambda(\mu), \nu(\mu), \eta(\mu))$ satisfying equations (33) is uniquely defined in some right-hand neighborhood $0 \leqslant \mu < \mu_1$ of zero; it is continuous, and $\theta(0) = \theta_0 = (\lambda_0, \nu_0, \eta_0)$. Therefore, the control $u^0(t, \mu) = u(t, \theta, \mu)$, $t \in T$, is admissible in problem (1)–(3) for sufficiently small μ, and it satisfies the maximum principle with the vector of adjoint variables $\psi_1^0(t, \mu) = \psi_1(t, \theta(\mu), \mu)$, $\psi_2^0(t, \mu) = \psi_2(t, \theta(\mu), \mu)$, $\psi_3^0(t, \mu) = \psi_3(t, \theta(\mu), \mu)$, $t \in T$; moreover, it represents a normal extremal. Hence, $u^0(t, \mu)$, $t \in T$, is the unique optimal control in problem (1)–(3).

As mentioned, the expansion (47) is uniform in the domain $\|\theta - \theta_0\| < \epsilon_0$, and its coefficients are linear functions. Then the solution $\theta(\mu) = (\lambda(\mu), \nu(\mu), \eta(\mu))$ of system (33) admits the asymptotic expansions (30). The proof of Theorem 1 is complete.

7. CONSTRUCTION OF ASYMPTOTICALLY SUBOPTIMAL CONTROLS

Using the theorem and related formulas, we present the algorithm for constructing asymptotic approximations to the solution of problem (1)–(3).

The vector function

$$u^{(0)}(t, \mu) = P^{-1}(D_1^{\mathrm{T}} \psi_1(t, \theta_0, \mu) + D_2^{\mathrm{T}} \psi_2(t, \theta_0, \mu) + D_3^{\mathrm{T}} \psi_3(t, \theta_0, \mu)), \quad t \in T,$$

is an asymptotically suboptimal control of zero order in the original problem. Note that it can be formed immediately after solving the basic problems. An asymptotically suboptimal control of order $N \geqslant 1$ has the form

$$u^{(N)}(t, \mu) = P^{-1}(D_1^{\mathrm{T}} \psi_1(t, \theta^{(N)}(\mu), \mu) + D_2^{\mathrm{T}} \psi_2(t, \theta^{(N)}(\mu), \mu) + D_3^{\mathrm{T}} \psi_3(t, \theta^{(N)}(\mu, \mu)), \quad t \in T, \quad (48)$$

where

$$\theta^{(N)}(\mu) = \sum_{k=0}^{N} \mu^k \theta_k, \quad \theta_k = (\lambda_k, \nu_k, \eta_k). \tag{49}$$

To construct the control (48), it is necessary to find the coefficients of the asymptotic series (30). This can be done by the method of undetermined coefficients with the expansion (47). More precisely, we expand, using Taylor's formula, the vector function

$$\sum_{k=0}^{N} \mu^k R_k(\theta^{(N)}(\mu))$$

in the powers of μ up to order N inclusive and equate the coefficients to zero, starting with that of μ. The resulting nonsingular systems of linear algebraic equations,

$$I_0\theta_1 = -R_1(\theta_0), \quad I_0\theta_k = -R_k(\theta_0) - \sum_{j=1}^{k-1} \frac{\partial R_j(\theta_0)}{\partial \theta}\theta_{k-j}, \quad k \geqslant 2, \tag{50}$$

serve to successively find the vectors θ_k, $k = \overline{1, N}$. (Here, we take into account that the coefficients of the expansions (47) are linear vector functions.) Note that, due to the structure (27) of the Jacobian matrix I_0, systems (50) decouple. Solving these systems successively yields the vectors θ_k, $k = \overline{1, N}$, and the polynomials (49). As mentioned above, the control (48) is asymptotically suboptimal of order N in the original problem. To construct this control, it is necessary to solve the Cauchy problem for the adjoint system (28) with the initial conditions

$$\psi_1(t^*) = \sum_{k=0}^{N} \mu^k \lambda_k, \quad \psi_2(t^*) = \sum_{k=0}^{N} \mu^k \nu_k, \quad \psi_3(t^*) = \sum_{k=0}^{N} \mu^k \eta_k.$$

The adjoint system is singularly perturbed and, consequently, stiff. Integration of stiff systems can be avoided by replacing the vector functions $\psi_i(t, \theta, \mu)$, $i = \overline{1, 3}$, in (48) by their asymptotic approximations

$$\psi_i^{(N)}(t, \theta, \mu) = \sum_{k=0}^{N} \mu^k (\psi_{ik}(t, \theta) + \Pi_k \psi_i(s, \theta) + Q_k \psi_i(\tau, \theta)),$$
$$s = (t - t*)/\mu, \quad \tau = (t - t^*)/\mu^2, \quad t \in T, \quad i = \overline{1, 3}.$$

The vector function

$$\overline{u}^{(N)}(t, \mu) = P^{-1}(D_1^{\mathrm{T}} \psi_1^{(N)}(t, \theta^{(N)}(\mu), \mu) + D_2^{\mathrm{T}} \psi_2^{(N)}(t, \theta^{(N)}(\mu), \mu) + D_3^{\mathrm{T}} \psi_3^{(N)}(t, \theta^{(N)}(\mu,)\mu)), \quad t \in T,$$

along with (48), is an asymptotically suboptimal control of order N in problem (1)–(3). In particular, it follows from (5), (6), (13), (20), (39), and (40) that

$$\begin{aligned} \overline{u}^{(0)}(t, \mu) &= P^{-1}\left(D_0^{\mathrm{T}} \psi^{(0)}(t) + D_4^{\mathrm{T}} \Pi\psi\left((t - t^*)/\mu\right) + D_3^{\mathrm{T}} Q\psi\left((t - t)/\mu^2\right)\right) \\ &= \overline{u}(t) + u^*\left((t - t^*)/\mu\right) + u_*\left((t - t)/\mu^2\right), \quad t \in T. \end{aligned} \tag{51}$$

Formulas (41) and (43) also indicate that this is an asymptotically suboptimal control of zero order. Note that the control (51) is independent of the initial states y_*, z_* of the fast variable vectors and, for small μ, substantially differs from the solution $\overline{u}(t)$, $t \in T$, of the first basic problem only in the boundary layer (i.e., in some left-hand neighborhood of the point t^*).

The constructed asymptotics can be used to find the optimal control $u^0(t, \mu)$, $t \in T$, in problem (1)–(3) for a given value of μ. For this purpose, one should follow the finishing procedure [13], i.e., find the roots of system (46) by Newton's method with $\theta^{(N)}(\mu)$ as an initial approximation. To avoid integrating stiff systems, one can take the approximation I_0 instead of the exact matrix $\partial R(\theta, \mu)/\partial\theta$.

8. ASYMPTOTICALLY SUBOPTIMAL DESIGN

Asymptotically suboptimal open-loop controls, of course, depend on the initial state (x_*, y_*, z_*, t_*) of the dynamic system. Previously, this dependence has been neglected by supposing a given initial state. In the current section, devoted to the asymptotically suboptimal design of a feedback control of zero order, we will be concerned with this dependence. From now on, let the matrix C be a function of t_*, $t_* < t^*$. The time instant t^* is still assumed given.

From equality (10) we have

$$\psi^0(t^*) = \lambda_0 = -S_1^{-1}(t_*)F_0(t_*)x_*, \tag{52}$$

which implies, together with (7),

$$\psi^0(t_*) = -F_0^{\mathrm{T}}(t_*)S_1^{-1}(t_*)F_0(t_*)x_*. \tag{53}$$

Due to (6) and (18),

$$\Pi\psi(0) = \sigma_0 = S_2^{-1}y^*(0) = -S_2^{-1}B^{-1}D_4\overline{u}(t^*) = -S_2^{-1}B^{-1}D_4P^{-1}D_0^{\mathrm{T}}\psi^0(t^*). \tag{54}$$

Hence, taking (14) and (52) into account, we obtain

$$\Pi\psi(s) = G^{\mathrm{T}}(s)S_2^{-1}B^{-1}D_4P^{-1}D_0^{\mathrm{T}}F_0^{\mathrm{T}}(t_*)S_1^{-1}(t_*)F_0(t_*)x_*, \quad s \leqslant 0. \tag{55}$$

In view of (6), (13), and (25),

$$\begin{aligned} Q\psi(0) = \chi_0 = S_3^{-1}z^*(0) &= -S_3^{-1}C_3^{-1}D_3(\overline{u}(t^*) + u^*(0)) \\ &= -S_3^{-1}C_3^{-1}D_3P^{-1}(D_0^{\mathrm{T}}\psi(t^*) + D_4^{\mathrm{T}}\mathrm{T}\Pi\psi(0)). \end{aligned}$$

Considering (21), (52), and (54), the above expression yields

$$Q\psi(\tau) = G_1^{\mathrm{T}}(\tau)S_3^{-1}C_3^{-1}D_3P^{-1}(D_0^{\mathrm{T}} + D_4^{\mathrm{T}}S_2^{-1}B^{-1}D_4P^{-1}D_0^{\mathrm{T}})C^{-1}(t_*)F_0(t_*)x_*.$$

According to formulas (51), (53), and (55), the asymptotically suboptimal control of zero order at the initial time instant can be represented as

$$\begin{aligned} \overline{u}^{(0)}(t_*, \mu) = P^{-1}\left(D_0^{\mathrm{T}} - D_4^{\mathrm{T}}G^{\mathrm{T}}((t_* - t*)/\mu\right)S_2^{-1}B^{-1}D_4P^{-1}D_0^{\mathrm{T}}F_0^{\mathrm{T}}(t_*) \\ -D_3^{\mathrm{T}}G_1\left((t_* - t^*)/\mu^2\right)S_3^{-1}C_3^{-1}D_3P^{-1}\left(D_0^{\mathrm{T}} + D_4^{\mathrm{T}}S_2^{-1}B^{-1}D_4P^{-1}D_0^{\mathrm{T}})\right)S_1^{-1}(t_*)F_0(t_*)x_*. \end{aligned}$$

Since (x_*, y_*, z_*, t_*) is an arbitrary initial state of the dynamic system, by Definition 2, the vector function

$$\begin{aligned} \overline{u}^{(0)}(x, y, z, t, \mu) = P^{-1}\Big(D_0^{\mathrm{T}} - D_4^{\mathrm{T}}G^{\mathrm{T}}\left((t - t*)/\mu\right)S_2^{-1}B^{-1}D_4P^{-1}D_0^{\mathrm{T}}F_0^{\mathrm{T}}(t) \\ - D_3^{\mathrm{T}}G_1\left((t - t^*)/\mu^2\right)S_3^{-1}C_3^{-1}D_3P^{-1}\left(D_0^{\mathrm{T}} + D_4^{\mathrm{T}}C_0^{-1}B^{-1}D_4P^{-1}D_0^{\mathrm{T}}\right)\Big)S_1^{-1}(t)F_0(t)x \end{aligned} \tag{56}$$

represents an asymptotically suboptimal feedback control of zero order in the original problem. Note that this feedback is independent of the current position of the fast variables.

9. EXAMPLE

As an illustrative example, we consider a control model for the horizontal rectilinear motion of two bodies with significantly different small masses and an elastic coupling:

$$\begin{gathered}\dot{x}_1 = y, \quad \dot{x}_2 = z,\\ \mu\dot{y} = -kx_1 + kx_2 - c_1 y + d_1 u_1, \quad \mu^2\dot{z} = kx_1 - kx_2 - c_2 z + d_2 u_2,\\ x_1(t_*) = x_{*1}, \quad x_2(t_*) = x_{*2}, \quad y(t_*) = y_*, \quad z(t_*) = z_*,\\ x_1(t^*) = x_2(t^*) = y(t^*) = z(t^*) = 0,\\ J(u) = \frac{1}{2}\int\limits_{t_*}^{t^*}(u_1^2 + u_2^2)dt \to \min.\end{gathered} \tag{57}$$

All constants in this problem are positive, and $\mu \ll 1$. Assumption 1 holds in this case.

The first basic problem has the form

$$\begin{gathered}\dot{x}_1 = -\frac{k}{c_1}x_1 + \frac{k}{c_1}x_2 + \frac{d_1}{c_1}u_1, \quad \dot{x}_2 = \frac{k}{c_2}x_1 - \frac{k}{c_2}x_2 + \frac{d_2}{c_2}u_2,\\ x_1(t_*) = x_{*1}, \quad x_2(t_*) = x_{*2},\\ x_1(t^*) = x_2(t^*) = 0, \quad J_1(u) = \frac{1}{2}\int\limits_{t_*}^{t^*}(u_1^2 + u_2^2)dt \to \min.\end{gathered} \tag{58}$$

As is easily verified, the dynamic system in the first basic problem is completely controllable. According to the maximum principle, the optimal control in problem (58) has the form

$$\begin{gathered}\overline{u}_1(t) = \frac{d_1}{c_1}\left(\frac{c_1(\lambda_{01}+\lambda_{02}) - (c_1\lambda_{02} - c_2\lambda_{01})\exp((k(c_1+c_2)(t-t_*))/c_1c_2)}{c_1+c_2}\right),\\ \overline{u}_2(t) = \frac{d_2}{c_2}\left(\frac{c_2(\lambda_{01}+\lambda_{02}) + (c_1\lambda_{02} - c_2\lambda_{01})\exp((k(c_1+c_2)(t-t_*))/c_1c_2)}{c_1+c_2}\right),\\ t \in [t_*, t^*].\end{gathered} \tag{59}$$

The vector $\lambda_0 = (\lambda_{01}, \lambda_{02})$ is calculated by

$$\lambda_0 = S_1^{-1}F_0(t^*)x_*, \tag{60}$$

where

$$\begin{gathered}F_0(t) = \begin{pmatrix} \dfrac{c_2\exp((k(c_1+c_2)(t^*-t))/c_1c_2) - c_1}{c_2-c_1} & \dfrac{c_2\exp((k(c_1+c_2)(t^*-t))/c_1c_2) - c_2}{c_1-c_2}\\ \dfrac{c_1\exp((k(c_1+c_2)(t^*-t))/c_1c_2) - c_1}{c_2-c_1} & \dfrac{c_1\exp((k(c_1+c_2)(t^*-t))/c_1c_2) - c_2}{c_1-c_2}\end{pmatrix},\\ t \in [t_*, t^*],\\ S_1 = \int\limits_{t_*}^{t^*} F_0(t)\begin{pmatrix} d_1^2/c_1^2 & 0\\ 0 & d_2^2/c_2^2\end{pmatrix}F_0^T(t)dt, \quad x_* = \begin{pmatrix} x_{*1}\\ x_{*2}\end{pmatrix}.\end{gathered}$$

After solving problem (58), we form the second basic problem

$$\begin{gathered}\frac{dy}{ds} = -c_1 y + d_1 u_1, \quad y(0) = -\frac{d_1}{c_1}\overline{u}(t^*),\\ y(-\infty) = 0, \quad J_2(u) = \frac{1}{2}\int\limits_{-\infty}^{0}(u_1^2 + u_2^2)ds \to \min.\end{gathered} \tag{61}$$

Here, $\overline{u}(t)$, $t \in [t_*, t^*]$, is the solution of problem (58). The dynamic system in the second basic problem is completely controllable. By the maximum principle, the solution of problem (61) is as follows:

$$u_1^*(s) = -2\overline{u}_1(t^*)\exp(c_1 s), \quad u_2^*(s) = 0, \quad s \leqslant 0.$$

The third basic problem has the form

$$\frac{dz}{d\tau} = -c_2 z + d_2 u_2, \quad z(0) = -\frac{d_2}{c_2}\overline{u}_2(t^*), \quad z(-\infty) = 0, \quad J_2(u) = \frac{1}{2}\int\limits_{-\infty}^{0}(u_1^2 + u_2^2)d\tau \to \min.$$

The dynamic system in this problem is also completely controllable. Based on the maximum principle, the optimal control in the third basic problem can be represented as

$$u_{*1}(\tau) = 0, \quad u_{*2}(\tau) = -2\overline{u}_2(t^*)\exp(c_2\tau), \quad \tau \leqslant 0.$$

According to (51), the control

$$\begin{gathered}\overline{u}_1^0(t,\mu) = \overline{u}_1(t) - 2\overline{u}_1(t^*)\exp(c_1(t-t^*)/\mu),\\ \overline{u}_2^0(t,\mu) = \overline{u}_2(t) - 2\overline{u}_2(t^*)\exp(c_2(t-t^*)/\mu^2), \quad t \in [t_*, t^*],\end{gathered} \tag{62}$$

is asymptotically suboptimal of zero order in problem (57).

When constructing an asymptotically suboptimal feedback control of zero order, we consider the matrix C in (60) as a function of t_* : $C = C(t_*)$. In the example,

$$D_0 = \begin{pmatrix} d_1/c_1 & 0 \\ 0 & d_2/c_2 \end{pmatrix}.$$

Then, due to (56), the vector function

$$\overline{u}^{(0)}(x,t,\mu) = D_0(E_2(1 - 2\exp((t-t^*)/\mu^2)) + 2\exp((t-t^*)/\mu)F_0^{\mathrm{T}}(t))C^{-1}(t)F_0(t)x$$

represents an asymptotically suboptimal feedback control of zero order in problem (57).

To estimate the quality of the above asymptotic approximations to the solution of problem (57), we computed the resulting states $x_1(t^*,\mu)$, $x_2(t^*,\mu)$, $y(t^*,\mu)$, and $z(t^*,\mu)$ of the dynamic system with the control (62) for particular values of the small parameter in the case where $k = 1$, $c_1 = 3$, $c_2 = 1$, $d_1 = 1$, $d_2 = 2$, $t_* = 0$, $t^* = 5$, $x_{*1} = 5$, $x_{*2} = -3.0$, $y_* = 2.0$, and $z_* = 1$. As it turned out,

$$\begin{gathered}x_1(5,\ 0.1) = 0.068161 \qquad x_2(5,\ 0.1) = 0.033727\\ y(5,\ 0.1) = -0.023323 \qquad z(5,\ 0.1) = -0.011998\\ x_1(5,\ 0.025) = 0.016755 \quad x_2(5,\ 0.025) = 0.014029\\ y(5,\ 0.025) = -0.004879 \quad z(5,\ 0.025) = -0.000235.\end{gathered}$$

The numerical results are provided with an accuracy of 10^{-6}.

10. CONCLUSIONS

This paper has proposed computational procedures for constructing asymptotic approximations to the solution of the problem in the form of open-loop control and feedback control without the need to integrate singularly perturbed systems. With the proposed algorithms, the original optimal control problem decomposes into three unperturbed ones of lower dimension. Such a decomposition enables efficient solution of optimization problems for dynamic systems with a large number of state variables. Also, the constructed asymptotic approximations have been used to solve the problem for given values of the small parameter.

According to the proof of Theorem 1, under Assumptions 2–4, there exists an admissible control in the original problem for any initial state of the dynamic system. Therefore, these assumptions ensure the complete controllability of system (1) for sufficiently small values of μ. This result was first established in [14].

The method serves to study problems containing several groups of fast variables with a hierarchy of rates of change in integer powers of the small parameter. In this case, the number of problems in the decomposition of the original problem is equal to that of groups of multi-tempo variables. Such a generalization introduces nonfundamental modifications into the algorithm, which are easily traced using the example of the considered problem. At the same time, it leads to cumbersome formulas. Note also that the above results can be extended to time-varying systems with sufficiently smooth coefficients without fundamental difficulties.

REFERENCES

1. Dmitriev, M.G. and Kurina, G.A., Singular Perturbations in Control Problems, *Autom. Remote Control*, 2006, vol. 67, no. 1, pp. 1–43.
2. Kalinin, A.I., Asymptotics of Solutions of Perturbed Optimal Control Problems, *J. Comput. Syst. Sci. Int.*, 1995, vol. 33, no. 6, pp. 75–84.
3. Zhang, Y., Naidu, D.S., Cai, C., and Zou, Y., Singular Perturbation and Time Scales in Control Theories and Applications. An Overview 2002–2012, *Int. J. Information and Systems Sciences*, 2014, vol. 9, no. 1, pp. 1–36.
4. Kokotovic, P.V. and Khalil, H.K., *Singular Perturbations in Systems and Control*, New York: IEEE Press, 1986.
5. Kurina, G.A. and Kalashnikova, M.A., Singularly Perturbed Problems with Multi-Tempo Fast Variables, *Autom. Remote Control*, 2022, vol. 83, no. 11, pp. 1679–1723.
6. Kalinin, A.I. and Lavrinovich, L.I., Asymptotics of the Solution of the Transient Optimization Problem in a Three-Tempo Singularly Perturbed System, *Trudy XIV Belorusskoi matematicheskoi konferentsii, posvyashchennoi 65-letiyu Instituta matematiki NAN Belarusi* (Proceedings of XIV Belarusian Mathematical Conference Dedicated to the 65th Anniversary of the Institute of Mathematics of the NAS of Belarus), Minsk, October 28–November 1, 2024, Minsk: Belaruskaya Navuka, 2024, part 1, pp. 116–117.
7. Kalinin, A.I. and Lavrinovich, L.I., Application of the Small Parameter Method to the Singularly Perturbed Linear-Quadratic Optimal Control Problem, *Autom. Remote Control*, 2016, vol. 77, no. 5, pp. 751–763.
8. Krasovskii, N.N., *Teoriya upravleniya dvizheniem: lineinye sistemy* (Theory of Control of Motion), Moscow: Nauka, 1968.
9. Mordukhovich, B.Sh., Existence of Optimal Controls, *Journal of Soviet Mathematics*, 1977, vol. 7, no. 5, pp. 795–849.
10. Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., and Mishchenko, E.F., *The Mathematical Theory of Optimal Processes*, New York: Interscience Publishers, 1962.
11. Gabasov, R. and Kirillova, F.M., *Optimization of Linear Systems*, New York: Plenum Press, 1979.
12. Vasil'eva, A.B., Asymptotic Methods in the Theory of Ordinary Differential Equations with Small Parameters Multiplying the Highest Derivatives, *USSR. Comput. Math. Math. Phys.*, 1963, vol. 3, no. 4, pp. 611–642.
13. Gabasov, R. and Kirillova, F.M., *Konstruktivnye metody optimizatsii* (Constructive Methods of Optimization), part 2, Minsk: Universitetskoe, 1984.
14. Kurina, G.A., Complete Controllability of Various-Speed Singularly Perturbed Systems, *Math. Notes*, 1992, vol. 52, no. 4, pp. 1029–1033.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 10, pp. 901–925.
© 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. 10, pp. 52–83.

NONLINEAR SYSTEMS

Robust Prescribed Performance Control under Given Constraints for Systems with Relative Degree One

K. A. Lastochkin

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

Received August 25, 2025
Revised February 5, 2026
Accepted February 10, 2026

Abstract—The problem of prescribed performance control under given constraints on the amplitude and rate of control signal is considered for dynamic systems with relative degree one in the presence of parametric uncertainty and external disturbances. Using the ideas of feedback design based on sat-function differentiable approximation and augmented perturbation estimation, a novel nonlinear robust control law is proposed, which, if the parameters are suitably chosen, provides a solution to the problem under consideration. The parameter calculation procedure is reduced to solving linear inequalities, which feasibility is proved analytically. Illustrative example corroborates theoretical results.

Keywords: robust control, transients, prescribed performance, control signal constraints, saturation function

DOI: 10.7868/S1608303226100033

1. INTRODUCTION

The problem of prescribed performance control in the presence of input saturation for systems with unknown parameters and external perturbations remains open in modern control theory. Among the known attempts to address this problem, approaches should be mentioned that design a control law on the basis of *i*) barrier functions [1–8], which prevent the state/output from violating the prescribed bounds, and *ii*) smooth approximations of a saturation function [9–13], which ensure prescribed performance of the tracking error with respect to the reference signal using control signal with bounded amplitude.

The formal problem statement of the above-mentioned papers includes only control signal amplitude constraint, while its first derivative (rate) boundedness is not taken into account. However, actuators of real-world plants usually have limited rate of the control signal implementation. For example, considering aircraft, the maximum instantaneous deflection rate of control surfaces (ailerons, elevators, rudders, etc.) is always limited. If such constraints are not considered when designing the control law, then only a part of the algorithmically generated control signal will actually act the system. This can lead to violation of the prescribed transient performance specifications or even to loss of closed-loop stability. Therefore, in this paper, considering the prescribed performance control problem, it is proposed to take into account constraints on both the amplitude and the rate of the control signal.

The aim of this paper is to ensure prescribed transient performance under given constraints on both the control signal amplitude and rate for dynamic systems with relative degree one in the presence of parametric uncertainty and external disturbances. The class of systems with relative

degree one is considered for simplicity of presentation, and in the future the obtained results can be used to solve control problems for more complex systems.

To achieve this goal, a new control law is proposed, in which the proportional feedback, as in [9–13], is designed using a special differentiable approximation of the saturation function, but it is additionally filtered through a nonlinear filter of the type [14]. This technique allows one to *a priori* bound both the amplitude and the rate of the control signal within the prescribed limits without the need to require boundedness of the amplitude and first derivative of the system output. Furthermore, in contrast to [9–13], in addition to the classical proportional feedback, the control law is proposed to include a dynamic estimate of the augmented disturbance obtained on the basis of A.M. Tsykunov's auxiliary loop method [15, p. 201] that is combined with a nonlinear filter [14] to bound the rate of the augmented disturbance estimate within the given limits. It is shown in the paper that such a nonlinear robust feedback allows one to *a priori* bound the amplitude and rate of the control signal by given values and reduce the prescribed performance control problem with amplitude/rate constraints to the one of the control law parameters calculation via solution of linear inequalities.

The main distinctive features of the proposed control law compared to those in [1–13] are as follows:

1) in addition to ensure the prescribed transient performance, the amplitude and rate of the control signal are bounded within the given limits;
2) the choice of the control law parameters is reduced to solving linear inequalities and can be implemented numerically;
3) the feasibility of the obtained linear inequalities to calculate the control law parameters is shown analytically rather than numerically;
4) in contrast to [1–8], barrier functions are not used for feedback design, which eliminates discontinuities in the control law caused by division by zero and/or taking the natural logarithm of zero (see Fig. 4 and the comments thereto in Section 6);
5) in contrast to [9–13], the proposed control law *i*) contains not only a summand that is proportional to the tracking error, but also a dynamic estimate of the augmented disturbance designed by the auxiliary loop method [15, p. 201], and *ii*) is combined with nonlinear filters intended to bound the rate of the control signal.

The paper is organized as follows. Section 2 provides notation and definitions. Section 3 presents the problem statement and comments. Section 4 elucidates the main result of the paper and derives the control law that guarantees the prescribed transient performance under bounded amplitude and rate of the control signal. Section 5 provides an analytical comparison of the developed control law with the one proposed in [10]. Section 6 presents the results of numerical experiments. Section 7 formulates the conclusions and outlines future research scope.

2. PRELIMINARIES

The following notation is adopted: $t \geq t_0 \geq 0$ denotes time, $f : [t_0, \infty) \mapsto \mathbb{R}^n$ is a function of time, $f(t)$ stands for a value of the function f at the time instant t, the notation $F := \{f(t) \in \mathbb{R}^n : |f(t)| \leq f^{\mathrm{U}},\ |\dot{f}(t)| \leq v^{\mathrm{U}}\}$ denotes the space of functions of time whose amplitude and first derivative at time t are bounded by the numbers $f^{\mathrm{U}} > 0$, $v^{\mathrm{U}} > 0$, σ: $\mathbb{R} \mapsto \mathbb{R}$ is any differentiable approximation of the saturation function that meets the following conditions[1]:

a) $|\sigma'(s)| = \left|\frac{\partial \sigma(s)}{\partial s}\right| \leq \kappa \in (1, 2)$ for all $s \in \mathbb{R}$,

b) $s\sigma(s) > 0$ for all $s \neq 0$, $\sigma(0) = 0$,

[1] Hereafter $\operatorname{sat}\{s\} = \operatorname{sign}\{s\} \min\{|s|, 1\}, \operatorname{sign}\{s\} \in \begin{cases} 1, & s > 0 \\ [-1, 1], & s = 0 \\ -1, & s < 0. \end{cases}$

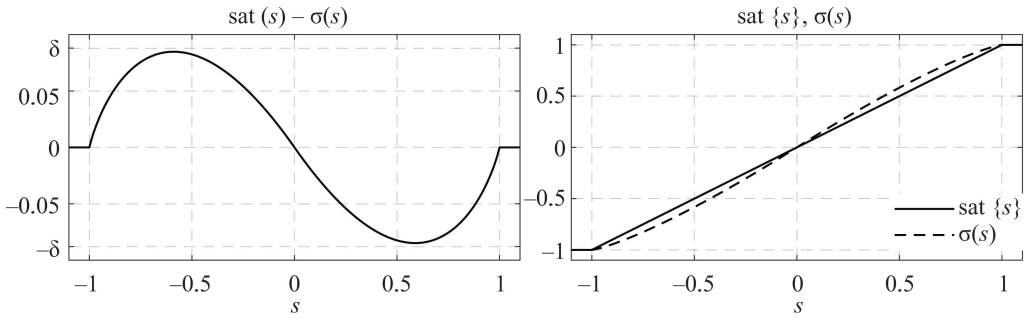


Fig. 1. Sat-function approximation error.

c) $\sigma(s) = \operatorname{sign}\{s\}$ for all $|s| \geq 1$,
d) $|s| < |\sigma(s)| < 1$ for $|s| < 1$,
e) there exists $\delta \in (0,\ 1)$ such that $|s - \sigma(s)| \leq \delta$ for all $|s| \leq 1$,
f) there exists $L > \kappa$ such that $|\sigma(a) - \sigma(b)| \leq L|a - b|$ for all $a,\ b \in \mathbb{R}$.

For instance, choosing $\kappa \simeq 1.26,\ \delta \simeq 0.088,\ L \simeq 1.3$, the following function meets the conditions *a*)–*f*):

$$\sigma(s) = \operatorname{sign}\{s\}\left(1 - (1 - \min\{|s|, 1\})^{1.26}\right).$$

Figure 1 demonstrates the comparison of sat-function with the above-given one.

3. PROBLEM STATEMENT

Consider a dynamic system

$$\dot{y}(t) = bu(t) + f(t, y, u),\ y(t_0) = y_0, \tag{3.1}$$

where $y(t) \in \mathbb{R}$ stands for a measurable output, $u(t) \in \mathbb{R}$ is a control signal, $f(t, y, u) \in \mathbb{R}$ denotes an unmeasurable augmented perturbation (a combination of parametric and external disturbances), $b > 0$ is a known estimate of the system gain, which is, without loss of generality, considered to be positive for simplicity.

The aim is to design a control law $u(t)$ for the system (3.1) in such a way that it holds for all $t \geq t_0$ that:

$$y(t) \in \mathcal{Y} := \left\{y(t) \in \mathbb{R}\colon \underline{g}(t) \leq y(t) \leq \overline{g}(t)\right\}, \tag{3.2}$$

$$u(t) \in \mathcal{U} := \left\{u(t) \in \mathbb{R}\colon\ |u(t)| \leq u^{\mathrm{U}},\ |\dot{u}(t)| \leq v^{\mathrm{U}}\right\}, \tag{3.3}$$

where $u^{\mathrm{U}} > 0,\ v^{\mathrm{U}} > 0$ are known scalars, $\underline{g}(t), \overline{g}(t)$ are measurable for all $t \geq t_0$.

Therefore, we consider the problem to ensure the prescribed transient performance for the system output using a control law from the class of signals with bounded amplitude and rate. To solve this problem, the following assumptions are adopted.

Assumption 1. The center of the set $\mathcal{Y}$ meets the following condition for all $t \geq t_0$:

$$g(t) \in \mathcal{G} := \left\{g(t) \in \mathbb{R}\colon |\dot{g}(t)| \leq g_1^{\Delta},\ |\ddot{g}(t)| \leq g_2^{\Delta}\right\},$$

where $g(t) = \frac{\overline{g}(t) + \underline{g}(t)}{2}$ is a measurable signal (the center of the set $\mathcal{Y}$), $g_1^{\Delta} > 0,\ g_2^{\Delta} > 0$ are known scalars.

Assumption 2. There exist known scalars $\gamma_u \in (0,\,1)$, $\gamma_v \in (0,\,1)$ such that for all $g(t) \in \mathcal{G}$, $u(t) \in \mathcal{U}$, $y(t) \in Y$, $\mathcal{Y} \subset \mathcal{Y}$ and all $t \geq t_0$ it holds that

$$F(t) \in \mathcal{F} := \left\{F(t) \in \mathbb{R}\colon |F(t)| \leq \gamma_u,\, \left|\dot{F}(t)\right| \leq \gamma_v \rho\right\},$$

where $F(t) := \frac{f(t,\,y,\,u)-\dot{g}(t)}{bu^{\mathrm{U}}}$, $\rho = \frac{v^{\mathrm{U}}}{u^{\mathrm{U}}}$.

Assumption 3. $y(t_0) \in \mathcal{Y}$, where $y(t)$ denotes a solution of (3.1).

Discussing the problem under consideration, first of all, we would like to note that knowledge of the gain b is not a restrictive requirement. Indeed, using decomposition

$$\beta(t,\,y,\,u)\,u(t) = bu(t) + \Delta_b(t,\,y,\,u)\,u(t)$$

in case the following conditions are met

$$\begin{gathered} |\Delta_b(t,y,u)| < b, \\ \frac{1}{b}\left|\frac{d\Delta_b(t,\,y,\,u)}{dt}u(t) + \Delta_b(t,y,u)\,\dot{u}(t)\right| < v^{\mathrm{U}} \end{gathered} \tag{3.4}$$

for all $t \geq t_0$ and $u(t) \in \mathcal{U}$, $g(t) \in \mathcal{G}$, the summand $\Delta_b(t,\,y,\,u)\,u(t)$, caused by the uncertainty of the true gain $\beta(t,\,y,\,u)$, can be included into the external perturbation $f(t,\,y,\,u)$. If inequalities (3.4) cannot be satisfied, it may mean that the system gain changes its sign (loss of controllability) and definitely means that Assumption 2 is not met. In general, Assumption 2 implies that the augmented perturbation (the combination of external and parametric disturbances), divided by the available control effort bu^{U}, belongs to the same class of signals as the available control signal over the set $\mathcal{Y}$.

Assumptions 1 and 2 are interrelated, and therefore the scalars $g_1^{\Delta} > 0$ and $g_2^{\Delta} > 0$ from the first assumption cannot be chosen arbitrarily and must satisfy the constraints

$$\frac{g_1^{\Delta}}{bu^{\mathrm{U}}} < \gamma_u,\ \frac{g_2^{\Delta}}{bv^{\mathrm{U}}} < \gamma_v, \tag{3.5}$$

i.e., the available control effort amplitude (bu^{U}) and rate (bv^{U}) must exceed the velocity and acceleration of the center of the set $\mathcal{Y}$. For given $y(t_0)$, b, u^{U}, v^{U}, inequalities (3.5) and Assumption 3 can always be satisfied by a special choice of the set $\mathcal{Y}$.

3.1. Comments

At first glance, the stated control problem for the first-order system (3.1) seems to be non-substantive — in practice, there exist not so many systems described by first-order differential equations. However, it should be kept in mind that, by means of special transformations and changes of variables, substantive control problems for various complex high-order systems can be reduced to the considered control problem for a first-order system. Further, we consider some examples of such parametrizations.

Example 1. Consider a linear system

$$\begin{aligned} &\dot{x}(t) = Ax(t) + B\Big(u(t) + d(t)\Big),\ x(t_0) = x_0, \\ &y(t) = Cx(t), \end{aligned} \tag{3.6}$$

where $x(t) \in \mathbb{R}^n$ stands for a system state, $y(t) \in \mathbb{R}$ denotes a measurable output, $u(t) \in \mathbb{R}$ is a control signal, $d(t) \in \mathbb{R}$ denotes an unmeasurable perturbation, $A \in \mathbb{R}^{n\times n}$ is a state matrix, $B \in \mathbb{R}^{n\times 1}$

stands for an input vector, $C \in \mathbb{R}^{1\times n}$ denotes an output vector, the pairs (A, B) and (A, C) are completely observable and controllable, respectively. The matrices A, B, C are unknown, but we know the relative degree of the system, which is arbitrary but time-invariant, *i.e.*, $1 \le r \le n$, where the scalar r is defined by $CB = 0$, $CAB = 0, \ldots, CA^{r-1}B \neq 0$.

According to the analysis from [16, p. 21]; [17], under the adopted assumptions, there exists an unknown but non-singular change of variables:

$$\begin{bmatrix}\eta(t)\\ \xi(t)\end{bmatrix} = \begin{bmatrix}T_0\\ T_1\end{bmatrix} x(t) = \begin{bmatrix}T_0\\ C\\ CA\\ \vdots\\ CA^{r-1}\end{bmatrix} x(t),$$

which transforms (3.6) into:

$$\begin{aligned}
&\dot\eta(t) = A_{00}\eta(t) + A_{01}\xi(t),\\
&\dot\xi_1(t) = \xi_2(t),\\
&\dot\xi_j(t) = \xi_{j+1}(t),\ j = 2,\ldots,r-1,\\
&\dot\xi_r(t) = \beta\Big[u(t) + d(t) + w(t,\xi)\Big],\\
&y(t) = \xi_1(t),
\end{aligned} \tag{3.7}$$

where $\begin{bmatrix}M_0 & M_1\end{bmatrix}\begin{bmatrix}T_0\\ T_1\end{bmatrix} = I$, $T_0B = 0$ and

$$\begin{aligned}
&A_{00} = T_0AM_0,\ A_{01} = T_0AM_1,\\
&A_{10} = CA^rM_0,\ A_{11} = CA^rM_1,\ \beta = CA^{r-1}B,\\
&w(t,\xi) = \beta^{-1}\Big(A_{10}\eta(t) + A_{11}\xi(t)\Big).
\end{aligned}$$

If zeros of the transfer function from $u(t) + d(t)$ to $y(t)$ lie in the left half-plane, then the matrix A_{00} is Hurwitz [16, 17] and, consequently, for all trajectories $\xi(t)$, the signal $\beta^{-1}A_{10}\eta(t)$ can be treated as a bounded external disturbance.

If additionally $r = 1$, then $\xi(t) := y(t)$ and we immediately have from (3.7) that:

$$\dot y(t) = bu(t) + f(t, y, u),$$

where

$$f(t, y, u) := \beta d(t) + A_{10}\eta(t) + A_{11}y(t) + (\beta - b)u(t),$$

and the signal $\beta d(t) + A_{10}\eta(t)$ is identified with the component $f(t, \cdot, \cdot)$.

Therefore, for $r = 1$, the problem of keeping the output of system (3.6) within the given set (3.2) under the constraints (3.3) is equivalent to the stated one-dimensional problem.

If $n \ge r > 1$, then we shall assume that by means of differentiators (derivative observers) that are well studied in control theory [14, 18, 19], for all $t \ge t_0$ the estimates $\hat\xi(t)$ of the states $\xi(t)$ are available for measurement. In this case, we also assume that the following holds for the estimation error $\tilde\xi(t) = \hat\xi(t) - \xi(t)$ and its derivative:

$$\left|\tilde\xi_i(t)\right| \le \delta_1,\ \left|\dot{\tilde\xi}_i(t)\right| \le \delta_2,\ i = 1,\ldots,r,$$

where $\delta_1 > 0,\ \delta_2 > 0$.

These requirements are not restrictive, since the existing derivative observers guarantee their satisfaction [14].

Application of the linear operator to $\hat{\xi}_1(t)$ yields:

$$\lambda^{r-1}s(t) = (p+\lambda)^{r-1}\left[\hat{\xi}_1(t)\right] = (p+\lambda)^{r-1}\left[y(t)\right] + (p+\lambda)^{r-1}\left[\tilde{\xi}_1(t)\right],$$

where $p := \frac{d}{dt}[.]$, and $s(t)$ is measurable, as it can be calculated using r derivatives of the system output.

Differentiating $s(t)$ with respect to time, we obtain:

$$\begin{aligned}\dot{s}(t) &= \lambda^{1-r}\left[\dot{\xi}_r(t) + \epsilon(\xi_1,\ldots,\xi_{r-1}) + (p+\lambda)^{r-1}\left[\dot{\tilde{\xi}}_1(t)\right]\right]\\ &= \lambda^{1-r}\left[\beta\left[u(t)+d(t)+w(t,\xi)\right] + \epsilon(t) + (p+\lambda)^{r-1}\left[\dot{\tilde{\xi}}_1(t)\right]\right]\\ &= bu(t) + f(t,s,u),\end{aligned} \tag{3.8}$$

where

$$\begin{aligned}f(t,s,u) &:= \left(\lambda^{1-r}\beta - b\right)u(t)\\ &+ \lambda^{1-r}\left(\beta d(t) + A_{10}\eta(t) + A_{11}\xi(t) + \epsilon(t) + (p+\lambda)^{r-1}p\left[\tilde{\xi}_1(t)\right]\right)\end{aligned}$$

and the term $\epsilon(\xi_1,\ \ldots,\xi_{r-1}) := \epsilon(t)$ contains all other derivatives.

As

$$y(t) = \frac{\lambda^{r-1}}{(p+\lambda)^{r-1}}\left[s(t)\right] - \tilde{\xi}_1(t),$$

then for a sufficiently large value of the parameter $\lambda > 0$ and a sufficiently small upper bound δ_1, owing to the monotonic dependence of $y(t)$ on $s(t)$, we obtain that the problem of keeping the output of system (3.6) within the given set (3.2) under constraints (3.3) for the case $n \geq r > 1$ is reduced to the stated one-dimensional problem.

By complete analogy, the control problem for a class of nonlinear minimum-phase systems with arbitrary relative degree can be reduced to the control problem of a first-order system [16, p. 167].

Example 2. Consider a cascade system:

$$\begin{aligned}&\dot{x}_1(t) = b_1x_2(t) + f_1(t,x,u),\\ &\dot{x}_j(t) = b_jx_{j+1}(t) + f_j(t,x,u), \quad j = 2,\ldots,n-1,\\ &\dot{x}_n(t) = b_nu(t) + f_n(t,x,u),\\ &y(t) = x_1(t),\end{aligned} \tag{3.9}$$

where $x(t) \in \mathbb{R}^n$ is a measurable state, $f_i(t,x,u)$ stands for an unknown perturbations, $b_i > 0$ denotes known estimates of the system gains.

It is required to keep the output of system (3.9) within the given set (3.2) under the constraints (3.3) and, additionally, ensure boundedness of all elements of the state vector for all $t \geq t_0$:

$$y(t) \in \mathcal{Y} \subset \mathcal{X}_1,\ x_i(t) \in \mathcal{X}_i,\ i = 2,\ldots,n,$$

where $\mathcal{X}_i := \left\{x_i(t) \in \mathbb{R} : |x_i(t)| \leq x_i^{\mathrm{U}}\right\}$.

The following change of variables

$$\begin{aligned}&s_1(t) = x_1(t) - g(t),\\ &s_i(t) = x_i(t) - u_{i-1}(t),\ i = 2,\ldots,n,\end{aligned}$$

allows one to rewrite (3.9) as:

$$\begin{aligned}
\dot{s}_1(t) &= b_1\Big(u_1(t) + s_2(t)\Big) + f_1(t, x, u_n) - \dot{g}(t),\\
\dot{s}_j(t) &= b_j\Big(u_j(t) + s_{j+1}(t)\Big) - \dot{u}_{j-1}(t) + f_j(t, x, u_n),\ j = 2, \ldots, n-1,\\
\dot{s}_n(t) &= b_n u_n(t) - \dot{u}_{n-1}(t) + f_n(t, x, u_n),
\end{aligned} \tag{3.10}$$

where $u_i(t)$ are virtual control signals, and, for generality, we adopt $u_n(t) := u(t)$.

Thus, the problem of keeping the output of system (3.9) within the given set (3.2) is reduced to solving one-dimensional and similar problems of keeping the states $s_i(t)$ within the given sets:

$$\begin{aligned}
\mathcal{S}_1 &:= \{s_1(t) \in \mathbb{R}\colon |s_1(t)| \le \varepsilon\},\\
\mathcal{S}_i &:= \Big\{s_i(t) \in \mathbb{R}\colon |s_i(t)| \le x_i^{\mathrm{U}} - u_{i-1}^{\mathrm{U}}\Big\},\ i = 2, \ldots, n
\end{aligned}$$

for all $t \ge t_0$ under conditions of boundedness of the amplitude and rate of a real control signal (3.3) and amplitude of the virtual ones:

$$U_i := \Big\{u_i(t) \in \mathbb{R} : |u_i(t)| \le u_i^{\mathrm{U}}\Big\},\ i = 1, \ldots, n-1,$$

where $u_i^{\mathrm{U}} < x_{i+1}^{\mathrm{U}}$ is a known scalar.

Moreover, if the amplitude constraint for the virtual control signals is augmented with the rate ones (where $v_i^{\mathrm{U}} > 0,\ i = 1, \ldots, n-1$, is a known scalar)

$$\mathcal{U}_i\colon = \Big\{u_i(t) \in \mathbb{R}\colon |u_i(t)| \le u_i^{\mathrm{U}},\ |\dot{u}_i(t)| \le v_i^{\mathrm{U}}\Big\},$$

then, for all $k = 2, \ldots, n$ in each kth equation of the system (3.10), the term $\dot{u}_{k-1}(t)$ is an external disturbance bounded by v_{k-1}^{U}. Consequently, solving the stabilization problems iteratively from the bottom to up (in the direction from $s_n(t)$ to $s_1(t)$) with the help of the control signals $u_i(t) \in \mathcal{U}_i$ for all $i = 1, \ldots, n-1$, we have that $s_{i+1}(t)$ in each ith equation is bounded by $\mathcal{S}_{i+1}$, and hence, at each step of the procedure, we have exactly the one-dimensional problem stated in this paper of keeping the output of system (3.1) within the given set (3.2) under the constraints on the amplitude and rate of the control signal.

Both considered examples demonstrate that sufficiently general and substantive control problems, which are often faced in practice, are reduced to solving the one-dimensional task considered in this paper. At the end of this subsection, it should be noted that the problems from Examples 1 and 2 are not solved in this study, but only given to illustrate the applicability of the main result of this paper for their solution. In other words, the results of this study are foundational on the way to solve the problems such as those considered in Examples 1 and 2. It is worth noting that, in fact, solving problem (3.2), (3.3) for system (3.1) means solving the same problem for the class of nonlinear systems with unknown parameters, relative degree one, and stable internal dynamics (such systems, according to [16, p. 167] and Example 1, are reduced to the form (3.1), and for them, under certain conditions, Assumption 2 will be satisfied). Defining the conditions on the right-hand side of the system equations, under which Assumption 2 holds, is also beyond the scope of this paper.

4. MAIN RESULT

In Section 4.1, the structure of the control law is derived, and the problem of choosing its parameters and achievement of the goal (3.2) under constraints (3.3) is reduced to finding a solution of a system of linear inequalities. In Section 4.2, the engineering interpretation of the obtained control law is discussed, and the MatLab code to solve the inequalities from Section 4.1 is provided.

4.1. Robust Prescribed Performance Control under Given Constraints

The problem of keeping the output $y(t)$ within the given set $\mathcal{Y}$ is weaker than the problem of tracking its center $g(t)$ with a dynamic error whose absolute value does not exceed the minimum width of the set, *i.e.*,

$$\text{IF } |y(t_0)-g(t_0)|\le\varepsilon \Rightarrow \forall t\ge t_0,\ |y(t)-g(t)|\le\varepsilon, \text{ THEN } y(t)\in\mathcal{Y},\ \forall t\ge t_0, \tag{4.1}$$

where $\varepsilon=\min\left\{\inf\limits_t|\overline{g}(t)-g(t)|,\ \inf\limits_t\left|\underline{g}(t)-g(t)\right|\right\}$ is a minimal width of $\mathcal{Y}$.

In other words, meeting the condition $|y(t)-g(t)|\le\varepsilon$ for all $t\ge t_0$ will immediately imply satisfaction of the prescribed transient performance (3.2), but not vice versa. Therefore, it is proposed to consider the prescribed performance control problem as a problem of tracking the set $\mathcal{Y}$ center with a given value of the dynamic error.

In order to solve such a problem under the constraints (3.3), the control law is portioned in the following way:

$$u(t)=(1-\gamma_u)\,u^{\mathrm{U}}u_s(t)+\gamma_u u^{\mathrm{U}}u_c(t), \tag{4.2}$$

where $u_s(t)$, $u_c(t)$ are stabilizing and compensation summands.

Then, owing to the following equalities

$$|u(t)|=\left|(1-\gamma_u)\,u^{\mathrm{U}}u_s(t)+\gamma_u u^{\mathrm{U}}u_c(t)\right|\le u^{\mathrm{U}},$$
$$|\dot u(t)|=\left|(1-\gamma_u)\,u^{\mathrm{U}}\dot u_s(t)+\gamma_u u^{\mathrm{U}}\dot u_c(t)\right|\le v^{\mathrm{U}},$$

the constraints (3.3) are weaker then

$$(1-\gamma_u)\,|u_s(t)|+\gamma_u\,|u_c(t)|\le 1, \tag{4.3a}$$
$$(1-\gamma_u)\,|\dot u_s(t)|+\gamma_u\,|\dot u_c(t)|\le \rho. \tag{4.3b}$$

Following decomposition (4.2), the problem to meet the control signal constraints (3.3) is reduced to satisfying inequalities (4.3a) and (4.3b) for all $t\ge t_0$.

To solve the tracking problem (4.1) under constraints (4.3a) and (4.3b), we define the control components as follows:

$$\begin{aligned}
\dot u_s(t)&=-\frac{(1-\mu\kappa\gamma_v)\,\rho}{1-\gamma_u}\sigma\left(\frac{1}{\lambda_1}\Big(u_s(t)+\sigma\big(\frac{K}{1-\gamma_u}z\big)\Big)\right),\quad |u_s(t_0)|\le 1,\\
\dot u_c(t)&=-\frac{\mu\kappa\gamma_v\rho}{\gamma_u}\sigma\left(\frac{1}{\lambda_2}\Big(u_c(t)+\sigma\big(\frac{1}{\gamma_u}\Phi\big)\Big)\right),\quad |u_c(t_0)|\le 1,
\end{aligned} \tag{4.4}$$

where $1<\mu<\frac{1}{\kappa\gamma_v}$, $\lambda_1>0$, $\lambda_2>0$, $K>0$, $z(t)=\frac{y(t)-g(t)}{bu^{\mathrm{U}}}$, and we will define $\Phi(t)$ further.

If $\Phi(t):=F(t)$, then $-\gamma_u u_c(t)$ is an augmented perturbation with the amplitude and rate bounded by given values. As the augmented perturbation $F(t)$ is unmeasurable, then $\Phi(t)$ is proposed to be obtained with the help of A.M. Tsykunov's auxiliary loop method [15, p. 201]:

$$\Phi(t):=\chi(t)-u_f(t), \tag{4.5}$$

where $l>0$ and

$$\chi(t)=l\Big(z(t)-z_f(t)\Big),$$
$$\dot z_f(t)=\chi(t),\quad z_f(t_0)=z(t_0),$$
$$\dot u_f(t)=l\left(\frac{1}{u^{\mathrm{U}}}u(t)-u_f(t)\right),\quad u_f(t_0)=u_{0f},$$

and then $-\gamma_u u_c(t)$ is an estimate of the augmented perturbation with the amplitude and rate bounded by given values.

It should be noted that, owing to (4.4), the absolute value of the rates of $u_s(t)$, $u_c(t)$ are bounded for all $t \geq t_0$ by given positive scalars $\frac{(1-\mu\kappa\gamma_v)\rho}{1-\gamma_u}$ and $\frac{\mu\kappa\gamma_v\rho}{\gamma_u}$, and therefore, inequality (4.3b) is an identity:

$$(1-\mu\kappa\gamma_v)\rho + \gamma_u \frac{1}{\gamma_u}\kappa\mu\gamma_v\rho \leq \rho. \tag{4.6a}$$

On the other hand, if for all $t \geq t_0$ we have that $|u_s(t)| \leq 1$, $|u_c(t)| \leq 1$, then inequality (4.3a) is also an identity:

$$1 - \gamma_u + \gamma_u \leq 1. \tag{4.6b}$$

Thus, the problem to meet the constraints (3.3) for the amplitude and rate of the control signal is reduced to satisfaction of the conditions $|u_s(t)| \leq 1$, $|u_c(t)| \leq 1$ for all $t \geq t_0$. The conditions, under which such requirements hold and constraints (3.3) are satisfied, are given in the following theorem.

Theorem 1. *Let*

$$\gamma_u \in (0,\ 1),\ \gamma_v \in (0,\ 1),\ \kappa \in (1,\ 2),\ \rho > 0,\ 1 < \mu < \tfrac{1}{\kappa\gamma_v},$$
$$\lambda_1 > 0,\ \lambda_2 > 0,\ K > 0,\ l > 0,$$

then the control law (4.4), (4.5) *with the initial conditions*

$$|u_s(t_0)| \leq 1,\ |u_c(t_0)| \leq 1 \tag{4.7}$$

ensures $u(t) \in \mathcal{U}$ *for all* $t \geq t_0$.

Proof of Theorem 1 is given in Appendix.

Since $u_s(t)$, $u_c(t)$ are internal states of the feedback law, it is always possible to set $u_s(t_0)$, $u_c(t_0)$ such that the constraint $u(t) \in \mathcal{U}$ is satisfied for all $t \geq t_0$. Despite the parameters γ_u, γ_v are used in the control law, the existence of bounds for the disturbance from Assumption 2 is not required for the proof of Theorem 1, and therefore boundedness of the amplitude and rate of the control signal is ensured independently of the satisfaction of Assumption 2 and, moreover, leads to the fulfillment of all its prerequisites. This result allows one to obtain conditions, under which the chosen control law provides a solution to the prescribed performance control problem (3.2) under the given constraints for the amplitude and rate of the control signal (3.3).

Theorem 2. *Let Assumptions 1–3 be met, under given parameters*

$$\gamma_u \in (0,\ 1),\ \gamma_v \in (0,\ 1),\ bu^{\mathrm{U}} > 0,\ \delta \in (0,\ 1),\ \kappa \in (1,\ 2),\ L > \kappa,\ \rho > 0,\ \varepsilon > 0 \tag{4.8a}$$

and some scalar

$$1 < \mu^* < \tfrac{1}{\kappa\gamma_v} \tag{4.8b}$$

the following linear inequalities with respect to $\Psi \in \mathbb{R}^5$ *be feasible:*

$$\begin{aligned} &e_i^\top \Gamma \Psi < e_i^\top \Theta,\ i = 1, \ldots, 4, \\ &e_j^\top \Psi > 0,\ j = 1, \ldots, 5, \end{aligned} \tag{4.9}$$

where $e_i \in \mathbb{R}^4$, $e_j \in \mathbb{R}^5$ *are coordinate vectors of the respective spaces,*

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 & 1-\mu^* & 1 \\ 0 & 1-\gamma_u & \gamma_u & L\gamma_v\rho & L\gamma_v\rho \\ -\frac{(1-\mu^*\kappa\gamma_v)\rho}{\kappa} & 1-\gamma_u & \gamma_u & L\gamma_v\rho & L\gamma_v\rho \\ (1-\gamma_u)\, bu^{\mathrm{U}} & 0 & 0 & 0 & 0 \end{bmatrix}, \Theta = \begin{bmatrix} 0 \\ 1-(1+\delta)\gamma_u \\ -1+(1-\delta)\gamma_u \\ \varepsilon \end{bmatrix}.$$

Then the control law (4.4), (4.5) *with the parameters*

$$K = \hat{\Psi}_1^{-1},\ \lambda_1 = \hat{\Psi}_2,\ \lambda_2 = \hat{\Psi}_3,\ l = \hat{\Psi}_4^{-1},\ \tau = \hat{\Psi}_5,\ \mu = \mu^*, \tag{4.10}$$

initial conditions (4.7) *and*

$$|z(t_0)| \le \frac{1-\gamma_u}{K}, \quad \left|u_s(t_0) + \sigma\left(\frac{K}{1-\gamma_u} z(t_0)\right)\right| \le \lambda_1, \tag{4.11a}$$

$$|-u_f(t_0) - F(t_0)| \le \tau\gamma_v\rho, \quad \left|u_c(t_0) - \sigma\left(\frac{1}{\gamma_u} u_f(t_0)\right)\right| \le \lambda_2 \tag{4.11b}$$

ensures $u(t) \in \mathcal{U}$ *and* $y(t) \in \mathcal{Y}$ *for all* $t \ge t_0$.

Proof of Theorem 2 is postponed to Appendix.

Theorem 2 shows that the control law parameters calculation with the help of (4.10) under conditions (4.8a), (4.8b) and (4.9) guarantees achievement of the goal (3.2) under the given constraints (3.3) for the amplitude and rate of the control signal. However, Theorem 2 does not address feasibility of the linear inequalities (4.9) for all values of parameters (4.8a) and scalar (4.8b). The following lemma states the conditions, under which inequalities (4.9) are feasible.

Lemma 1. *Let*

$$\gamma_u \in \left(0,\ \frac{1}{1+(1+\alpha)\,\delta}\right), \quad \alpha \in (0,\ 1), \quad \gamma_v \in (0\ \ 1),$$
$$bu^{\mathrm{U}} > 0, \quad \delta \in (0,\ 1), \quad \kappa \in (1,\ 2), \quad L > \kappa, \quad \rho > 0$$

and

$$1 < \mu^* < \frac{1}{\kappa\gamma_v}.$$

Then inequalities (4.9) *are feasible wih respect to* $\Psi \in \mathbb{R}^5$ *for any* $\varepsilon > \underline{\varepsilon} > 0$.

Proof of Lemma 1 and equation for $\underline{\varepsilon}$ are presented in Appendix.

Thus, in accordance with Lemma 1, there exists ε, for which inequalities (4.9) are feasible. However, according to Lemma 1, in order to ensure existence of inequalities (4.9) solution, it is required that $\gamma_u \in \left(0,\ \frac{1}{1+(1+\alpha)\delta}\right)$, $\alpha \in (0,\ 1)$. This reduced range of admissible values of γ_u does not coincide with the one given in Assumption 2. However, this circumstance is not a serious obstacle to achieve the goal (3.2) under constraints (3.3), since the parameter δ is a parameter of the function $\sigma(.)$ from the condition e). For the following differentiable function satisfying conditions a)–f)

$$\sigma(s) = \operatorname{sign}\{s\}\left(1 - (1 - \min\{|s|, 1\})^{1.0178}\right) \tag{4.12}$$

it holds that $\delta = 6.5 \cdot 10^{-3}$, and consequently, $\frac{1}{1+(1+\alpha)\delta} \le 0.9872$, and the range from Assumption 2 coincides with sufficient accuracy with the interval, for which inequalities (4.9) are feasible.

The conditions of Lemma 1 are only sufficient, and their violation does not imply that (4.9) is infeasible, *i.e.*, Lemma 1 conditions violation does not mean that the proposed control law cannot ensure achievement of the goal (3.2) under the given constraints (3.3) and the required $\varepsilon < \underline{\varepsilon}$. However, the result of Lemma 2 is important from a theoretical point of view, since it allows one to conclude that it is fundamentally possible to solve (4.9) for a set of given width.

Inequalities (4.9) can be solved using various algorithms and software. For example, one can apply algorithms to solve linear (linprog function in MatLab) or semidefinite programming problems (cvx, SEDUMI, YALMIP in MatLab).

According to Theorem 2, any solution of the inequalities allows one to obtain a control law that ensures the prescribed transient performance for the output (3.2) under the given control constraints (3.3). Therefore, the question of how to choose the most suitable solver or obtain the optimal solution of the considered inequalities is beyond the scope of the theoretical problem statement of this paper and requires its refinement or extension by some optimality criterion. From practical point of view, the solutions obtained via different numerical algorithms are identical with respect to the performance requirements (3.2) and constraints (3.3), but may differ, for example, in terms of the oscillations of the system output within the set $\mathcal{Y}$ or the boundedness of the control signal acceleration. In the author's opinion, the most successful solutions of inequalities (4.9) with respect to the mentioned criteria can be obtained using the interior-point method (feasp function in MatLab).

The program code below, written in the MatLab language using the feasp function, implements the solution of inequalities (4.9) for given parameters (4.8a) and a chosen scalar (4.8b). The result of the code execution is the control law parameters.

```
% SYSTEM PARAMETERS
rho=1.5; g_u=0.98; g_v=0.45; bU=0.8*(1.4522e+04)*1/3;
kappa=1.04; delta=6.5*10^-3; L=1.1; mu=1.01; eps=5;
% INEQUALITIES INITIALIZATION AND SOLUTION
Gamma=[0 0 0 1-mu 1; 0 1-g_u g_u L*g_v*rho L*g_v*rho;
-(1-mu*kappa*g_v)*rho/kappa 1-g_u g_u L*g_v*rho L*g_v*rho;
(1-g_u)*bU 0 0 0 0];
Theta=[0; 1-(1+delta)*g_u;-1+(1-delta)*g_u; eps];
e1=[1 0 0 0]; e2=[0 1 0 0]; e3=[0 0 1 0]; e4=[0 0 0 1];
setlmis([]); Psi=lmivar(2,[5 1]);
lmiterm([-1 1 1 Psi],.5*[1,0,0,0,0],1,'s');
lmiterm([-2 1 1 Psi],.5*[0,1,0,0,0],1,'s');
lmiterm([-3 1 1 Psi],.5*[0,0,1,0,0],1,'s');
lmiterm([-4 1 1 Psi],.5*[0,0,0,1,0],1,'s');
lmiterm([-5 1 1 Psi],.5*[0,0,0,0,1],1,'s');
lmiterm([6 1 1 Psi],.5*e1*Gamma,1,'s');
lmiterm([7 1 1 Psi],.5*e2*Gamma,1,'s');
lmiterm([8 1 1 Psi],.5*e3*Gamma,1,'s');
lmiterm([9 1 1 Psi],.5*e4*Gamma,1,'s');
lmiterm([-6 1 1 0],e1*Theta); lmiterm([-7 1 1 0],e2*Theta);
lmiterm([-8 1 1 0],e3*Theta); lmiterm([-9 1 1 0],e4*Theta);
LM1=getlmis;[tmin,xfeas] = feasp(LM1, [0 1000 1e22 0 0]);
% if tmin ¡ 0, then the inequalities (4.9) is feasible
PS1 = [1 0 0 0 0]*dec2mat(LM1,xfeas,Psi);
PS2 = [0 1 0 0 0]*dec2mat(LM1,xfeas,Psi);
PS3 = [0 0 1 0 0]*dec2mat(LM1,xfeas,Psi);
PS4 = [0 0 0 1 0]*dec2mat(LM1,xfeas,Psi);
PS5 = [0 0 0 0 1]*dec2mat(LM1,xfeas,Psi);
K=PS1^-1; lambda1=PS2; lambda2=PS3; l=PS4^-1; mu=mu; tau=PS5;
```

Remark 1. By choosing the initial condition $y(t_0)$ inside a sufficiently small neighborhood of the center of the set $\mathcal{Y}$, the inequality $|z(t_0)| \le \frac{1-\gamma_u}{K}$ can be satisfied. Conditions (4.11b) are not restrictive for practical scenarios, since usually, at the initial time instant t_0, we have $y(t_0) = 0$, $u(t_0) = 0$, $\dot{g}(t_0) = 0$ and the disturbance $f(t, y, u)$ satisfies $f(t_0, 0, 0) = 0$. Therefore, to satisfy conditions (4.11b), it is usually sufficient to choose $u_f(t_0) = u_c(t_0) = 0$. If this is not the case,

then the following simple technique is proposed. Since $\Phi(t)$ is an estimate of $F(t)$, one can obtain the estimate $\Phi(t)$ over $t < t_0$ and, at the moment of the control system initialization t_0, set $u_f(t_0) = -\Phi(t_0)$. If, for $t < t_0$, a sufficiently large parameter $l > 0$ is used in (4.5), then the equality $u_f(t_0) = -\Phi(t_0)$ will be satisfied with any prescribed accuracy [15]. In other words, if nothing is known *a priori* about whether conditions (4.11b) are satisfied, then it makes sense to first run the disturbance observer (4.5), and only then initialize the control system (4.4) taking into account the disturbance estimate, which has already been obtained by the time t_0.

Remark 2. In accordance with the proof of Theorem 2, if the constraints for the initial conditions (4.11a), (4.11b) are not satisfied, but inequalities (4.7), (4.9) hold, then $u(t) \in \mathcal{U}$ for all $t \geq t_0$, and for all $y(t_0) \in Y$ there exists a time instant $T > t_0$ such that $y(t) \in \mathcal{Y}$ for all $t \geq T$.

Remark 3. To reduce the values of the the control law (4.4) parameters, inequalities (4.9) can be augmented with constraints of the form $\Psi_i \geq \underline{\Psi}_i > 0$, where $\underline{\Psi}_i > 0$ are given numbers.

4.1.1. Estimation of augmented perturbation initial condition. In accordance with Theorem 2, the goal (3.2) achievement essentially depends on the satisfaction of inequalities (4.11a) and (4.11b). According to the arguments presented in Remark 1, inequalities (4.11a) can always be satisfied by choosing $\overline{g}(t_0), \underline{g}(t_0), u_s(t_0)$ for a given $y(t_0)$. In this subsection, we develop the idea proposed in the same remark on how to satisfy inequalities (4.11b) via estimation of the initial condition $F(t_0)$ of the augmented disturbance. To this end, let $t_0 > 0$ be the time instant of initialization of the control law (4.4), (4.5) for system (3.1), and let $t = 0$ be the time instant when the states of system (3.1) start being measured. Then, to satisfy inequalities (4.11b), the following initialization of the control law initial conditions is proposed:

$$\begin{aligned} u_f(t_0) &= -\hat{F}(t_0), \\ u_c(t_0) &= \sigma\left(\frac{1}{\gamma_u} u_f(t_0)\right), \end{aligned} \tag{4.13}$$

where $\hat{F}(t)$ is an estimate of $F(t_0)$ obtained over $[0,\ t_0]$.

Therefore, satisfaction of (4.11b) is reduced to estimation of the augmented perturbation over $[0,\ t_0]$ in such a way that

$$\left|\hat{F}(t_0) - F(t_0)\right| \leq \tau\gamma_v\rho,\ \tau > 0. \tag{4.14}$$

To achieve this goal, we introduce a disturbance observer that is structurally identical to (4.5) but starts its operation at $t = 0$:

$$\hat{F}(t) = \hat{\chi}(t) - \hat{u}(t) - \frac{\dot{g}(t)}{bu^{\mathrm{U}}}, \tag{4.15}$$

where

$$\begin{aligned} \hat{\chi}(t) &= k\Big(z(t) - \hat{z}(t)\Big), \\ \dot{\hat{z}}(t) &= \hat{\chi}(t), \quad \hat{z}(0) = z(0), \\ \dot{\hat{u}}(t) &= k\left(\frac{1}{u^{\mathrm{U}}} u(t) - \hat{u}(t) - \frac{\dot{g}(t)}{bu^{\mathrm{U}}}\right), \quad \hat{u}(0) = 0 \end{aligned}$$

and $k > 0$, $g(t) = 0$ and $\dot{g}(t) = 0$ for all $t_0 > t \geq 0$.

Then the estimation error $\tilde{F}(t) = \hat{F}(t) - F(t)$ satisfies the following equation for all $t_0 \geq t \geq 0$:

$$\dot{\tilde{F}}(t) = k\left(\dot{z}(t) - \dot{\hat{z}}(t)\right) - \dot{\hat{u}}(t) - \frac{\ddot{g}(t)}{bu^{\mathrm{U}}} - \dot{F}(t) = -k\tilde{F}(t) - \frac{\dot{f}(t,x,u)}{bu^{\mathrm{U}}}, \tag{4.16}$$

where $\tilde{F}(0) = -F(0)$.

Owing to equation (4.16), it is easy to show the following.

Proposition 1. *Let* $\left|\frac{\dot{f}(t,\,x,\,u)}{bu^{\mathrm{U}}}\right| \leq \gamma_v\rho$ *for all* $0 \leq t \leq t_0$ *and the initial conditions for a compensation component of the control law* (4.4), (4.5) *be defined in accordance with* (4.13), (4.15). *Then, for all* $F(0)$ *and* $t_0 > 0$, *there exists* $k^* > 0$ *such that the choice*

$$k > k^* \tag{4.17}$$

ensures that inequality (4.11b) *holds.*

Proof of proposition 1 is presented in Appendix.

Thus, using the proposed initialization algorithm for the compensation component of the control law, it is always possible to satisfy inequalities (4.11b). This allows one to conclude that all requirements on the initial conditions (4.7), (4.11a), (4.11b) from Theorem 2 are not restrictive and can be satisfied by: *a*) choosing $\overline{g}(t_0), \underline{g}(t_0), u_s(t_0)$ for a given $y(t_0)$, *b*) initializing the initial conditions of the compensation component of the control law (4.4), (4.5) using the algorithm (4.13), (4.15), (4.17). It should also be noted that inequality $\left|\frac{f(t,\,x,\,u)}{bu^{\mathrm{U}}}\right| \leqslant \gamma_v\rho$ can be satisfied even for unstable systems, and hence, proposition 1 does not require the system to be stable and/or be at the equilibrium over $[0, t_0]$.

4.2. Engineering Interpretation

The control law (4.4), (4.5) is contains a proportional feedback, whose amplitude and rate are bounded at all time instants by $(1-\gamma_u)u^{\mathrm{U}}$ and $(1-\mu\kappa\gamma_v)v^{\mathrm{U}}$ and the estimate of the augmented disturbance, whose amplitude and rate are bounded at all time instants by $\gamma_u u^{\mathrm{U}}$ and $\mu\kappa\gamma_v v^{\mathrm{U}}$, respectively.

To calculate the parameters of such a control law, Theorem 2 requires to solve linear inequalities (4.9) and know the following parameters:

— the maximum value of the control effort bu^{U}, where b is a known estimate of the unknown system gain (see Section 3.1.2);
— the constraints u^{U} and v^{U} for the amplitude and rate of the control signal;
— the scalars γ_u and γ_v, which indicate how many times the amplitude and rate of the disturbance, divided by maximum control effort, are less than unity;
— the minimum width ε of the set $\mathcal{Y}$.

Considering practical scenarios, an assumption that all these parameters are known is not restrictive, since some of the parameters are chosen at the stage of the system actuator design (u^{U}, v^{U}, b) or at the specification stage of the control system development (ε). The other part, namely the scalars $\gamma_u \in (0,\,1)$, $\gamma_v \in (0,\,1)$, from theoretical point of view, can be chosen arbitrarily close to the right boundaries of their admissible intervals. However, from practical point of view and due to the conservativeness imposed by derivation of inequalities (4.9), there is no need in that, and the choice of γ_u, γ_v should be aimed at inequalities (3.5) satisfaction.

To solve inequalities (4.9), it is also necessary to know the parameters κ, L, δ from properties *a*)–*f*) of the saturation function approximation. This requirement is not restrictive, since for each specific approximation these parameters can always be calculated numerically or found analytically. Moreover, in this paper, for two functions (see Section 2 and equation (4.12)), the values of these parameters have already been given.

If inequalities (4.9) are feasible, then, according to Theorem 2, the control law (4.4), (4.5) with parameters (4.10) and initial conditions (4.7), (4.11a), (4.11b) is guaranteed to ensure the following property:

If Assumptions 1–3 are met, then for any augmented disturbances $F(t)$, *whose amplitude does not exceed* $100\gamma_u$ % *of* u^{U}, *and rate is not higher than* $100\gamma_v$ % *of* v^{U}, *it is ensured that* $y(t) \in \mathcal{Y}$, $u(t) \in \mathcal{U}$ *for all* $t \geq t_0$.

5. ANALYTIC COMPARISON

In terms of structure and properties, the control law (4.2) is most closely related to the one from [10]. In accordance with the results of [10], the following control law is introduced to ensure a given bound for the tracking error with respect to the center of the set $\mathcal{Y}$:

$$u(t) = -u^{\mathrm{U}} m \sigma(kz(t)), \tag{5.1}$$

where $m > 0,\ k > 0$ and

$$\sigma(s) := \frac{2}{1+\exp(-s)} - 1$$

is a sigmoid function (in contrast to the saturation function approximation used in this paper, for the sigmoid one, condition c) from Section 2 is not satisfied).

According to lemma and Remark 1 from [10], if there exists $\gamma_u \in (0,\ 1)$ such that for all $t \geq t_0$ it holds that $|F(t)| \leq \gamma_u$, and the following inequalities are satisfied:

$$\frac{\gamma_u}{\sigma(2.2)} < m \leq 1, \tag{5.2a}$$

$$\frac{2.2}{\Delta} \leq k < \frac{\rho}{m^2}, \quad \Delta > 0, \tag{5.2b}$$

$$bu^{\mathrm{U}}\Delta \leq \varepsilon \tag{5.2c}$$

then the control law (5.1) ensures that $y(t) \in \mathcal{Y}$ and $u(t) \in \mathcal{U}$ for all $t \geq t_0$.

First of all, it should be noted that inequalities (5.2a) and (5.2b) are feasible only if it holds that

$$\begin{aligned} v^{\mathrm{U}} &> \frac{2.2}{\Delta} m^2 u^{\mathrm{U}}, \\ \varepsilon &> 2.2 b u^{\mathrm{U}} \frac{\gamma_u^2}{\sigma^2(2.2)\rho} = \underline{\varepsilon}, \end{aligned} \tag{5.3}$$

which are rather restrictive requirements: i) not for all $v^{\mathrm{U}} > 0$ and $u^{\mathrm{U}} > 0$ it is possible to bound the control signal rate, ii) the comparison of the lower bounds (5.3) and (A.14) allows one to conclude that the proposed control law, for the same $u^{\mathrm{U}} > 0,\ v^{\mathrm{U}} > 0$ and $\gamma_u \to 1$, ensures $y(t) \in \mathcal{Y}$ for all $t \geq t_0$ for sets of substantially smaller width ε.

Secondly, the following upper bound was used to obtain (5.2b) in [10]:

$$\begin{aligned} |\dot{u}(t)| &= \left| -\frac{u^{\mathrm{U}} m k}{2} \left[1 - \sigma^2(kz(t))\right] \dot{z}(t) \right| \\ &\leq \left| -\frac{u^{\mathrm{U}} m k}{2} \left[1 - \sigma^2(kz(t))\right] \right| \left| -m\sigma^2(kz(t)) + F \right| \\ &< \left| -u^{\mathrm{U}} m^2 k \left[1 - \sigma^2(kz(t))\right] \right| < u^{\mathrm{U}} m^2 k. \end{aligned} \tag{5.4}$$

Such an estimate can only be obtained when the output signal $y(t)$ is not affected by the measurement noise, *i.e.*, when inequalities (5.2a), (5.2b), (5.2c) are met, but the output is corrupted by the measurement noise, the control law (5.1) does not guarantee boundedness of the control signal rate by a given value. As follows from (4.4), the control law proposed in this paper, owing to the nonlinear filters of the type [14], guarantees boundedness of the control signal rate regardless of the presence of measurement noise, which is a significant advantage for practical scenarios.

Therefore, the main advantages of the control law (4.4), (4.5) in comparison with (5.1) are as follows:

1) the given constraints for the amplitude and rate of the control signal are satisfied in addition to the one for the tracking error;
2) the proposed control law contains not only a component proportional to the tracking error, but also a dynamic estimate of the external augmented disturbance designed with the help of the auxiliary loop method [15, p. 201];
3) the procedure to choose the control law parameters is reduced to solving linear inequalities, which can be done numerically using software packages (the inequalities from [10] are solved manually).

6. NUMERICAL EXPERIMENTS

Simulations have been conducted in MatLab Simulink using forward Euler integration method with a fixed step of 10^{-4} seconds.

To demonstrate the theoretical results, we considered a linearized model of the longitudinal motion of an aircraft at an altitude of 1200 meters with an airspeed of Mach 2 [20, p. 565]:

$$\begin{aligned} &\dot{x}(t) = Ax(t) + Bu(t) + Dd(t),\ x(t_0) = x_0, \\ &y(t) = Cx(t), \end{aligned} \tag{6.1}$$

where

$$A = \begin{bmatrix} -1.364 & -92.82 & -128.46 \\ 1 & -4.68 & -0.087 \\ 0 & 0 & -190 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ 0 \\ 190 \end{bmatrix},\ D = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 1.36 & -184.26 & 76.43 \end{bmatrix},\ x(t_0) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix},\ d(t) = 0,$$

$u(t)$ is the elevator control signal, rad, $y(t)$ stands for the vertical acceleration, m/s^2, $x_1(t)$ denotes the pitch rate, rad/s, $x_2(t)$ is the angle of attack, rad, $x_3(t)$ denotes the elevator deflection angle, rad, $d(t)$ is an external disturbance. The maximum elevator deflection angle is $\frac{1}{3}$ rad, and its rate does not exceed $\frac{1}{2}$ rad/s, *i.e.* the control constraints are defined as $u^{\mathrm{U}} = \frac{1}{3}$, $v^{\mathrm{U}} = \frac{1}{2}$. The matrices A, B, C and D are assumed to be unknown, but we know the estimate $b = 1.16 \times 10^4$ of the gain $CB = 1.45 \times 10^4$. The relative degree of system (6.1) is equal to one, and therefore, according to Example 1, it is reducible to the form (3.1).

The prescribed transient performance for the system (6.1) is defined by the set $\mathcal{Y}$ with the following bounds:

$$\overline{g}(t) = 75\sin(0.1\pi t) + 5,\ \underline{g}(t) = 75\sin(0.1\pi t) - 5, \tag{6.2}$$

and then $\varepsilon = 5$ is the minimal width of the set $\mathcal{Y}$.

Assume that the amplitude of the augmented disturbance $F(t)$ does not exceed 98% of u^{U}, and its rate is not higher than 45% of v^{U} over the trajectories of the closed-loop system, *i.e.*,

$$\gamma_u = 0.98,\ \gamma_v = 0.45. \tag{6.3}$$

This hypothesis corresponds to Assumption 2 and cannot be verified *a priori* for the system (6.1) with unknown parameters and affected by external disturbances. The verification of this hypothesis can only done *a posteriori* by application of the control law (4.4), (4.5) and analysis of the system output behavior. From theoretical point of view, one can choose $\gamma_u \to 1$ and $\gamma_v \to 1$, then the class of admissible augmented disturbances is maximally extended. From practical point of view, there is no need for such a choice, since in many practical problems the augmented disturbance rarely exceeds $50 - 60\%$ of the maximal amplitude and rate of the control signal. Moreover, the solution of inequalities (4.9) may result in relatively large values of the control law parameters for $\gamma_u \to 1$ and $\gamma_v \to 1$.

In fact, a particular choice of γ_u and γ_v defines a class of systems with unknown parameters and external disturbances, for which the proposed control law ensures achievement of the goal (3.2) under constraints (3.3). Unfortunately, the fact that a particular system with unknown parameters and external disturbances is a member of such a class can only be verified *a posteriori* after testing the control law. For example, for system (6.1), the values from (6.3) were chosen via iterative process consisted of *i*) solution of inequalities (4.9) for various values of γ_u and γ_v, and *ii*) application of the control law (4.4), (4.5) with the calculated parameters. Such iterations were terminated, when we obtained such values of parameters (4.10) that the application of the control law ensured achievement of the stated goal. Note that the discussed *a posteriori* validation is conventional for many robust control laws. The only difference is that, usually, the ranges for the system parameters and the amplitude of external disturbances are chosen on the basis of some hypothesis, whereas in this paper, we choose the maximum values of the amplitude and rate of the augmented disturbance, which is an unknown function of the system output, its parameters and disturbances (see Example 1).

Thus, the given sets $\mathcal{Y}$ and $\mathcal{U}$, together with the chosen parameters (6.3), correspond to the following specification for the automatic control system design: *it is required to ensure tracking of the reference signal for the vertical acceleration with a dynamic error not exceeding 5 m/s^2 in the presence of matched augmented disturbances that are not higher than* 98% *of the maximum amplitude and* 45% *of the maximum rate of the control signal.*

To solve such a problem using the control law (4.4), (4.5), in accordance with Theorem 2, it is necessary to solve inequalities (4.9). To this end, we have chosen the function (4.12) and pick $\mu^* = 1.01$.

As a result, the following solution of (4.9) was obtained:

$$K = 17.7436,\ \lambda_1 = 0.1434,\ \lambda_2 = 0.0029,\ l = 194.912,\ \tau = 3.5 \times 10^{-5}. \tag{6.4}$$

The inequalities were solved in MatLab using code presented in Section 4.2. To achieve the goal (3.2) under constraints (3.3), in accordance with Theorems 1 and 2, the initial conditions of the control law were picked as follows:

$$z_f(t_0) = z(t_0) = 0,\ u_s(t_0) = 0,\ u_f(t_0) = -F(t_0) = 0,\ u_c(t_0) = 0, \tag{6.5}$$

which ensured satisfaction of the conditions (4.7) and (4.11a), (4.11b).

Figures 2 and 3 depicts the behavior of the control signal $u(t)$ and its rate $\dot{u}(t)$, the output $y(t)$, as well as the internal variables $\eta_1(t)$, $\eta_2(t)$, $\eta_3(t)$ of the control law (see the proof of Theorem 2).

Figures 2 and 3 validate theoretical results obtained in Theorem 2. The satisfaction of inequalities (4.9) and other prerequisites of Theorem 2 leads to the achievement of the stated goal (3.2) under constraints (3.3). Moreover, the internal variables of the control law also do not leave the range specified by inequalities (4.9) (see the proof of Theorem 2).

Next the proposed control law was tested for a system affected by measurement noise and discontinuous external disturbances. For such an experiment, the output of the system (6.1) and the perturbation were defined as follows:

$$\begin{gathered} y(t) = Cx(t) + w_0(t) + w_1(t), \\ d(t) = 0.1 + \operatorname{sign}\{\sin(0.4\pi t)\}, \\ w_1(t) = \begin{cases} -Cx(t) - w_0(t) + \overline{g}(t), & \text{if} \quad t \in [7,\ 7.1), \\ 0 \quad \text{otherwise}, \end{cases} \end{gathered}$$

where the signal $w_0(t)$ was implemented in the MatLab Simulink with the help of the block "Random Number" with zero mean, variance of 10^{-1} and sampling time equaled to 10^{-4}. All other parameters of the system and control law were left unchanged.

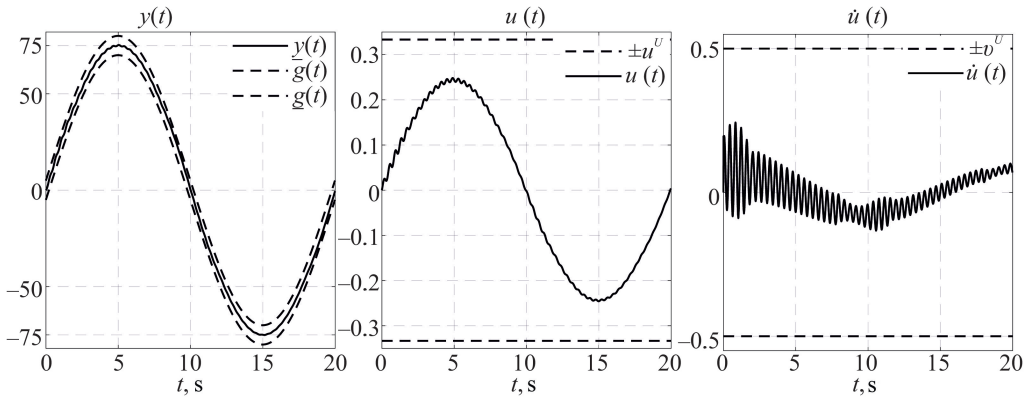


Fig. 2. Behavior of $u\,(t)$, $\dot{u}\,(t)$ and $y\,(t)$.

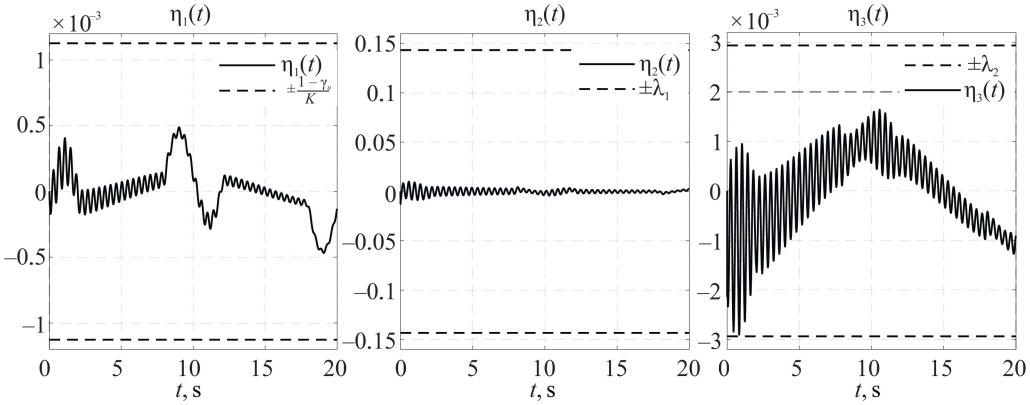


Fig. 3. Behavior of $\eta_1\,(t)$, $\eta_2\,(t)$, $\eta_3\,(t)$.

For comparison purposes, we chose the control law (5.1) with parameters $m = 1,\ k = 1692$ calculated via (5.2) with $\varepsilon = 5$, $b = 1.16 \times 10^4$ and $u^{\mathrm{U}} = \frac{1}{3}$, and the law from Section 4 of [1]:

$$u(t) = -(CB)^{-1}\Big[CAx(t) + K\varepsilon(t)\Big], \quad \varepsilon(t) := ln\left(\frac{\underline{g}(t) - y(t)}{y(t) - \overline{g}(t)}\right), \tag{6.6}$$

where $K > 0$.

The control law (6.6) requires knowledge of the system (6.1) parameters and does not take into account the constraints (3.3) for the control signal and its rate. However, despite these circumstances, its comparison with the proposed control law is admissible and correct, since the purpose of the comparison is primarily to verify the sensitivity to measurement noise and discontinuous perturbations. The conclusions drawn will remain valid both for adaptive/robust versions of the control law (6.6) and for versions that take into account the control input constraint [1, 2].

Figure 4 depicts that the control law (6.6) ensures that the system output remains in the set $\mathcal{Y}$ in the presence of measurement noise for all $t < 7$. However, when $y\,(t)$ approaches the boundaries of the set $\mathcal{Y}$, the rate of the control signal increases. As a result, a control signal that cannot be implemented by the actuator may be generated. Moreover, at the time instant $t = 7$, due to the influence of the noise component $w_1\,(t)$, a feedback discontinuity occurs (see (6.6)), and therefore, the simulation cannot be continued for all $t > 7$. Considering practical scenarios, even a potential occurrence of such discontinuities is unacceptable, since feedback discontinuity can lead to equally unpredictable and catastrophic consequences. The control law proposed in this paper is free from this drawback.

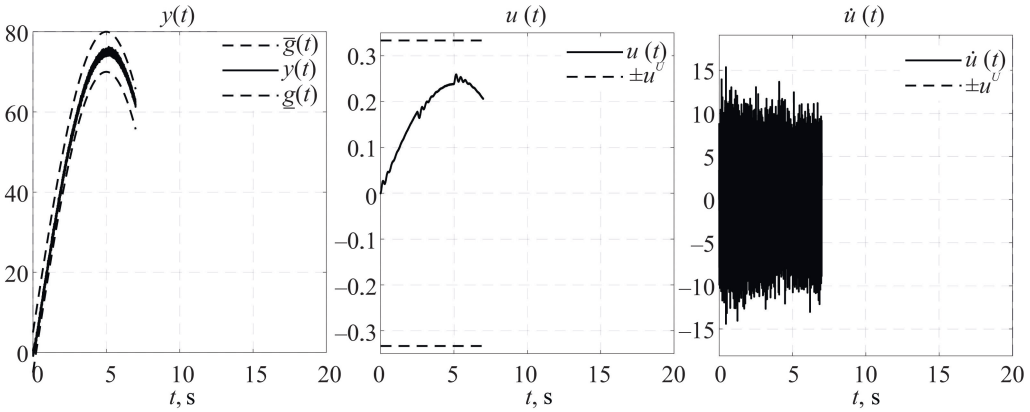


Fig. 4. Behavior of $u(t)$, $\dot{u}(t)$ and $y(t)$ for (6.6) when $K = 2.5$.

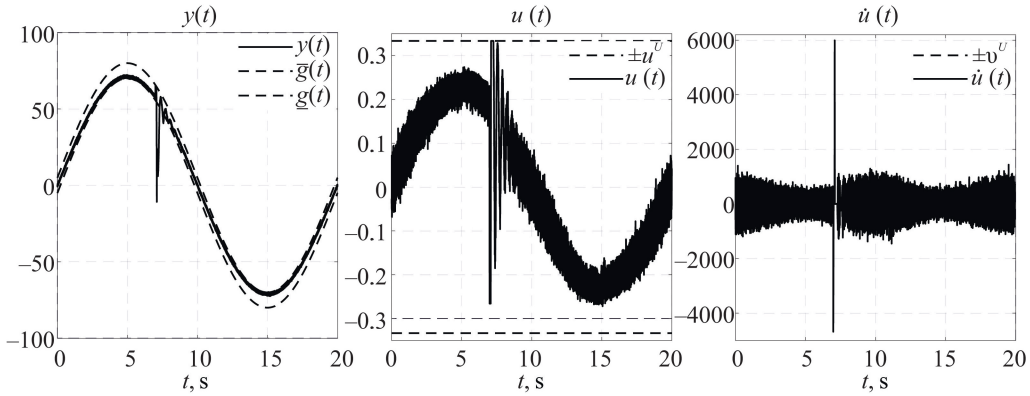


Fig. 5. Behavior of $u(t)$, $\dot{u}(t)$ and $y(t)$ for (5.1) when $m = 1,\ k = 1692$.

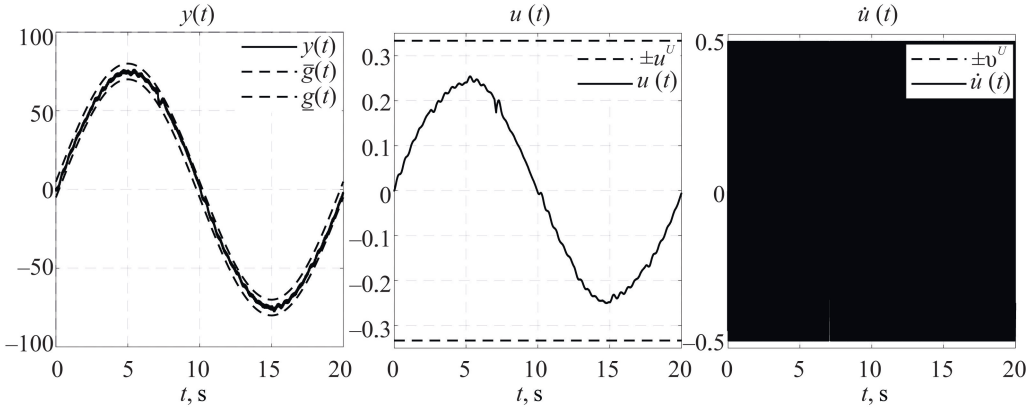


Fig. 6. Behavior of $u(t)$, $\dot{u}(t)$ and $y(t)$ for (4.4), (4.5).

In Figure 5, the behavior of $u(t)$, $\dot{u}(t)$ and $y(t)$ for (5.1) is presented.

The control law (5.1) solves the tracking problem with prescribed accuracy, which leads to keeping the system output within the given set $\mathcal{Y}$. However, the rate of the control signal is not bounded by a given value, and the control signal is significantly affected by noise, which illustrates the conclusions of the analysis from Section 5.

Figure 6 presents the behavior of $u(t)$, $\dot{u}(t)$ and $y(t)$ for the proposed control law for the system affected by discontinuous disturbances and measurement noise.

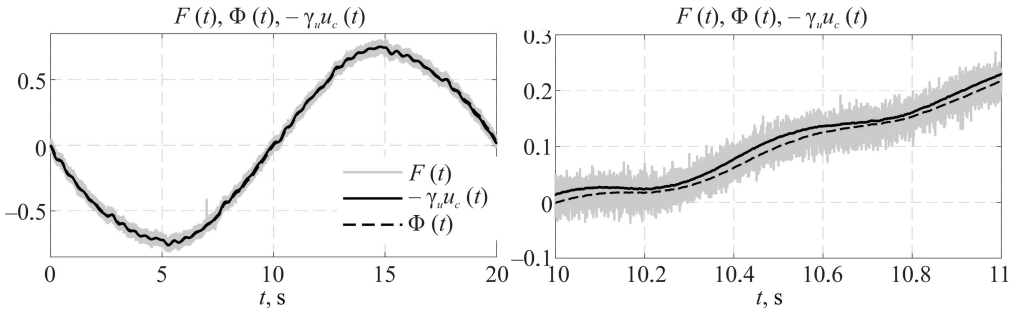


Fig. 7. Behavior of $F(t)$ and $\Phi(t)$, $-\gamma_u u_c(t)$.

According to Assumption 2, the proposed control law requires all external disturbances to be differentiable. In this simulation, this requirement is violated, but the prescribed performance (3.2) under the given constraints (3.3) is still ensured. This allows us to conclude that the conditions adopted for the proof of Theorem 2 are conservative and stimulates further theoretical research aimed at their relaxation. Moreover, in comparison with the control law (6.6), the proposed one is free from the problem of feedback discontinuities and ensures that the given constraints for the amplitude and rate of the control signal are met even in the presence of measurement noise. Comparing to the control law (5.1), the proposed one, even under such conditions, ensures boundedness of the control rate by a given value. It should be noted that the properties of the proposed control law (4.4), (4.5) are strictly justified only for the situation when we have no discontinuous external disturbances and measurement noise. Therefore, the conclusions drawn from Figures 3-5 are valid only for the considered system (6.1) under the given specific initial conditions and the integration algorithm in use.

Figure 7 depicts behavior of $F(t)$ and estimates $\Phi(t)$, $-\gamma_u u_c(t)$.

It follows from Fig. 7 that the disturbance estimate obtained by the auxiliary loop method [15, p. 201] is noisy due to the presence of noise in the measured output. In practice, using such disturbance estimates for compensation purposes can result in system instability because of *i*) the bounded control signal rate and *ii*) the excitation of system unmodeled dynamics. To reduce the influence of noise on the estimate $\Phi(t)$, the auxiliary loop method requires to decrease the parameter l. However, the developed control law employs a nonlinear filter of the type [14], which forms the estimate $\Phi(t)$ with a rate not exceeding the control signal rate constraint. This makes it possible, regardless of the values of the parameter l, to attenuate the noise influence and, as a result, satisfy the control signal rate constraint without causing excitation of the unmodeled dynamics. An analytical study of the filtering properties of the nonlinear filter used in the proposed control law is one of the promising directions of future research.

7. CONCLUSION

A method is proposed to solve a robust prescribed performance control problem under given constraints for the amplitude and rate of the control signal for a class of nonlinear minimum-phase systems with unknown parameters and relative degree equal to one (such systems, according to [16, p. 167], can be reduced to the form (3.1), and Assumption 2 can be satisfied for them). It follows from Example 1 that one of the representatives of such a class of systems is linear systems with unknown parameters and relative degree equal to one.

The scope of further theoretical research is, first, to extend the obtained results to the class of nonlinear systems with stable internal dynamics and relative degree greater than one (see Example 1), and second, to use the results of the paper to solve prescribed performance control problem for cascade systems under given constraints on the states and the amplitude/rate of the control signal (see Example 2).

APPENDIX

Proof of Theorem 1. Consider a function:

$$V = \frac{1}{2}u_s^2. \tag{A.1}$$

The derivative of (A.1) is written as

$$\dot{V} = -\frac{(1-\mu\kappa\gamma_v)\,\rho}{1-\gamma_u}u_s\sigma\left(\frac{1}{\lambda_1}\left(u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right)\right)$$
$$= -\frac{(1-\mu\kappa\gamma_v)\,\rho}{1-\gamma_u}u_s\operatorname{sign}\left\{u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right\}\left|\sigma\left(\frac{1}{\lambda_1}\left(u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right)\right)\right|.$$

For all $u_s \notin \Lambda = \{u_s \in \mathbb{R}\colon |u_s| \le 1\}$, it holds that

$$\operatorname{sign}\left\{u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right\} = \operatorname{sign}\{u_s\},$$

therefore, we have

$$\dot{V} = -\frac{(1-\mu\kappa\gamma_v)\,\rho}{1-\gamma_u}\left|u_s\sigma\left(\frac{1}{\lambda_1}\left(u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right)\right)\right| \le 0$$

for all $u_s \notin \Lambda = \{u_s \in \mathbb{R}\colon |u_s| \le 1\}$.

On the other hand, it holds for all $u_s \in \Lambda$ that

$$\dot{V} = -\frac{(1-\mu\kappa\gamma_v)\,\rho}{1-\gamma_u}u_s\sigma\left(\frac{1}{\lambda_1}\left(u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right)\right)$$
$$= -\frac{(1-\mu\kappa\gamma_v)\,\rho}{1-\gamma_u}u_s\operatorname{sign}\left\{u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right\}\left|\sigma\left(\frac{1}{\lambda_1}\left(u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right)\right)\right|$$
$$\le \frac{(1-\mu\kappa\gamma_v)\,\rho}{1-\gamma_u}\left|u_s\sigma\left(\frac{1}{\lambda_1}\left(u_s+\sigma\left(\frac{K}{1-\gamma_u}z\right)\right)\right)\right|,$$

thus, when $|u_s(t_0)| \le 1$, we have that $|u_s(t)| \le 1$ for all $t \ge t_0$.

Having considered the function $V = \frac{1}{2}u_c^2$ and repeated the above-given derivations, we obtain that, when $|u_c(t_0)| \le 1$, it holds that $|u_c(t)| \le 1$ for all $t \ge t_0$, and therefore, in accordance with the derivations (4.6a) and (4.6b), we have $u(t) \in \mathcal{U}$ for all $t \ge t_0$.

Proof of Theorem 2. Consider the following change of variables:

$$\eta_1 = z,$$
$$\eta_2 = u_s + \sigma\left(\frac{K}{1-\gamma_u}\eta_1\right),$$
$$\eta_3 = u_c + \sigma\left(\frac{1}{\gamma_u}\Phi\right)$$

and obtain the following differential equations with respect to η_1, η_2, η_3:

$$\begin{aligned}
\dot{\eta}_1 &= (1-\gamma_u)\,u_s + \gamma_u u_c + F(t)\\
&= (1-\gamma_u)\left(u_s \pm \sigma\left(\frac{K}{1-\gamma_u}\eta_1\right)\right) + \gamma_u u_c + F(t)\\
&= (1-\gamma_u)\left(-\sigma\left(\frac{K}{1-\gamma_u}\eta_1\right)+\eta_2\right) + \gamma_u u_c + F(t),\\
\dot{\eta}_2 &= -\frac{(1-\mu\kappa\gamma_v)\,\rho}{1-\gamma_u}\sigma\left(\frac{1}{\lambda_1}\eta_2\right) + \frac{K}{1-\gamma_u}\sigma'\left(\frac{K}{1-\gamma_u}\eta_1\right)\dot{\eta}_1,\\
\dot{\eta}_3 &= -\frac{\mu\kappa\gamma_v\rho}{\gamma_u}\sigma\left(\frac{1}{\lambda_2}\eta_3\right) + \sigma'\left(\frac{1}{\gamma_u}\Phi\right)\frac{1}{\gamma_u}\dot{\Phi}.
\end{aligned} \tag{A.2}$$

Consider the sets ($\eta = \begin{bmatrix}\eta_1 & \eta_2\end{bmatrix}^\top$):

$$\Omega_1 = \left\{\eta \in \mathbb{R}^2\colon\ |\eta_1| \le \frac{1-\gamma_u}{K},\quad |\eta_2| \le \lambda_1\right\},$$
$$\Omega_2 = \left\{\eta \in \mathbb{R}^2\colon\ |\eta_2| \le \lambda_1\right\},$$

for which, when it holds that

$$bu^{\mathrm{U}}(1-\gamma_u) < K\varepsilon, \tag{A.3}$$

then $y \in \mathcal{Y}$ follows from $\eta \in \Omega_1$.

First of all, we study boundedness of η_3 in case $\eta \in \Omega_1$. To that end, the estimate of $\dot{\Phi}(t)$ is obtained as follows. Having differentiated $\Phi(t)$, it is written:

$$\dot{\Phi}(t) = l\Big(\dot{z}(t) - \dot{z}_f(t)\Big) - \dot{u}_f(t) = -l\Big(\Phi(t) - F(t)\Big) = -l\tilde{\Phi}(t), \tag{A.4}$$

where $\Phi(t_0) = -u_f(t_0)$ and $\tilde{\Phi}(t) = \Phi(t) - F(t)$.

Then $\dot{\Phi}(t)$ is bounded if the error $\tilde{\Phi}(t)$ is bounded. A differential equation for $\tilde{\Phi}(t)$ is written as:

$$\dot{\tilde{\Phi}}(t) = -l\tilde{\Phi}(t) - \dot{F}(t),\ \tilde{\Phi}(t_0) = -u_f(t_0) - F(t_0). \tag{A.5}$$

When $g(t) \in \mathcal{G}$, $u(t) \in \mathcal{U}$, ($\eta(t) \in \Omega_1 \Rightarrow y(t) \in \mathcal{Y}$) for all $t \ge t_0$, Assumptions 1, 2 and condition $|-u_f(t_0) - F(t_0)| \le \tau\gamma_v\rho$, $\tau > 0$ are met, then there exists the following upper bound for the solution of equation (A.5):

$$\begin{aligned}\left|\tilde{\Phi}(t)\right| &= \left|e^{-l(t-t_0)}\tilde{\Phi}(t_0) - \int_{t_0}^{t} e^{-l(t-\tau)}\dot{F}(\tau)\,d\tau\right| \\ &\le \tau\gamma_v\rho + \gamma_v\rho\int_{t_0}^{t} e^{-l(t-\tau)}d\tau \le \left(\tau + l^{-1}\right)\gamma_v\rho\end{aligned} \tag{A.6}$$

for all $t \ge t_0$.

Then the derivative of the quadratic function $V = \frac{1}{2}\eta_3^2$ for all $\eta \in \Omega_1$ is obtained as:

$$\begin{aligned}\dot{V} &= \eta_3\left[-\frac{\mu\kappa\gamma_v\rho}{\gamma_u}\sigma\left(\frac{1}{\lambda_2}\eta_3\right) - \sigma'\left(\frac{1}{\gamma_u}\Phi(t)\right)\frac{1}{\gamma_u}l\tilde{\Phi}(t)\right] \\ &\le |\eta_3|\left[-\frac{\mu\kappa\gamma_v\rho}{\gamma_u}\left|\sigma\left(\frac{1}{\lambda_2}\eta_3\right)\right| + \kappa\frac{1}{\gamma_u}l\left|\tilde{\Phi}\right|\right] \\ &\le |\eta_3|\left[-\frac{\mu\kappa\gamma_v\rho}{\gamma_u}\left|\sigma\left(\frac{1}{\lambda_2}\eta_3\right)\right| + \frac{\kappa(l\tau+1)\gamma_v\rho}{\gamma_u}\right].\end{aligned} \tag{A.7}$$

On the other hand, for all $\eta_3 \notin \{\eta_3 \in \mathbb{R}\colon\ |\eta_3| < \lambda_2\}$, owing to property c) of the function $\sigma(.)$ and conditions $\mu > 1$, $\tau > 0$, $\tau + \frac{1-\mu}{l} < 0$, it holds that:

$$\dot{V} \le -\underbrace{\Big(\mu - (l\tau+1)\Big)}_{>0}\frac{\kappa\gamma_v\rho}{\gamma_u}|\eta_3|.$$

Therefore, when $|\eta_3(t_0)| \le \lambda_2$, we have $|\eta_3(t)| \le \lambda_2$ for $\eta \in \Omega_1$ and all $t \ge t_0$ independently from the sign of $\dot{V}$ over a set $\{\eta_3 \in \mathbb{R}\colon |\eta_3| < \lambda_2\}$.

Next aim is to prove that $\eta(t) \in \Omega_1$ for all $t \geq t_0$ follows from $\eta(t_0) \in \Omega_1$.

The derivative of the quadratic function $V_1 = \frac{1}{2}\eta_1^2$ for all $\eta \in \Omega_1$ is written as:

$$\begin{aligned}\dot{V}_1 &= \eta_1\left[(1-\gamma_u)\left(-\sigma\left(\frac{K}{1-\gamma_u}\eta_1\right)+\eta_2\right)+F+\gamma_u u_c \pm \gamma_u\sigma\left(\frac{F}{\gamma_u}\right)\right]\\ &= \eta_1(1-\gamma_u)\left(-\sigma\left(\frac{K}{1-\gamma_u}\eta_1\right)+\eta_2\right)\\ &+\eta_1\left[F-\gamma_u\sigma\left(\frac{F}{\gamma_u}\right)+\gamma_u\left(u_c \pm \sigma\left(\frac{\Phi}{\gamma_u}\right)+\sigma\left(\frac{F}{\gamma_u}\right)\right)\right].\end{aligned} \tag{A.8}$$

On the other hand, when we satisfy (4.7), Assumptions 1 and 2 for some

$$\delta := \sup_{|F|\leq\gamma_u}\left|\frac{F}{\gamma_u}-\sigma\left(\frac{F}{\gamma_u}\right)\right|$$

and all $\eta \in \Omega_1$, the following inequality holds:

$$\left|F-\gamma_u\sigma\left(\frac{1}{\gamma_u}F\right)\right| = \left|\frac{F}{\gamma_u}-\sigma\left(\frac{F}{\gamma_u}\right)\right|\gamma_u \leq \delta\gamma_u.$$

On the other hand, by virtue of the Lipchitz property of the function $\sigma(\cdot)$ for some $L > \kappa$, when inequalities (4.7) and Assumptions 1 and 2 are met, owing to the above-given derivations, for all $\eta \in \Omega_1$ the following inequality holds:

$$\left|\sigma\left(\frac{1}{\gamma_u}\Phi\right)-\sigma\left(\frac{1}{\gamma_u}F\right)\right| \leq \frac{L}{\gamma_u}|\Phi-F| = \frac{L}{\gamma_u}\left|\tilde{\Phi}\right| \leq \frac{L\left(\tau+l^{-1}\right)\gamma_v\rho}{\gamma_u}.$$

On the third hand, by what has been proved above, when $|\eta_3(t_0)| \leq \lambda_2$ and inequalities (4.7), Assumptions 1 and 2 are met, for all $\eta \in \Omega_1$ the following holds:

$$\left|u_c+\sigma\left(\frac{1}{\gamma_u}\Phi\right)\right| = |\eta_3| \leq \lambda_2.$$

Then the upper bound for the derivative (A.8) for all $\eta \in \Omega_1$ is written as

$$\dot{V}_1 \leq |\eta_1|\left[(1-\gamma_u)\left(-\sigma\left(\frac{K}{1-\gamma_u}\eta_1\right)+|\eta_2|\right)+\gamma_u\delta+L\left(\tau+l^{-1}\right)\gamma_v\rho+\gamma_u\lambda_2\right].$$

If $|\eta_1| = \frac{1-\gamma_u}{K}$ and $|\eta_2| \leq \lambda_1$, then the condition $\dot{V}_1 < 0$ is met when

$$(1-\gamma_u)(-1+\lambda_1)+\delta\gamma_u+L\left(\tau+l^{-1}\right)\gamma_v\rho+\gamma_u\lambda_2 < 0. \tag{A.9}$$

Moreover, the following upper bound of $|\dot{\eta}_1|$ for all $\eta \in \Omega_1$holds:

$$|\dot{\eta}_1| \leq (1-\gamma_u)(1+\lambda_1)+\delta\gamma_u+L\left(\tau+l^{-1}\right)\gamma_v\rho+\gamma_u\lambda_2.$$

The derivative of the quadratic function $V_2 = \frac{1}{2}\eta_2^2$ for all $\eta \in \Omega_2$ is written as

$$\begin{aligned}\dot{V}_2 &= \eta_2\left[-\frac{(1-\mu\kappa\gamma_v)\rho}{1-\gamma_u}\sigma\left(\frac{1}{\lambda_1}\eta_2\right)+\frac{K}{1-\gamma_u}\sigma'\left(\frac{K}{1-\gamma_u}\eta_1\right)\dot{\eta}_1\right]\\ &\leq |\eta_2|\left[-\frac{(1-\mu\kappa\gamma_v)\rho}{1-\gamma_u}\left|\sigma\left(\frac{1}{\lambda_1}\eta_2\right)\right|+\left|\frac{K}{1-\gamma_u}\sigma'\left(\frac{K}{1-\gamma_u}\eta_1\right)\dot{\eta}_1\right|\right].\end{aligned} \tag{A.10}$$

If $|\eta_1| > \frac{1-\gamma_u}{K}$, than, following the property a) of the function $\sigma(.)$, it holds that $\sigma'\left(\frac{K}{1-\gamma_u}\eta_1\right) = 0$ and, when $|\eta_2| = \lambda_1$, we have $\dot{V}_2 < 0$. If $|\eta_1| \le \frac{1-\gamma_u}{K}$, then, by virtue of $\eta \in \Omega_2$, we also have that $\eta \in \Omega_1$ and

$$\left|\frac{K}{1-\gamma_u}\sigma'\left(\frac{K}{1-\gamma_u}\eta_1\right)\dot{\eta}_1\right| \le \frac{\kappa K}{1-\gamma_u}\left[(1-\gamma_u)(1+\lambda_1) + \delta\gamma_u + L\left(\tau + l^{-1}\right)\gamma_v\rho + \gamma_u\lambda_2\right],$$

therefore, when $|\eta_2| = \lambda_1$, the condition $\dot{V}_2 < 0$ is met if:

$$-\frac{(1-\mu\kappa\gamma_v)\rho}{1-\gamma_u} + \frac{\kappa K}{1-\gamma_u}\left[(1-\gamma_u)(1+\lambda_1) + \delta\gamma_u + L\left(\tau + l^{-1}\right)\gamma_v\rho + \gamma_u\lambda_2\right] < 0. \tag{A.11}$$

If $\eta \in \partial\Omega_i$, $i = 1, 2$, then $|\eta_1| = \frac{1-\gamma_u}{K}$, $|\eta_2| = \lambda_1$ for $i = 1$ or $|\eta_2| = \lambda_1$ for $i = 2$. As $\Omega_1 \subset \Omega_2$, then, owing to the proven above, $\dot{V}_2 < 0$, therefore, if $\eta(t_0) \in \Omega_1$, then $\eta(t) \in \Omega_1$ for all $t \ge t_0$, which, when (A.3) holds, means that $y(t) \in \mathcal{Y}$ for all $t \ge t_0$.

As a result, combining (A.3), (A.9) and (A.11), we have that, when Assumptions 1–3 are met and, under given parameters

$$\gamma_u \in (0, 1), \gamma_v \in (0, 1), bu^{\mathrm{U}} > 0, \delta \in (0, 1), \kappa \in (1, 2), L > \kappa, \rho > 0, \varepsilon > 0$$

and some scalar $1 < \mu^* < \frac{1}{\kappa\gamma_v}$, inequalities (4.9) are feasible with respect to

$$\Psi_1 = \frac{1}{K}, \Psi_2 = \lambda_1, \Psi_3 = \lambda_2, \Psi_4 = l^{-1}, \Psi_5 = \tau,$$

then the law (4.4), (4.5) with parameters (4.10), initial conditions (4.7) and (4.11) for all $t \ge t_0$ ensures $u(t) \in \mathcal{U}$ and $y(t) \in \mathcal{Y}$.

Proof of Lemma 1. Equation (4.9) is rewritten in the following form:

$$\begin{gathered}
(1-\mu^*)\Psi_4 + \Psi_5 < 0, \\
(1-\gamma_u)\Psi_2 + \gamma_u\Psi_3 + L\gamma_v\rho(\Psi_4 + \Psi_5) + (1+\delta)\gamma_u - 1 < 0, \\
-\frac{(1-\mu^*\kappa\gamma_v)\rho}{\kappa}\Psi_1 + (1-\gamma_u)\Psi_2 + \gamma_u\Psi_3 + L\gamma_v\rho(\Psi_4 + \Psi_5) + 1 - (1-\delta)\gamma_u < 0, \\
(1-\gamma_u)bu^{\mathrm{U}}\Psi_1 < \varepsilon.
\end{gathered} \tag{A.12}$$

Considering sufficiently small Ψ_2, Ψ_3, Ψ_4, Ψ_5, inequalities (A.12) are satisfied if the following inequalities hold:

$$\begin{gathered}
\Psi_4 > \frac{\alpha\delta\gamma_u}{\mu - 1}, \quad \gamma_u < \frac{1}{1 + (1+\alpha)\delta}, \\
\frac{1 - (1 - (1+\alpha)\delta)\gamma_u}{(1-\mu\kappa\gamma_v)\rho}\kappa < \Psi_1 < \frac{\varepsilon}{(1-\gamma_u)bu^{\mathrm{U}}}
\end{gathered} \tag{A.13}$$

for $1 > \alpha > 0$.

The third inequality from (A.13) can be met, if it holds for $\varepsilon > 0$ that

$$\varepsilon > \frac{1 - (1 - (1+\alpha)\delta)\gamma_u}{(1-\mu\kappa\gamma_v)\rho}(1-\gamma_u)bu^{\mathrm{U}}\kappa = \underline{\varepsilon}, \tag{A.14}$$

which completes the proof of lemma.

Proof of Proposition 1. To prove the proposition, the upper bound of the solution of equation (4.16) is written as

$$\left|\tilde{F}(t)\right| = \left|-e^{-kt}F(0) - \int_0^t e^{-k(t-\tau)}\frac{\dot{f}(t,\ x,\ u)}{bu^{\mathrm{U}}}d\tau\right| \le e^{-kt}F(0) + \frac{\gamma_v\rho}{k}. \tag{A.15}$$

Then, using the initialization algorithm (4.13), (4.15), the satisfaction of inequality (4.11b) will follow from the satisfaction of the inequality

$$e^{-kt_0}F(0) + \frac{\gamma_v\rho}{k} \le \tau\gamma_v\rho. \tag{A.16}$$

Conducting simple reasoning, we obtain that inequality (A.16) holds for all

$$k > k^* = \max\left\{-t_0^{-1}\ln\left(\frac{\vartheta\tau\gamma_v\rho}{F(0)}\right), \frac{1}{(1-\vartheta)\tau}\right\}, \vartheta \in (0,\ 1),$$

which completes the proof of proposition.

REFERENCES

1. Furtat, I.B. and Gushchin, P.A., Control of dynamical plants with a guarantee for the controlled signal to stay in a given set, *Autom. Remote Control*, 2021, vol. 82, no. 9, pp. 654–669.
2. Furtat, I.B., Gushchin, P.A., and Nguyen, B.H., Control of dynamic systems under input and output constraints, *Autom. Remote Control*, 2023, vol. 84, no. 4, pp. 355–368.
3. Furtat, I.B., Density systems: Analysis and control, *Autom. Remote Control*, 2023, vol. 84, no. 11, pp. 1175–1190.
4. Bechlioulis, C.P. and Rovithakis, G.A., Adaptive control with guaranteed transient and steady state tracking error bounds for strict feedback systems, *Automatica*, 2009, vol. 45, no. 2, pp. 532–538.
5. Bechlioulis, C.P. and Rovithakis, G.A., Robust adaptive control of feedback linearizable MIMO nonlinear systems with prescribed performance, *IEEE Trans. Autom. Control*, 2008, vol. 53, no. 9, pp. 2090–2099.
6. Tee, K.P., Ge, S.S., and Tay, E.H., Barrier Lyapunov functions for the control of output-constrained nonlinear systems, *Automatica*, 2009, vol. 45, no. 4, pp. 918–927.
7. Tee, K.P., Ren, B., and Ge, S.S., Control of nonlinear systems with time-varying output constraints, *Automatica*, 2011, vol. 47, no. 11, pp. 2511–2516.
8. Glushchenko, A.I. and Lastochkin, K.A., Placement of input and output in given sets for one class of systems, *Large-Scale Syst. Control*, 2025, vol. 113, pp. 95–119.
9. Marconi, L. and Isidori, A., Robust global stabilization of a class of uncertain feedforward nonlinear systems, *Syst. Control Lett.*, 2000, vol. 41, no. 4, pp. 281–290.
10. Antipov, A.S., Krasnova, S.A., and Utkin, V.A., Synthesis of invariant nonlinear single-channel sigmoid feedback tracking systems ensuring given tracking accuracy, *Autom. Remote Control*, 2022, vol. 83, no. 1, pp. 32–53.
11. Utkin, V.A. and Utkin, A.V., Synthesis of invariant systems ensuring design constraints, *Proc. XIII All-Russ. Conf. Control Probl. (VSPU-2019)*, Moscow: ICS RAS, 2019, pp. 68–72.
12. Antipov, A.S., Krasnova, S.A., and Pivneva, S.V., Control of an uncertain electromechanical system ensuring design constraints, *Proc. 14th Int. Conf. Manag. Large-Scale Syst. Dev. (MLSD'2021)*, 2021, pp. 567–579.

13. Pesterev, A.V. and Morozov, Y.V., Stabilization of a chain of three integrators subject to a phase constraint, *Autom. Remote Control*, 2024, vol. 85, no. 7, pp. 597–603.

14. Kokunko, Y.G., Krasnova, S.A., and Utkin, V.A., Cascade synthesis of differentiators with piecewise linear correction signals, *Autom. Remote Control*, 2021, vol. 82, no. 7, pp. 1144–1168.

15. Tsykunov, A.M., *Adaptive and Robust Output-Feedback Control of Dynamic Plants*, Moscow; Fizmatlit, 2009, 268 p.

16. Isidori, A., *Lectures in Feedback Design for Multivariable Systems*, Basel: Springer International Publishing, 2017.

17. Utkin, V.A. and Utkin, A.V., Problem of tracking in linear systems with parametric uncertainties under unstable zero dynamics, *Autom. Remote Control*, 2014, vol. 75, pp. 1577–1592.

18. Khalil, H.K., Cascade high-gain observers in output feedback control, *Automatica*, 2017, vol. 80, pp. 110–118.

19. Astolfi, D. and Marconi, L., A high-gain nonlinear observer with limited gain power, *IEEE Trans. Autom. Control*, 2015, vol. 60, no. 11, pp. 3059–3064.

20. Afanasyev, V.N., Kolmanovskii, V.B., and Nosov, V.R., *Mathematical Theory of Control Systems Design*, Moscow: Vysshaya Shkola, 2003.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 10, pp. 926–939.
© 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. 10, pp. 84–103.

NONLINEAR SYSTEMS

An Integrated Approach to Fault Accommodation in Fuzzy Discrete-Event Systems

A.E. Shumsky[*,**,a] **and A.N. Zhirabok**[*,**,b]

[*] *Far Eastern Federal University, Vladivostok, Russia*
[**] *Institute of Marine Technology Problems, Far Eastern Branch, Russian Academy of Sciences, Vladivostok, Russia*
e-mail: [a] *alexeyshumsky@yandex.ru,* [b] *zhirabok@mail.ru*

Received July 17, 2025
Revised January 12, 2026
Accepted January 23, 2026

Abstract—This paper considers the problem of fault accommodation in complex safety-critical discrete-event systems described by the fuzzy finite state automaton model. The problem is solved by forming control that ensures complete decoupling from fault-induced effects. The fault-tolerant control scheme contains several channels activated depending on the diagnosed fault. Each channel includes two main blocks: 1) an auxiliary deterministic finite state automaton (observer) estimating the fuzzy vector of possibilities of system transitions to various states and 2) a converter forming the input of the controlled system based on a new external input and the observer state. An integrated approach (the pair algebra of fuzzy coverings) is proposed to describe the blocks and the new input. A distinctive feature of this approach is the joint use of mathematical constructs of the pair algebra of coverings (system of sets) and fuzzy logic.

Keywords: fault-tolerant control, fault accommodation, discrete-event systems, fuzzy finite state automata, observers, pair algebra of fuzzy coverings, optimization approach

DOI: 10.7868/S1608303226100044

1. INTRODUCTION

The fault-tolerance requirement imposed on controlled safety-critical systems assumes the ability of such systems to adapt their behavior to faults arising during operation. Following the classification of adaptive systems into self-tuning and self-organizing systems, one distinguishes between two fundamentally different approaches to ensuring fault tolerance. The first approach (self-tuning or fault accommodation) is to form a fault-tolerant control law under which a faulty system will continue operation while preserving its most important performance indices within admissible limits. The second approach (self-organization) is to introduce structural changes into the system and its reconfiguration in order to remove failed elements and use reserve ones.

In this paper, we address the fault accommodation problem. In the monograph [1], it was solved using an original approach with the following stages: 1) fault detection, 2) fault isolation, 3) fault identification (determination of the fault value), 4) construction of a model of the faulty system, and, finally, 5) fault-tolerant control design based on the model. The approach to accommodation proposed therein involves mathematical models of controlled systems in the form of linear ordinary differential equations. Stages 1–3 are implemented using well-developed fault diagnosis methods for linear dynamic systems, and an appropriate control law (stage 5) is constructed by optimal control methods.

The efficiency of fault accommodation is primarily determined by the speed of forming the fault-tolerant control law. Therefore, the possibility of reducing the time costs of implementing each accommodation stage, or even excluding some stages, is crucial for improving accommodation. The second factor impacting the efficiency of accommodation is the accuracy of the controlled system model used to form the fault-tolerant control law. Consequently, it is necessary to apply nonlinear dynamic models, which most accurately describe the behavior of real systems, and robust methods for fault diagnosis and fault-tolerant control, which reduce the impact of uncontrolled factors (exogenous disturbances and model errors).

Several works, e.g., see [2, 3], are remarkable for the use of adaptive sliding mode observers for simultaneously estimating the state and parameters of the system and fault-tolerant control design methods based on sliding modes. Such a framework provides acceptable time costs for implementing the accommodation stages.

For controlled systems modeled by nonlinear ordinary differential equations, an approach without stages 3 and 4 was proposed in [4]. Its peculiarity is the use of a control law ensuring complete decoupling from the effects caused by faults. This approach was later extended to controlled systems described by discrete-time models in the form of nonlinear difference equations [5].

In addition to the above classes of continuous- or discrete-time controlled systems described by linear or nonlinear differential or difference equations, there exists an important class of so-called discrete-event systems (DESs) [6, 7]. These include not only systems that naturally fall into the class (e.g., digital information processing and control systems). Many complex systems conventionally classified as continuous, such as physical, technical (including human–machine ones), and socioeconomic systems, can be treated as DESs at the upper level of their hierarchy.

DESs possess the following distinctive features:

1. DESs are discrete in time and in the values of state variables.

2. The state space of a DES is finite; states may have different semantic content, e.g., “the DES is idle,” “the DES operates in a certain mode” (possibly caused by a system fault), “the operation of the DES is completed,” etc.

3. The operation of a DES is determined by events that may be consequences of certain external actions (inputs or commands) and faults, e.g., “start operation,” “change the operating mode of the fault-free DES,” “correct the control law when a certain fault is detected in the DES,” “terminate the operation of the DES,” etc.

4. As a rule, a DES behaves randomly, which is associated with the possibility of realizing different transitions from one state to another initiated by the same event.

The following mathematical models are commonly employed to describe DESs for subsequent fault diagnosis and fault-tolerant control design [8]:

1) deterministic finite state automata,

2) probabilistic finite state automata and Markov chains, and

3) Petri nets.

This paper focuses on the use of finite state automaton (FSA) models. By assumption, the inputs and outputs of an FSA are formed as observable events (external and internal, respectively), while the occurrence of a DES fault is treated as a directly unobservable internal event. A solution of the fault accommodation problem in a DES based on the deterministic FSA (DFSA) model and the complete decoupling principle was presented in [9]. This solution involves the mathematical apparatus of the pair algebra of partitions [10].

Whenever the DFSA model of a DES is unavailable, along with the probabilistic FSA model, models of nondeterministic and fuzzy FSAs (NFSA and FFSA, respectively) can also be applied. Compared with their probabilistic counterparts, nondeterministic and fuzzy models reduce the

volume of required computations, leading to lower time costs for fault accommodation. NFSA and FFSA models reflect the situation of transitions to different states from a certain fixed state and a certain fixed input of the automaton. Note that in the NFSA case, no priority can be assigned to the possibility of realizing a particular transition. In contrast, in the FFSA case, additional (e.g., statistical) information, expert estimates, learning results, etc., can be utilized to obtain the degree of confidence in the realization of each possible transition.

The aim of this paper is to extend the results of [9] to the case of FFSA models of DESs. The problem under consideration is to develop fault-tolerant control design methods oriented toward FFSA models of DESs. To solve this problem, we employ mathematical constructs of the pair algebra of coverings [10], fuzzy logic [11], and also the concepts of a fuzzy automaton [12] and its determinizer [13].

2. PROBLEM STATEMENT

Let the fault-free operation of a DES be described by the FFSA model

$$A = (U, X, Y, \delta, \lambda, x(0)), \tag{1}$$

where $U = \{u_1, \ldots, u_m\}$, $X = \{x_1, \ldots, x_n\}$, and $Y = \{y_1, \ldots, y_l\}$ are finite sets of inputs, states, and outputs, respectively; $x(0) \in X$ is a known initial state; $\delta : X \times U \to \mu(X)$ and $\lambda : X \to \mu(Y)$ are fuzzy transition and output functions, respectively; $\mu(X) \in \{\mu(x_i) \in [0,1], 1 \leqslant i \leqslant n\}$ and $\mu(Y) \in \{\mu(y_j) \in [0,1], 1 \leqslant j \leqslant l\}$ are fuzzy sets. For the fuzzy transition function δ, we will adopt a matrix representation in which transitions performed under an input $u_k \in U$ are described by a matrix M^k of dimensions $n \times n$: its element $M^k_{i,j} \in [0,1]$ characterizes the degree of confidence that the automaton A will pass from a state $x_i \in X$ to a state $x_j \in X$ under an input u_k. Let $S(U)$ denote the set of all matrices M^k, $1 \leqslant k \leqslant m$. To specify the fuzzy output function λ, we will use a matrix L of dimensions $l \times n$: its element $L_{i,j}$ determines the degree of confidence that the output $y_i \in Y$ will be formed by the automaton A in a state $x_j \in X$. (Note that the same output can be formed in different DES states.) Let $S(Y)$ denote the set of all rows L_i, $1 \leqslant i \leqslant l$, of the matrix L.

Model (1) corresponds to a special case of an FFSA with deterministic inputs and fuzzy transition and output functions [15]. If it is difficult to justify the degree of confidence in the possibility of transitions, one can proceed by setting $M^k_{i,j} = 1$ for a correct transition and $M^k_{i,j} = 0$ for an incorrect transition. Thereby, one will pass from the FFSA model of the DES to its NFSA model.

Let $X_{i,k}$, $X_{i,k} \subseteq X$, be the set of all states reachable from a state $x_i \in X$ under an input $u_k \in U$. Suppose that a fault f_s, $1 \leqslant s \leqslant N$, in the DES (1) can be represented by a distortion of the transition function δ such that under an input $u_k \in U$, an incorrect transition from a state x_i to a state $x_t \notin X_{i,k}$ is realized instead of an correct one to a state $x_j \in X_{i,k}$. We will write this fact as follows: $f_s : (x_j \to x_t)_{i,k}$. In this case, the matrix M^k_s of the DES with the fault f_s is obtained from the matrix M^k of the fault-free DES by changing the value of the element $M^k_{i,t}$ from 0 to 1. The analysis below is based on the hypothesis of single faults from a given list $F = \{f_1, \ldots, f_N\}$.

Consider the sequence and content of actions necessary for forming a control law that ensures complete decoupling from the effects caused by a fault f_s. We introduce the FFSA model of the DES with a fault f_s :

$$A_s = (U, X, Y, \delta_s, \lambda, x(0)). \tag{2}$$

The transition function δ_s of model (2) can be obtained by supplementing the set of transitions determined by the function δ of model (1) with a transition initiated by a fault $f_s \in F$ through changing the matrix M^k in the manner described above. Owing to the features of constructing the matrix M^k, model (2) reflects the possibility of a correct transition during the fault-free operation

of the DES and also the possibility of an incorrect transition to a state caused by a DES fault. Hence, we can speak of a certain correspondence between the behavior of model (2) and that of both the fault-free DES and the DES with a fault.

Following [9], let the input u ensuring accommodation to a fault f_s be formed in accordance with the law

$$u = g(x_0, y, u_*), \tag{3}$$

where g is a mapping to be determined; u_* is a new input; x_0 is the state of the deterministic FSA

$$A_0 = (X_0, U \times Y, \delta_0, x_0(0)), \tag{4}$$

whose transition function δ_0 is defined by the mapping $X_0 \times U \times Y \to X_0$ and is found by determinizing the automaton A_s; $x_0(0) \in X_0$ is the initial state of the automaton A_0; the output function of the automaton A_0 is not considered an independent object. Note that the description of the deterministic FSA A_0, hereinafter called the observer, does not depend on the (non-)occurrence of a DES fault f_s.

To find the control law ensuring accommodation of the DES (1) to a fault $f_s \in F$, it is necessary to solve the following problems sequentially:

1) construct model (2) of the DES with a fault f_s;

2) find the observer (4) by determinizing the automaton (2);

3) find the mapping g and form the control law (3).

These problems define the structure of the paper. In Section 3, we present an approach to determinizing an FFSA by fuzzy logic methods and reveal its drawback. To provide theoretical grounds for an approach that eliminates the drawback, we elaborate the algebra of fuzzy coverings (Section 4) and propose an algorithm for constructing an observer (Section 5). The issues of fault-tolerant control design are discussed in Section 6. An illustrative example is given in Section 7. The outcomes of this research are summarized in the Conclusions (Section 8).

3. DETERMINIZATION OF THE FUZZY FINITE STATE AUTOMATON A_s

Recall several operations of fuzzy logic that are important for the subsequent presentation. A fuzzy matrix $B = \{B_{ij}\}$ is a matrix with elements $B_{ij} \in [0, 1]$. Hence, the matrices from the set $S(U)$ introduced above are fuzzy matrices. Now, let B and C be fuzzy matrices of dimensions $a \times b$ and $b \times c$, respectively. The product of fuzzy matrices is defined as [13]

$$(BC)_{ij} = \max\limits_{ij} \min(B_{ih} C_{hj}) = \max(\min(B_{i1} C_{1j}), \ldots, \min(B_{ib} C_{bj}),$$

where $1 \leqslant i \leqslant a$, $1 \leqslant h \leqslant b$, and $1 \leqslant j \leqslant c$. This formula generalizes the well-known matrix multiplication one [16, p. 24]; it is obtained by replacing the product (sum) of matrix elements with their minimum (maximum, respectively). Here is a simple example.

Example 1. Consider matrices of the form

$$B = \begin{pmatrix} 0.1 & 0.9 \\ 0.7 & 0.2 \end{pmatrix}, \quad C = \begin{pmatrix} 0.6 & 0.3 \\ 0.4 & 0.8 \end{pmatrix}.$$

Straightforward calculations yield

$$\begin{aligned} (BC)_{11} &= \max(\min(0.1\ 0.6), \min(0.9\ 0.4)) = 0.4, \\ (BC)_{12} &= \max(\min(0.1\ 0.3), \min(0.9\ 0.8)) = 0.8, \\ (BC)_{21} &= (\min(0.7\ 0.6), \min(0.9\ 0.4)) = 0.6, \\ (BC)_{22} &= \max(\min(0.7\ 0.3), \min(0.2\ 0.8)) = 0.3. \end{aligned}$$

As a result, the product of fuzzy matrices is

$$BC = \begin{pmatrix} 0.4 & 0.8 \\ 0.6 & 0.3 \end{pmatrix}.$$

We denote by E the set of fuzzy vectors of dimensions $1 \times n$; their components can take values in the interval $[0, 1]$. Let $A' = (U, X, \delta)$ be a fuzzy semiautomaton (i.e., an automaton A without the output function) and $S(U)$ be the set of fuzzy transition matrices of the automaton. Suppose that the signature (the set of admissible operations) Σ includes all operations involving multiplication of a fuzzy vector of dimensions $1 \times n$ on the right by a matrix from the set $S(U)$. Also, let $E^0 \subseteq E$ be the set of all unit vectors of dimensions $1 \times n$, called generating states of the determinizer. The closure $[E]_\Sigma$ of the set E with respect to the signature Σ is the set of all vectors (including those of E) that can be obtained by applying operations from Σ to vectors from E. We propose the following algorithm.

Algorithm 1 (construction of the closure).

1. Set $i = 0$.
2. Find the set of vectors $E^{i+1} = E^i \cup [E^i]_\Sigma$.
3. If $E^{i+1} = E^i$, set $[E]_\Sigma = E^i$. End.
4. Assign $i := i + 1$ and return to step 2.

According to [13], the determinizer $D(A')$ of the FFSA A' is a DFSA described by the triple

$$D(A') = (S(X), U, \delta_D), \tag{5}$$

where $S(X) = [E]_{S(U)}$ and $\delta_D : S(X) \times S(U) \to S(X)$ is the determinizer's transition function defined as

$$\delta_D(\mu_i, u_k) = \mu_i M^k, \quad \mu_i \in S(X), \quad M^k \in S(U). \tag{6}$$

Thus, each state μ_i of the determinizer is associated with a certain fuzzy vector of membership grades characterizing the possibility of reaching certain states of the controlled DES during its operation.

In [14], the maximum number of states of the determinizer was estimated as

$$\sharp S(X) \leqslant q^n - 1, \tag{7}$$

where the symbol $\sharp$ denotes cardinality; q is the number of different values of elements of fuzzy matrices from the set $S(U)$; as before, n is the number of states of the FFSA.

According to the relation (7), the number of states of the determinizer can substantially exceed that of the original automaton and grows exponentially with the latter. This fact has minimal impact on fault diagnosis since such problems are solved in real time without requiring the complete description of the determinizer [14]. However, in the case of fault accommodation, additional computations are needed to find the mapping g and choose the control u_* based on the determinizer's description. The volume and feasibility of computations will be directly related to the determinizer's dimension. As a consequence, there may exist a restriction on the dimension of the original system model. To overcome this restriction, we propose to design the observer A_0 with another set of states, i.e., to replace the closure E with a certain set not exceeding, by the dimension, the original model (1), while preserving its closedness with respect to operations from the signature Σ. To find such a set, the pair algebra of fuzzy coverings will be introduced below.

Another difference between the observer and the determinizer is that the former's transition function also depends on the DES output. Therefore, it is possible to refine the membership grades (the components of fuzzy vectors associated with x_0) considering the additional information contained in the DES output. This feature explains the term "observer" used for the automaton A_0.

4. THE PAIR ALGEBRA OF FUZZY COVERINGS

In [17], constructs of the pair algebra of coverings were introduced for solving the determinization problem of NFSAs. Below, these constructs are extended to fuzzy sets to propose an approach to determinizing FFSAs considering the information contained in the fuzzy model of DESs.

Recall that a covering π of a set X is a set of its subsets (blocks of the covering) $\pi = \{B_{\pi,1}, \ldots, B_{\pi,\nu}\}$ such that

$$B_{\pi,k} \subseteq X, \qquad \bigcup_{k=1}^{\nu} B_{\pi,k} = X, \qquad B_{\pi,k} \subseteq B_{\pi,j} \Rightarrow B_{\pi,k} = B_{\pi,j}.$$

Note that under the additional condition

$$B_{\pi,k} \cap B_{\pi,j} = \emptyset, \quad k \neq j,$$

the above constructs define a partition of X. The set of all coverings of a set X is denoted by Π_X.

Assume that the blocks of the coverings considered below are fuzzy sets. It is convenient to specify such blocks by a row vector of membership functions. In this case, the notation $\mu_{B_{\pi,j}} = (0\ \ 0.1\ \ 0.7\ \ 0\ \ 0\ \ 0.5)$ means that (a) the cardinality of X is 6, (b) the jth block of the covering π contains the states x_2, x_3, and x_6, and (c) the nonzero membership grades of these states for the jth block of the covering are $\mu_{B_{\pi,j}}(x_2) = 0.1$, $\mu_{B_{\pi,j}}(x_3) = 0.7$, and $\mu_{B_{\pi,j}}(x_6) = 0.5$.

To perform operations on coverings, it is necessary to find the intersections and unions of their blocks. Let π and σ be certain coverings of a set X. In accordance with the rules of operations on fuzzy sets [14], we write

$$\mu_{B_{\pi,j} \cap B_{\sigma,k}}(x_i) = \min(\mu_{B_{\pi,j}}(x_i), \mu_{B_{\sigma,k}}(x_i)),$$
$$\mu_{B_{\pi,j} \cup B_{\sigma,k}}(x_i) = \max(\mu_{B_{\pi,j}}(x_i), \mu_{B_{\sigma,k}}(x_i)).$$

A covering π is said to be less than or equal to a covering σ, $\pi \leqslant \sigma$, if, for each block $B_{\pi,j}$ of π, there always exists a block $B_{\sigma,k}$ of σ such that $B_{\pi,j} \subseteq B_{\sigma,k}$. In this case, $B_{\pi,j} \cup B_{\sigma,k} = B_{\sigma,k}$, and hence the inclusion $B_{\pi,j} \subseteq B_{\sigma,k}$ requires the fuzzy set $B_{\sigma,k}$ to dominate the fuzzy set $B_{\pi,j}$, i.e.,

$$\mu_{B_{\pi,j}}(x_i) \leqslant \mu_{B_{\sigma,k}}(x_i), \quad 1 \leqslant i \leqslant n. \tag{8}$$

There are two special coverings, denoted by $\mathbf{0}$ and $\mathbf{1}$, representing partitions such that each block of $\mathbf{0}$ contains only one element of a set X, while $\mathbf{1}$ includes only one block containing all elements of X. For an arbitrary covering π of a set X, we have

$$\mathbf{0} \leqslant \pi \leqslant \mathbf{1}.$$

As is known [10], the set of all coverings of a set X with a partial order relation forms a lattice. Consequently, for each pair of coverings π and σ, one can find the two coverings $\inf(\pi, \sigma)$ and $\sup(\pi, \sigma)$. These coverings are conventionally denoted by $\pi \times \sigma$ and $\pi + \sigma$, respectively (the product and sum of coverings). To calculate them, we use the rules formulated in [17].

1. Each block of the covering $\pi \times \sigma$ is the intersection of certain blocks of the coverings π and σ.
2. Each block of the covering $\pi + \sigma$ is the union of intersecting blocks of the coverings π and σ.

We emphasize a peculiarity that may arise during the operation $\times$. Let the zero partition be the result of this operation in accordance with the rules of traditional algebra of coverings [10]. However, with formal application of the above method considering membership functions, blocks of the zero partition may appear for which the values of the membership functions of states are strictly less than 1. This indicates a singularity in the process of transforming the FFSA model of

the system (e.g., during its determinization). If necessary, a possible recipe here is to pass (fully or partially) to the NFSA model of the system by replacing the nonzero values of the corresponding membership functions with 1.

Example 2. Consider $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and coverings $\pi = \{(x_1); (x_3); (x_2, x_4); (x_2, x_5, x_6)\}$, $\sigma = \{(x_1); (x_2, x_5); (x_3, x_4); (x_4, x_6)\}$ with the membership functions of states for the corresponding blocks $\mu_{B_{\pi,4}}(x_2) = 0.5$ and $\mu_{B_{\sigma,3}}(x_4) = 0.3$; the other membership functions are equal to 1. By applying the rule for the intersection of fuzzy sets to the blocks of the coverings π and σ, we obtain the membership functions $\mu_{B_{\pi,1}\cap B_{\sigma,1}}(x_1) = \mu_{B_{\pi,2}\cap B_{\sigma,3}}(x_3) = \mu_{B_{\pi,3}\cap B_{\sigma,2}}(x_2) = \mu_{B_{\pi,3}\cap B_{\sigma,4}}(x_4) = \mu_{B_{\pi,4}\cap B_{\sigma,4}}(x_6) = 1$, $\mu_{B_{\pi,4}\cap B_{\sigma,2}}(x_5) = 0.5$, and $\mu_{B_{\pi,3}\cap B_{\sigma,3}}(x_4) = 0.3$. The membership functions of states for the empty sets obtained by intersecting blocks of the partitions π and σ are zero. Note that $B_{\pi,3\cap\sigma,2} \subseteq B_{\pi,4\cap\sigma,2}$. Following the definition of a covering, it is necessary to replace both of these blocks by one block $B_{\pi,3\cap\sigma,2} \cup B_{\pi,4\cap\sigma,2}$ dominating them. Moreover, the fuzzy set $B_{\pi,3} \cap B_{\sigma,4}$ dominates the fuzzy set $B_{\pi,3} \cap B_{\sigma,3}$. As a result, we have arrived at the partition $\pi \times \sigma = \{(x_1); (x_3); (x_2, x_5); (x_4); (x_6)\}$.

Now, let us find the sum of these coverings, $\pi+\sigma$. By uniting the intersecting blocks of coverings, we obtain

$$\begin{gathered}\mu_{B_{\pi,1}\cup B_{\sigma,1}}(x_1) = \mu_{B_{\pi,2}\cup B_{\sigma,3}}(x_3) = \mu_{B_{\pi,2}\cup B_{\sigma,3}}(x_4) = \mu_{B_{\pi,4}\cup B_{\sigma,2}}(x_2)\\ = \mu_{B_{\pi,3}\cup B_{\sigma,2}}(x_4) = \mu_{B_{\pi,3}\cup B_{\sigma,2}}(x_5) = \mu_{B_{\pi,3}\cup B_{\sigma,3}}(x_3) = \mu_{B_{\pi,3}\cup B_{\sigma,3}}(x_4)\\ = \mu_{B_{\pi,4}\cup B_{\sigma,2}}(x_2) = \mu_{B_{\pi,4}\cup B_{\sigma,2}}(x_5) = \mu_{B_{\pi,4}\cup B_{\sigma,2}}(x_6) = \mu_{B_{\pi,4}\cup B_{\sigma,3}}(x_2)\\ = \mu_{B_{\pi,4}\cup B_{\sigma,3}}(x_3) = \mu_{B_{\pi,4}\cup B_{\sigma,3}}(x_4) = \mu_{B_{\pi,4}\cup B_{\sigma,3}}(x_5) = \mu_{B_{\pi,4}\cup B_{\sigma,3}}(x_6)\\ = \mu_{B_{\pi,2}\cup B_{\sigma,4}}(x_2) = \mu_{B_{\pi,4}\cup B_{\sigma,4}}(x_5) = \mu_{B_{\pi,4}\cup B_{\sigma,4}}(x_6) = 1,\\ \mu_{B_{\pi,3}\cup B_{\sigma,3}}(x_2) = 0.5, \quad \mu_{B_{\pi,4}\cup B_{\sigma,4}}(x_3) = 0.3.\end{gathered}$$

When constructing the resulting covering, we take into account that $B_{\pi,3} \cup B_{\sigma,3}$ dominates $B_{\pi,2} \cup B_{\sigma,3}$ and also $B_{\pi,3} \cup B_{\sigma,2} \subseteq B_{\pi,4} \cup B_{\sigma,4}$, $B_{\pi,3} \cup B_{\sigma,4} \subseteq B_{\pi,4} \cup B_{\sigma,4}$, and $B_{\pi,4} \cup B_{\sigma,2} \subseteq B_{\pi,4} \cup B_{\sigma,4}$. As a consequence, the dominating fuzzy sets $B_{\pi,3} \cup B_{\sigma,3}$ and $B_{\pi,3} \cup B_{\pi,4} \cup B_{\sigma,2} \cup B_{\sigma,4}$ are chosen as the independent blocks of the covering $\pi + \sigma$. The final covering is $\pi + \sigma = \{(x_1); (x_2, x_3, x_4); (x_2, x_4, x_5, x_6)\}$, where the second block represents a fuzzy set since $\mu B_{\pi+\sigma,2}(x_2) = 0.5$.

Let π and σ be certain coverings of a set X. For the convenience of further presentation, we give a concise description only for the main constructs of the pair algebra of coverings (used in this paper), i.e., the binary relation Δ, the concept of pair coverings, and the operator $\mathbf{m}$. More complete information on this issue can be found, e.g., in [10].

Coverings π and σ are said to satisfy the relation Δ, $(\pi, \sigma) \in \Delta$, and form a pair if

$$x \equiv x'(\pi) \Rightarrow \delta(x, u_k) \equiv \delta(x', u_k)(\sigma) \quad \forall u_k \in U. \tag{9}$$

Let $x, x' \in B_{\pi,j}$. According to [17], for each $u_k \in U$, we can write $\delta(x, u_k), \delta(x', u_k) \in B^k_{\sigma,j}$, where the vector of membership functions $\mu_{B^k_{\sigma,j}}$ is defined by

$$\mu_{B^k_{\sigma,j}} = \mu_{B(\pi,j)} M^k, \quad M^k \in M(U). \tag{10}$$

For a given covering π, there may exist several coverings satisfying the relation $(\pi, \sigma) \in \Delta$. In particular, $(\pi, \mathbf{1}) \in \Delta$ is valid for any covering π. We introduce the operator $\mathbf{m}$ that bijectively associates the covering $\mathbf{m}(\pi)$ with a covering π as follows:

$$(\pi, \mathbf{m}(\pi)) \in \Delta, \quad (\pi, \sigma) \in \Delta \Rightarrow \mathbf{m}(\pi) \leqslant \sigma.$$

The operator $\mathbf{m}(\pi)$ is computed by the following algorithm.

Algorithm 2 (computation of the operator $\mathbf{m}(\pi)$).

1. For each $u_k \in U$ and each j, $1 \leqslant j \leqslant \nu$, where ν is the number of blocks of a covering π, find the subset $B^k_{\sigma,j}$ using (10).

2. Write the set of non-empty subsets $Q = \{B^k_{\sigma,j}, 1 \leqslant k \leqslant m, 1 \leqslant j \leqslant \nu\}$.

3. Form the covering $\mathbf{m}(\pi)$ from Q by uniting those subsets for which one is contained in another. End.

Example 3. Consider an FFSA described by the matrices

$$M^1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad M^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad M^3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

and set $\pi = \mathbf{0}$. It is required to compute $\sigma = \mathbf{m}(\pi)$. When executing step 1 of Algorithm 2, with the first block $\mu_{B_{\pi,1}} = (1\ 0\ 0\ 0\ 0)$ of the covering π taken as a basis, we obtain the vector membership function $\mu_{B^1_{\sigma,1}} = \mu_{B_{\pi,1}} M^1 = (0\,1\,0\,0\,0)$ corresponding to the subset $B^1_{\sigma,1} = \{x_2\}$. Next, the vector membership functions $\mu_{B^2_{\sigma,1}} = \mu_{B_{\pi,1}} M^2 = (0\ 0\ 0\ 0\ 0)$ and $\mu_{B^3_{\sigma,1}} = \mu_{B_{\pi,1}} M^3 = (0\ 0\ 0\ 0\ 0)$, corresponding to the empty subsets $B^2_{\sigma,1}$ and $B^3_{\sigma,1}$, are found. For the remaining blocks of the covering π, similar computations yield non-empty fuzzy subsets with the membership functions $\mu_{B^3_{\sigma,1}} = (0\ 0\ 1\ 0\ 0)$, $\mu_{B^2_{\sigma,2}} = (0\ 0\ 0.7\ 0.3\ \ 0)$, $\mu_{B^3_{\sigma,2}} = (1\ 0.7\ 0\ 0\ 0)$, $\mu_{B^2_{\sigma,3}} = (0\ 0.4\ 0.6\ 0\ 0)$, $\mu_{B^3_{\sigma,3}} = (0\ 0\ 0.9\ 0\ 0.1)$, $\mu_{B^3_{\sigma,4}} = (0\ 0\ 0\ 0\ 1)$, and $\mu_{B^3_{\sigma,5}} = (1\ 0\ 0\ 0\ 0)$. In accordance with step 2 of Algorithm 2, $Q = \{B^3_{\sigma,5}; B^1_{\sigma,1}; B^2_{\sigma,3}; B^2_{\sigma,2}; B^3_{\sigma,3}; B^3_{\sigma,4}\} = \{(x_1); (x_2); (x_2, x_3); (x_3, x_4); (x_5)\}$. For step 3 of the algorithm, note that $B^1_{\sigma,1} \subseteq B^2_{\sigma,3}$ and $B^3_{\sigma,4} \subseteq B^3_{\sigma,3}$. We replace the above fuzzy sets with the dominating fuzzy ones $B^* = B^1_{\sigma,1} \cup B^2_{\sigma,3}$ and $B^{**} = B^3_{\sigma,4} \cup B^3_{\sigma,3}$ with the membership functions $\mu_{B^*} = (0\ 1\ 0.6\ 0\ 0)$ and $\mu_{B^{**}} = (0\ 0\ 0.9\ 0\ 0.1)$, respectively. Finally, we arrive at the fuzzy covering $\sigma = \{B^3_{\sigma,5}; B*; B^2_{\sigma,3}; B^{**}\} = \{(x_1); (x_2, x_3); (x_3, x_4); (x_3, x_5)\}$ with the corresponding membership functions.

Note that, in contrast to DFSAs, the relation $\mathbf{m}(\mathbf{0}) \neq \mathbf{0}$ always holds for NFSAs or FFSAs.

5. DESIGN OF THE OBSERVER A_0

For an arbitrary covering φ, by analogy with the constructs of the algebra of functions for continuous-time dynamic systems [4], we introduce the concepts of δ-invariance,

$$(\varphi, \varphi) \in \Delta,$$

and (λ, δ)-invariance,

$$(\pi_\lambda \times \varphi, \varphi) \in \Delta, \tag{11}$$

where the blocks of a fuzzy covering π_λ are specified by the rows of the matrix L. Obviously, δ-invariance is a special case of (λ, δ)-invariance (for $\pi_\lambda = \mathbf{1}$). The latter concept will be used to design the observer A_0. It has the following informational meaning [10]. Assume that at some discrete time instant, there is complete information about the DES output or, equivalently, information about the current DES state up to some block of a covering π_λ, as well as additional information about the current state up to some block of a covering φ. Then the DES state at the next discrete time instant can surely be calculated up to some block of the covering φ.

A (λ, δ)-invariant covering such that

$$\varphi^0 \leqslant \varphi, \tag{12}$$

where φ^0 is a given covering, can be found using an algorithm similar to the one presented in [15]; by assumption, the operations and operators of the algebra of coverings are computed in accordance with the rules of fuzzy logic introduced above.

Algorithm 3 (computation of a (λ, δ)-invariant covering φ).

1. Set $i = 0$ and specify φ^0.
2. Compute the covering

$$\varphi^{i+1} = \varphi^i + \mathbf{m}(\pi_\lambda \times \varphi); \tag{13}$$

if $\varphi^{i+1} = \varphi^i$, then go to step 3; otherwise, assign $i := i + 1$ and repeat step 2.

3. Assign $\varphi^{i+1} := \varphi^i$. End.

The convergence of Algorithm 3 and the property of the resulting covering are established by the following result.

Theorem 1 (convergence theorem). *There exists a finite number $i = i_{\max}$ such that the covering $\varphi = \varphi^i = \varphi^{i+1} = \varphi^{i_{\max}}$ determined by* (13) *is the smallest one satisfying conditions* (11) *and* (12).

Proof. For the case of partitions, it was provided in [9]. When using the algebra of coverings, the proof remains the same [17].

We introduce the operation

$$< L_i, \mu_{B_(\varphi,j)} >= (\min(L_{i,1}, \mu_{B_{\varphi,j}}(x_1)), \ldots, \min(L_{i,n}, \mu_{B_{\varphi,j}}(x_n))). \tag{14}$$

The result of this operation defines the row vector of membership functions of DES states for the block of the covering $\pi_\lambda \times \varphi$ obtained by intersecting the ith and jth blocks of the coverings π_λ and φ, respectively. Then, in view of (10), we have

$$\mu_{B^k_{\pi_\lambda \times \varphi^i}} =< L_i, \mu_{B_{\varphi^i_j}} > M^k_s, \tag{15}$$

where $B^k_{\pi_\lambda \times \varphi^i}$ is the subset from step 2 of Algorithm 2 when computing the operator $\mathbf{m}(\pi_\lambda \times \varphi^i)$. The relationship (15) is convenient to use when computing the fuzzy covering φ^{i+1} by formula (13).

By setting $\varphi^0 = \mathbf{0}$, we apply Algorithm 2 to find the covering φ. For each state $x_{0,j}$ of the observer A_0, we assign a block of the covering φ by using a representation in the form of a unit vector of dimensions $1 \times n_0$ with the jth element equal to 1. Here, n_0, $n_0 \leqslant n$, is the number of blocks of the covering φ, and the set X_0 of the unit vectors has a dimension not exceeding that of the state set of the original automaton A.

The equality

$$< L_i, \mu_{B_{\varphi,j}} >= 0 \tag{16}$$

indicates that the ith DES output is incompatible with the jth state of the DFSA A_0.

We will specify the transition function δ_0 of the observer by the family of matrices $M^{k,i}_0$ ($1 \leqslant k \leqslant m$, $1 \leqslant i \leqslant l$), each corresponding to a certain combination of values of the DES inputs and outputs. Next, we find

$$\mu_{B^+_{\varphi},j} =< L_i, \mu_{B_{\varphi,j}} > M^k_s. \tag{17}$$

If the inclusion

$$B^+_{\varphi,j} \subseteq B_{\varphi,t} \tag{18}$$

holds, the DES states from the block $B_{\varphi,j}$ of the covering φ under the output y_i and input u_k are transferred to those belonging to the block $B_{\varphi,t}$. Hence, the matrices $M_0^{k,i}$ can be found by computing (17) and checking the inclusion (18) for various combinations of the DES inputs and outputs: $1 \leqslant k \leqslant m$, $1 \leqslant i \leqslant l$. Such a check can be performed using the following algorithm. (The variables j and t correspond to the states of the observer A_0 before and after the transition.)

Algorithm 4 (computation of the matrices $M_0^{k,i}$).

1. Set $k = 1$.
2. Set $i = 1$.
3. Set $M_0^{k,i} = 0_{n_0 \times n_0}$.
4. Set $j = 1,\ t = 1$.
5. If (16) holds, assign $j := j + 1,\ t := 1$, and repeat step 5.
6. Find the set $B^+_{\varphi,j}$ from the row vector of the membership functions computed by (17).
7. Check the inclusion (18). If it fails, go to step 9.
8. Write "1" for the element of the matrix $M_0^{k,i}(i,t)$ standing at the intersection of the jth row and tth column. Go to step 10.
9. If $t = n_0$, assign $t := t + 1$ and return to step 7.
10. If $j = n_0$, assign $j := j + 1,\ t := 1$ and return to step 5.
11. If $i = l$, assign $i := i + 1$ and return to step 3.
12. If $k = m$, assign $k := k + 1$ and return to step 2. End.

The algorithm assumes execution of several nested loops for elementary actions on matrices; the total number of loops does not exceed lmn_0^2, which characterizes its computational complexity. Steps 1–4 specify the initial values of the variables t, j, i, and k, which control the loop execution sequence of Algorithm 4. Steps 9–12 serve to correct and control the limit values of these variables, ensuring the convergence of the algorithm. Since all computations are off-line, the algorithm can be implemented for real systems without significant computational resources, on a standard PC. The main part of the algorithm (steps 5–8) is illustrated below.

Example 4. Let the DES dynamics during fault-free operation be described by the FFSA model from Example 3. In addition, let the output function of the DES be specified by a fuzzy matrix L. Suppose that a fault $f_1 : (x_5 \to x_1)_{4,3}$ is possible in the system. As a consequence, the matrix M^3 takes the form of M_1^3. Here are the newly introduced matrices for the DES with the fault:

$$L = \begin{pmatrix} 1 & 0.5 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad M_1^3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0.1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

(The matrices M^1 and M^2 retain their previous form.)

For the FFSA model under consideration, we apply Algorithm 3 to find the (λ, δ)-invariant covering for $\varphi^0 = 0$. The operator $\mathbf{m}$ will be computed in accordance with Algorithm 2. By setting $i = 0$ and using (13), we obtain

$$\varphi^1 = \mathbf{0} + \mathbf{m}(\lambda \times \mathbf{0}) = \mathbf{m}(\mathbf{0}) = \{B^{***}; B^{*}; B^{2}_{\sigma,2}; B^{**}\} = \{(x_1, x_5); (x_2, x_3); (x_3, x_4); (x_3, x_5)\},$$

where $\mu_{B^{***}} = (1\ 0\ 0\ 0\ 1)$. Continuing computations based on (13), one can show that $\varphi^2 = \varphi^1 + \mathbf{m}(\pi_\lambda \times \varphi^1) = \varphi^1$. Following step 3 of Algorithm 3, we set $\varphi = \varphi^1$.

For the sake of convenience, we begin the illustration of Algorithm 4 by presenting some intermediate results. The table below demonstrates the results of checking the inclusion (18) (steps 5–8 of the algorithm).

The results of Algorithm 4 (steps 1–7)

$B_{\varphi 1} \cap B_{\pi_\lambda 1}$	M^1	M^2	M^3	$B_{\varphi 1} \cap B_{\pi_\lambda 2}$	M^1	M^2	M^3	$B_{\varphi 1} \cap B_{\pi_\lambda 3}$	M^1	M^2	M^3
$B^+_{\varphi 1} \subseteq$	$B_{\varphi 2}$	$\emptyset$	$\emptyset$	$B^+_{\varphi 1} \subseteq$	$\emptyset$	$\emptyset$	$B_{\varphi 1}$	$B^+_{\varphi 1} \subseteq$	$\emptyset$	$\emptyset$	$\emptyset$
$B_{\varphi 2} \cap B_{\pi_\lambda 1}$	M^1	M^2	M^3	$B_{\varphi 2} \cap B_{\pi_\lambda 2}$	M^1	M^2	M^3	$B_{\varphi 2} \cap B_{\pi_\lambda 3}$	M^1	M^2	M^3
$B^+_{\varphi 2} \subseteq$	$\emptyset$	$B_{\varphi 3}$	$\emptyset$	$B^+_{\varphi 2} \subseteq$	$\emptyset$	$B_{\varphi 3}$	$\emptyset$	$B^+_{\varphi 2} \subseteq$	$\emptyset$	$B_{\varphi 2}$	$B_{\varphi 4}$
$B_{\varphi 3} \cap B_{\pi_\lambda 1}$	M^1	M^2	M^3	$B_{\varphi 3} \cap B_{\pi_\lambda 2}$	M^1	M^2	M^3	$B_{\varphi 3} \cap B_{\pi_\lambda 3}$	M^1	M^2	M^3
$B^+_{\varphi 3} \subseteq$	$\emptyset$	$\emptyset$	$B_{\varphi 1}$	$B^+_{\varphi 3} \subseteq$	$\emptyset$	$\emptyset$	$\emptyset$	$B^+_{\varphi 3} \subseteq$	$\emptyset$	$B_{\varphi 2}$	$B_{\varphi 4}$
$B_{\varphi 4} \cap B_{\pi_\lambda 1}$	M^1	M^2	M^3	$B_{\varphi 4} \cap B_{\pi_\lambda 2}$	M^1	M^2	M^3	$B_{\varphi 4} \cap B_{\pi_\lambda 3}$	M^1	M^2	M^3
$B^+_{\varphi 4} \subseteq$	$\emptyset$	$\emptyset$	$\emptyset$	$B^+_{\varphi 4} \subseteq B_{\varphi 2}$	$\emptyset$	$B_{\varphi 2}$	$B_{\varphi 1}$	$B^+_{\varphi 4} \subseteq$	$\emptyset$	$B_{\varphi 2}$	$B_{\varphi 4}$

Horizontally, this table is split into four parts, each corresponding to a certain covering block $B_{\varphi,j}$ (the observer's state $x_{0,j}$), $1 \leqslant j \leqslant 4$, existing before the transition. Each part of the table contains two rows. The odd rows describe the arguments: combinations of the initial covering block $B_{\varphi,j}$ with the blocks of the covering π_λ (corresponding to the DES outputs) and with the matrices M^k_s used in the computations (corresponding to the DES inputs). The even rows present the blocks of the covering φ corresponding to the next states of the observer after the transition. Based on this table, we consider the process of finding the rows of matrices describing the dynamics of the observer A_0. Note that steps 1–4 of the algorithm yield $k = 1$, $i = 1$, $M_0^{1,1} = 0_{4\times 4}$, $j = 1$, and $t = 1$. Equality (16) fails, and (in accordance with step 5 of the algorithm) it is necessary to go to step 6. Computations of step 6 and a check of the inclusion (18) (step 7) lead to step 9. The subsequent return to step 7 and a check of the inclusion (18) confirm the latter's validity for $j = 1$ and $t = 2$. After executing step 8, the first row of the matrix $M_0^{1,1}$ will contain 1 in the second position. By continuing Algorithm 4 for the other values of the variable j, $2 \leqslant j \leqslant 4$, one will easily verify that the remaining rows of this matrix are 0. Thus, for a certain pair of inputs and outputs, the form of all rows of the matrix $M_0^{k,i}$ is completely defined by the table. For example, the elements of the third row of the matrix $M_0^{k,2}$ are 0 for all $k \in \{1, 2, 3\}$, while the elements of the fourth row of the matrices $M_0^{2,3}$ and $M_0^{3,3}$ contain 1 in the second and fourth positions, respectively.

6. FORMATION OF THE CONTROL LAW

Consider the following terminal control problem for a DES: it is required to transfer the system from an initial state $x(0) \in X$ to a certain final state x^r, $x^r \in X$. Note that the solution of this problem cannot be guaranteed based on a single reference trajectory due to an uncertain response of the DES to the same input actions. In other words, one can speak of controlling the system states only up to blocks of a certain covering of the set X. As a result, in practice, the existence of several correct trajectories may be required, with the possibility of transition between them depending on the DES response. The problem of forming correct alternative trajectories is closely related to the specifics of a particular system and will not be considered in detail here.

If a fault f_s is diagnosed in the DES, the mentioned covering is a covering φ [9]. Hence, the original terminal control problem is replaced by that of transferring the system state from a block B^0_φ of the covering φ, $x(0) \in B^0_\varphi$, to a block B^r_φ of the covering φ, $x^r \in B^r_\varphi$. Next, it is necessary to form the required trajectory in terms of blocks of the covering φ, representing a sequence of transitions: $B^0_\varphi \to B^1_\varphi \to \cdots \to B^r_\varphi$, where at least one state of the subsequent block is reachable from at least one state of the previous block. Note that the specified sequence will be realized regardless of the errors in DES transitions caused by the fault f_s since the resulting distortions of its state lie within the same block of the covering φ. Also, we emphasize that this sequence bijectively corresponds to a certain sequence of transitions of the observer $A_0 : x^0_0 \to x^1_0 \to \cdots \to x^r_0$.

Following [9], we will associate the value of the generated new input u_* with the current value of the transition function δ_0 of the observer or, equivalently, with the block $B_{\varphi,t}$ of the covering φ

whose states are reachable from the states of the block $B_{\varphi,j}$ (see (17) and (18)). For example, if the above chain of transitions contains a transition $B^i_\varphi \to B^{i+1}_\varphi$, $1 \leqslant i \leqslant r-1$, then the block B^i_φ here can be treated as the block $B_{\varphi,j}$, and B^{i+1}_φ as $B_{\varphi,t}$. Consequently, the new input u_* is completely defined by the desired next block of the covering φ in the chain of transitions.

Consider now the rule for forming the mapping g. This rule is based on analysis of the transition function δ_0 of the observer. The problem here is that for the same pair of combinations of the DES output y and the observer's state x^i_0 associated with the block B^i_φ of the covering φ, there may exist several inputs implementing the required transition of the DES to the state $x_* \in X$. To eliminate the arising ambiguity, we choose an input ensuring the highest degree of confidence in the transition of the system to the specified state x_* in accordance with the criterion

$$J = \max_u \mu_{B^{i+1}_\varphi}(x_*). \tag{19}$$

Example 5. We return to the DES with the fault f_1 from Examples 3 and 4. Let the terminal control problem consist in sequentially transferring this system from the initial state $x(0) = x_1 \in B_{\varphi,1}$ to the final state $x^r = x_3$ by realizing the chain of transitions $x_1 \to x_2 \to x_3$. We associate this chain with a sequence of blocks of the covering φ. Since $x_2 \in B_{\varphi,2}$, the only correct transition is ensured by the DES input u_1. This can be verified using the table as a basis for forming the mapping g. However, if the output y_3 is formed instead, then the transition is implemented to the state x_3 rather than to the state x_2. In other words, the originally specified trajectory (chain of DES transitions) is violated, and a transition to another correct trajectory is needed.

If another DES output is formed (not y_3), then the transition to the state x_2 can be ensured. In this case, according to the table, the input u_2 will implement the next transition to the block $B_{\varphi,3}$, containing the state x_3. The formation of the output y_3 will then confirm the successful transition.

Let us illustrate the optimization approach based on the criterion (19). Suppose now that $x_3 \in B_{\varphi,2}$, the DES output is y_3, and while realizing the alternative trajectory $x_1 \to x_3 \to x_3$, it is required to preserve the same state ($x_* = x_3$) with the greatest confidence. According to the table, the state can be preserved with the inputs u_2 or u_3. After a successful transition, we have $x_3 \in B_{\varphi,2}$ or $x_3 \in B_{\varphi,4}$, respectively. However, in the latter case, the value of the criterion J is greater than in the former: $J = \mu_{B_{\varphi,4}}(x_3) = 0.9$ vs. $J = \mu_{B_{\varphi,2}}(x_3) = 0.6$. Therefore, the choice u_3 is preferable.

7. AN ILLUSTRATIVE EXAMPLE

Consider the process of change management in an IT system [17, Table 1]. This process is one of the most important for IT systems: it is responsible for managing the life cycle of all changes and facilitates the implementation of useful changes with minimal interruption of IT services. The agents of the process are:

1) the initiator (an IT department representative who performs the initial processing, assignment, and control of changes);

2) the executor (an engineer who makes changes to configuration elements or coordinates the contractor's work on these changes);

3) the Advisory Change Committee (ACC, an advisory body that meets regularly to assess and plan changes);

4) the process manager (an IT department representative who controls the change management process and forms suggestions for its improvement).

In [14], the fuzzy finite state automaton model of a process was introduced, and a solution of the fault diagnosis problem of the process was proposed. Omitting the detailed description of the model itself, we present the list of the variables used. The following external events are considered to be model inputs: u_1 (work plan completion), u_2 (plan approval by the ACC), u_3 (transfer of the

non-approved plan for revision), u_4 (transfer of a task and a work plan to the executor), u_5 (work completion), and u_6 (entering of the changes made in the IT system library). The following stages of the regulations are considered to be model states: x_1 (formation of a task and an implementation plan), x_2 (coordination of a task and a work plan by the ACC), x_3 (coordination of a task and a work plan by the process manager), x_4 (carrying out works by the executor), x_5 (check of the carried out works by the process manager), and x_6 (completed works). Available information about some stages of the regulations is used as outputs. The model allows noncritical deviations in the actions of the process agents from the original (reference) regulations. The initiator's error $f_1 : (x_2 \to x_3)_{x_1,u_1}$ is considered. The fuzzy covering φ is specified by the membership functions $\mu_{B^0_\varphi} = (1\ 0\ 0\ 0\ 0\ 0)$, $\mu_{B^1_\varphi} = (0\ 1\ 1\ 0\ 0\ 0)$, $\mu_{B^2_\varphi} = (0.1\ 0.1\ 0\ 0.8\ 0\ 0)$, $\mu_{B^3_\varphi} = (0.2\ 0.1\ 0\ 0\ 0.7\ 0)$, and $\mu_{B^4_\varphi} = (0\ 0\ 0\ 0\ 0\ 1)$. Recall that the blocks of the covering correspond to the observer's states x^0_0, x^1_0, x^2_0, x^3_0, and x^4_0.

The reference trajectory has the form $x_1 \to x_2 \to x_3 \to x_4 \to x_5 \to x_6$. This trajectory corresponds to the following sequence of transitions between the blocks of the covering $\varphi : B^0_\varphi \to B^1_\varphi \to B^1_\varphi \to B^2_\varphi \to B^3_\varphi \to B^4_\varphi$, initiated by the sequence of inputs u_1, u_2, u_4, u_5, u_6 and leading to the sequence of outputs $y_1, y_2, y_3, y_4, y_2, y_5$ (see [14, Table 4]).

The trajectory $x_1 \to x_2 \to x_3 \to x_2 \to x_3 \to x_4 \to x_5 \to x_6$ is also correct: the process manager transfers the task and work plan for repeated coordination to the ACC. It corresponds to the following sequence of transitions between the blocks of the covering $B^\varphi : B^0_\varphi \to B^1_\varphi \to B^1_\varphi \to B^1_\varphi \to B^1_\varphi \to B^2_\varphi \to B^3_\varphi \to B^4_\varphi$, initiated by the sequence of inputs u_1, u_2, u_4, u_2, u_4, u_5, u_6 and leading to the sequence of outputs y_1, y_2, y_3, y_2, y_3, y_4, y_2, y_5 (see [14, Table 5]).

The trajectory $x_1 \to x_3 \to x_4 \to x_5 \to x_6$ is incorrect, caused by an error of the process initiator. It corresponds to the following sequence of transitions between the blocks of the covering φ: $B^0_\varphi : B^1_\varphi \to B^2_\varphi \to B^3_\varphi \to B^4_\varphi$, initiated by the sequence of inputs u_1, u_2, u_4, u_5 and leading to the sequence of outputs y_1, y_3, y_4, y_2, y_5 (see [14, Table 6]). If this error is detected, it is proposed to make a transition to the reference trajectory, starting from the state $x_2, x_2 \in B^1_\varphi$, using the input u_4.

The real-time formation algorithm for the new input u_* and the mapping g ensuring accommodation to the process initiator's error can be obtained based on the results of this paper. It is omitted here because of simplicity and page limits.

8. CONCLUSIONS

In this paper, the following results have been obtained when solving the general fault accommodation problem for DESs.

1. The mathematical apparatus of the pair algebra of coverings previously proposed by the authors [17] has been extended to the case of covering blocks specified by fuzzy sets. The pair algebra of fuzzy coverings has been elaborated. The necessary computations of the operations and operators of the algebra can be supported by existing applied software packages, e.g., MATLAB.

2. The concept of an observer as applied to an FFSA has been introduced. New mathematical constructs of the pair algebra of fuzzy coverings have been used to develop an algorithm for designing an FFSA observer (Algorithm 4).

3. An optimization approach has been proposed to find the new input u_*, mapping g, and input u of the DES in order to ensure a high degree of implementability of the fault-tolerant control law.

FUNDING

This work was supported by the Russian Science Foundation, project no. 25-19-00590, https://rscf.ru/project/25-19-00590/.

REFERENCES

1. Blanke, M., Kinnaert, M., Lunze, J., and Staroswiecki, M., *Diagnosis and Fault-Tolerant Control*, Berlin: Springer-Verlag, 2003.
2. Edwards, C., Spurgeon, S.K., and Patton, R.J., Sliding Mode Observers for Fault Detection and Isolation, *Automatica*, 2000, vol. 36, no. 4, pp. 541–553.
3. Feng, W., Li, M., Zhang, W., et al., Sliding Mode Fault-Tolerant Control Based on Fast Convergence Law for Faults of PMSM, *Chinese Journal of Electrical Engineering*, 2024, vol. 10, no. 3, pp. 147–157.
4. Zhirabok, A.N. and Shumsky, A.E., Fault Accommodation in Nonlinear Dynamic Systems, *Journal of Computer and Systems Sciences International*, 2009, vol. 48, no. 4, pp. 559–566.
5. Zhirabok, A.N., Kotta, Ü., and Shumsky, A.E., Accommodation to Defects in the Discrete Dynamic Systems, *Automation and Remote Control*, 2014, vol. 75, no. 6, pp. 977–990.
6. Wonham, W.M., Cai, K., and Rudie, K., Supervisory Control of Discrete-Event Systems: A Brief History — 1980–2015, *IFAC-PapersOnLine*, 2017, vol. 50, no. 1, pp. 1791–1797.
7. Cassandras, Ch. and Lafortune, S., *Introduction to Discrete Event Systems*, 2nd ed., New York: Springer Science+Business Media, 2008.
8. Zaytoon, J. and Lafortune, S., Overview of Fault Diagnosis Methods for Discrete Event Systems, *Annual Reviews in Control*, 2013, vol. 37, no. 2, pp. 308–320.
9. Shumskii, A.E. and Zhirabok, A.N., Method of Accommodation to the Defects in Finite Automata, *Automation and Remote Control*, 2010, vol. 71, no. 5, pp. 837–846.
10. Hartmanis, J. and Stearns, R., *Algebraic Structure Theory of Sequential Machines*, Englewood Cliffs, New Jersey: Prentice-Hall, 1966.
11. Zadeh, L.A., Fu, K.S., Tanaka, K., and Shimura, M., *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, New York: Academic Press, 1975.
12. Wee, W.G. and Fu, K.S., A Formulation of Fuzzy Automata and Its Application as a Model of Learning Systems, *IEEE Trans. Syst. Science and Cybernetics*, 1969, vol. SSC-5, no. 3, pp. 215–223.
13. Ulzutuev, I.E. and Maximov, A.A., The Properties of Subautomata Lattices of Fuzzy Semiautomata and Their Determinizators, *Vestnik of Saratov State Technical University*, 2015, vol. 79, no. 2, pp. 117–126.
14. Shumsky, A.E. and Zhirabok, A.N., A Fault Diagnosis Method for Discrete-Event Systems Based on the Fuzzy Finite State Automaton Model, *Control Sciences*, 2025, no. 2, pp. 31–41.
15. Borisov, V.V., Kruglov, V.V., and Fedulov, A.S., *Nechetkie modeli i seti* (Fuzzy Models and Networks), 2nd ed., Moscow: Goryachaya Liniya – Telekom, 2012.
16. Voevodin, V.V. and Kuznetsov, Yu.A., *Matritsy i vychisleniya* (Matrices and Calculations), Moscow: Nauka, 1984.
17. Zhirabok, A.N., Kalinina, N.A., and Shumskii, A.E., Method for the Functional Diagnosis of Non-Deterministic Finite Automata, *Journal of Computer and Systems Sciences International*, 2020, vol. 59, no. 4, pp. 565–574.

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

ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 10, pp. 940–951.
© 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. 10, pp. 104–118.

Defense Planning for a Moving Target Against Multiple Attackers Using a Defender via the Monte Carlo Tree Search

E. A. Ryabushev[*,a] **and A. A. Galyaev**[*,b]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a] *forteamofdota@mail.ru,* [b] *galaev@ipu.ru*

Received June 24, 2026
Revised July 6, 2026
Accepted July 8, 2026

Abstract—The problem of protecting a moving target from an attack by multiple pursuing attackers using an interceptor defender is considered. A rigorous mathematical formulation of the problem is presented, quality criteria for defense plans are introduced, and mathematical properties of the problem are formulated. Based on these properties, heuristic rules and algorithms for their application are proposed in order to construct suboptimal solutions.

Keywords: mathematical optimization, operations research, Monte Carlo tree search (MCTS), local rules

DOI: 10.7868/S1608303226100055

1. INTRODUCTION

The world is undergoing major changes. Drones dominate on land and in the air, while unmanned surface vessels and underwater vehicles pose a danger at sea. Autonomous systems are capable of solving an increasing number of both military and civilian tasks without any human participation. This is most noticeable on the battlefield, where unmanned vehicles have already changed the nature of war and define new rules. Massed drone attacks on infrastructure and logistics facilities represent a significant and urgent problem. It is becoming obvious that economically feasible protection against this threat requires interceptors capable of sequentially destroying attacking drones one after another. An important role here is played by the choice of the best interception order, which is a problem close in meaning to the "dynamic traveling salesman problem" (DTSP) [1, 2] or Moving Targets TSP (MTTSP) [3, 4]. At present, there are no effective methods for the exact solution of DTSP, and the problem itself is at the initial stages of investigation [5]. Modern formulations of DTSP concentrate on conditions of uncertainty in information about the positions of targets and limited time windows for their observation [6, 7].

The traveling salesman problem and, as a consequence, many of its dynamic generalizations are NP-complete [8], which means that there are no effective exact solution algorithms for them. Building a defense requires decision-making under limited time, which necessitates the development of methods for fast suboptimal planning. Algorithms of this kind were already considered by the authors for the case of defending a stationary point in [9]; the goal of the present work is to generalize them to the problem of defending a moving point. For this purpose, a generalized formulation of the problem is proposed that includes both planning the order of visiting targets and constructing an evasion maneuver from attackers threatening the protected moving point. It is proposed to evaluate the quality of plans on the basis of vector criteria generalizing those from [10] for the defense of a static point to the dynamic case. The motion of the moving defense object is planned

on the basis of local rules similar to those used in [11], and the interception order of the attackers is chosen by means of Monte Carlo tree search (MCTS) [12, 13].

MCTS is a combinatorial optimization method that became known after its successful application in the AlphaGo project [12]. Classical MCTS applications assumed game formulations in finite games, but the algorithm has also proven itself for combinatorial optimization in non-game problems [14]. Monte Carlo tree search with PUCT estimates (Predictor + Upper Confidence bounds for Trees) is usually applied together with a neural network and is used for its training. However, MCTS is also capable of operating on the basis of local rules that can evaluate actions instead of a neural network [9]. Such an implementation of the algorithm not only allows it to be applied without preliminary training, but also makes it possible to use it for computing q-estimates for reinforcement learning. Thus, the adaptation of Monte Carlo tree search not only provides a ready-made algorithm for solving the planning problem, but also opens a path for effective training of neural-network models capable of solving this problem.

2. MATHEMATICAL MODEL, PROBLEM STATEMENT, AND SOLUTION METHOD

This section discusses the formulation of the problem of protecting a moving target by an interceptor defender against an attack by multiple attackers, which generalizes the problem formulations from [9, 10] to the case of a moving defense object.

2.1. Motion Model of the Protected Point, Attacking Targets, and Interceptor Defender

Suppose it is necessary to protect a moving target T, capable of moving with speed V_T, from an attack by n attackers that chase T with speed $V_A > V_T$. We assume that T must be defended by an interceptor defender D, which can pursue the attackers with speed $V_D > V_A$. Motion occurs in a plane; the target T and the defender D are located at the origin at the initial time, and the attackers are on the boundary of a circle of radius R in a layer of thickness $2\Delta R$.

The initial positions of the attackers are specified by radius vectors $\mathbf{r}_1^0, \ldots, \mathbf{r}_n^0$, and their velocities $\mathbf{V}_1, \ldots, \mathbf{V}_n$ are constantly directed toward T:

$$\mathbf{V}_i = V_A \frac{\mathbf{r}_T - \mathbf{r}_i}{|\mathbf{r}_T - \mathbf{r}_i|}, \tag{1}$$

where $\mathbf{r}_i$ and $\mathbf{r}_T$ are the current positions of the ith attacker and the target T. An attacker is considered missed if at some time its coordinates coincide with the position of T.

The dynamics of the protected target T is described by the simple motion model

$$\begin{aligned} \dot{x}_T &= V_T \cos(u), \\ \dot{y}_T &= V_T \sin(u), \end{aligned} \tag{2}$$

where x_T and y_T are the coordinates of T, and $u \in [0, 2\pi]$ is the control of the direction of motion.

The velocity vector of the defender D is determined by the attacker that D is intercepting at the given time:

$$\mathbf{V}_D = V_D \frac{\mathbf{r}_i - \mathbf{r}_D}{|\mathbf{r}_D - \mathbf{r}_i|}, \tag{3}$$

where $\mathbf{r}_D$ and $\mathbf{r}_i$ are the positions of D and of the attacker whose interception D is moving toward. An attacker is considered destroyed if at some time its coordinates coincide with the position of D.

2.2. Concept of a Defense Plan, Its Quality Criteria, and Problem Statement

An interception plan $\pi = [\pi_1, \ldots, \pi_n]$ is a sequence of targets, where π_i is the number of the target that D visits in the ith turn. Let $\tau^*(\pi)$ be the execution time of π; then on $[0, \tau^*(\pi)]$ it is required to construct a control $u(t)$ of the motion of T according to (2). To simplify the problem, it is proposed to specify the motion of the point T in discrete time with step Δt. We introduce the following

Definition 1. A defense plan Π is a pair (π, ρ), where π is the sequence of target interceptions by the defender D, and $\rho = \{\mathbf{r}_T(t_k)\}_{k=0}^K$ is the set of points that T occupies at times $t_k = k\Delta T$. Moreover, the following constraints are imposed on ρ:

$$|\mathbf{r}_T(t_{k+1}) - \mathbf{r}_T(t_k)| = V_T \Delta t,$$

which express the reachability conditions for the points ρ; without loss of generality, we always assume that $\mathbf{r}_T(0) = \mathbf{r}_D(0) = (0, 0)$.

We introduce the loss vector $L(\Pi) = (m_0, \ldots, m_5)$, where m_0 is the number of missed targets, and m_i is the number of targets intercepted at a distance of at least R_i from T, where $R_1 < R_2 < \cdots < R_5$. Then we formulate the problem.

Problem 1. Given the initial positions $\mathbf{r}_1^0, \ldots, \mathbf{r}_n^0$ of the attacking targets, construct a defense plan Π according to Definition 1 with the smallest possible loss vector $L(\Pi)$, where comparison is performed in lexicographic order, i.e.,

$$(m_0, \ldots, m_5) < (m_0', \ldots, m_5') \to \exists k : \forall i < k : m_i = m_i' \wedge m_k < m_k'.$$

In practice, finding an optimal plan Π is practically impossible, and it is required to find a suboptimal Π with an acceptable $L(\Pi)$. In [10], the following is proved

Theorem 1. *For any initial arrangement of attackers, in the presence of a dynamic advantage $(V_D > V_A)$ of the defender over the attackers, for a stationary protected target there exists an interception plan π that ensures interception of all attackers without misses.*

Corollary 1. *For any initial arrangement of attackers, in the presence of a dynamic advantage $(V_D > V_A)$ of the defender over the attackers, for a moving protected target there exists an interception plan π that ensures interception of all attackers without misses.*

Proof. Indeed, since a dynamic target is able to remain stationary, its ability to move cannot worsen the result of Theorem 1, and therefore there always exists a defense plan without missed attackers.

Thus, a reasonable requirement for the sub-optimality of a defense plan is $m_0 = 0$, i.e., the absence of missed targets.

2.3. Positional Properties of Attackers and Local Decision Rules

Suppose that at time t the defender D and the protected target T occupy positions $\mathbf{r}_I(t)$ and $\mathbf{r}_T(t)$, $\mathbf{V}_T(t)$ is the velocity of T, and $\mathbf{r}_i(t)$ are the radius vectors of the attackers. For further discussion, the following quantities are required:

$$\tau_i(t) = \frac{V_A|\mathbf{r}_T(t) - \mathbf{r}_i(t)| + \mathbf{V}_T(t)[\mathbf{r}_T(t) - \mathbf{r}_i(t)]}{V_A^2 - V_T^2},$$

where $\tau_i(t)$ is the time required by the ith attacker to hit T, provided that T continues moving in a straight line with velocity $\mathbf{V}_T$, and

$$\theta_i(t) = \frac{V_I|\mathbf{r}_T(t) - \mathbf{r}_i(t)| + \mathbf{V}_i(t)[\mathbf{r}_T(t) - \mathbf{r}_i(t)]}{V_I^2 - V_A^2}, \tag{4}$$

where $\theta_i(t)$ is the time required by the defender D to intercept the ith attacker, provided that the latter moves in a straight line with velocity $\mathbf{V}_i$ determined according to (1) from the current values of $\mathbf{r}_i(t)$ and $\mathbf{r}_T(t)$.

On the basis of $\tau_i(t)$ and $\theta_i(t)$, we introduce the following definitions.

Definition 2. The danger of the ith attacker is understood as the quantity

$$K_i(t) = \frac{1}{\tau_i(t)},$$

the reciprocal of the hit time $\tau_i(t)$.

Danger shows how strongly the attacker threatens T: the less time it needs to hit T, the higher its danger.

Definition 3. The convenience of the ith attacker is understood as the quantity

$$U_i(t) = \frac{1}{\theta_i(t)},$$

the reciprocal of the interception time $\theta_i(t)$.

Convenience shows how easy it is for the defender to intercept the attacker: the less time required for this, the higher its convenience.

Definition 4. The advantage over the ith attacker is understood as the quantity

$$A_i(t) = \tau_i(t) - \theta_i(t),$$

the difference between the hit time $\tau_i(t)$ and the interception time $\theta_i(t)$.

Advantage shows the time reserve for interception: the higher the advantage, the longer the decision on assigning this target can be postponed.

When planning the motion of T, it is reasonable to primarily evade attackers that pose the greatest threat in order to gain time for their interception by D. The degree of threat from the ith attacker is estimated by its danger $K_i(t)$; therefore, to find the velocity vector of the target, it is logical to use the rule

$$\mathbf{V}_T(t) \sim \sum_{i=1}^{n} \frac{K_i(t)}{\sum\limits_{j=1}^{n} K_j(t)} \frac{\mathbf{r}_T(t) - \mathbf{r}_i(t)}{|\mathbf{r}_i(t) - \mathbf{r}_T(t)|}, \tag{5}$$

where the proportionality sign means that the quantity on the right determines the direction of the velocity $\mathbf{V}_T(t)$, while its magnitude coincides with V_T. This rule corresponds to priority escape from attackers with the greatest danger $K_i(t)$. An important caveat should be made here: the use of this rule deprives Corollary 1 of force, since now the protected target cannot stop moving, which means that there may be initial conditions under which no interception plan without misses exists, but such configurations are extremely rare.

In addition, the instantaneous characteristics of the attackers make it possible to formulate local rules for the defender D:

$$P_Q(i) = \frac{Q_i(t)}{\sum\limits_{j=1}^{n} Q_j(t)}, \tag{6}$$

where instead of $Q_i(t)$ one substitutes the danger $K_i(t)$ or the convenience $U_i(t)$ of the ith target. The quantities $P_Q(i)$ can be used both for deterministic assignment of attackers to the defender by $\operatorname{argmax} P_Q(i)$ and for constructing a stochastic policy, where $P_Q(i)$ is treated as the probability distribution of assignments.

2.4. Concept of a Partial Plan and Properties of the Planning Problem

When rule (5) is used for the motion of the target T, its trajectory is completely determined by the initial positions and the interception plan π. For $\pi = [\pi_1 \dots \pi_n]$, we introduce the sequence of times $\mathbf{t} = [t_1 \dots t_n]$ required by D to service the first k attackers from π. Here, servicing an attacker means either its successful interception or its hitting T before capture. Then the motion of T up to time t_k depends only on the first k elements of the plan π, which motivates the following definitions.

Definition 5. A partial plan $\pi = [\pi_1, \dots, \pi_k]$ is an ordered set of $k \leqslant n$ distinct targets specifying the order of their interception, to which the intermediate state $S(\pi)$ corresponds. By $S(\pi)$ we mean the coordinates $\mathbf{r}_T(t_k)$, $\mathbf{r}_D(t_k)$, and $\mathbf{r}_j(t_k)$, where t_k is the execution time of the partial plan π, $\mathbf{r}_j(t_k)$ are the positions of those attackers that were not intercepted by D and have not yet hit T by the time π is completed, together with the current loss vector $L(\pi) = (m_0(t_k), \dots, m_5(t_k)$, where $m_0(t_k)$ is the number of missed targets, and $m_i(t_k)$ are the numbers of attackers intercepted in the corresponding zone by time t_k.

For a partial plan, we introduce the estimate $l(\pi) = L(\pi) + Z(S)$, where $L(\pi)$ is the current loss vector of the plan π, and $Z(S) = (0, z_1, \dots, z_5)$, where z_i is the number of targets located in the corresponding loss zone in the intermediate state $S(\pi)$. Then the following lower-bound theorem holds.

Theorem 2. *The final loss vector $L(\pi^*)$ for any continuation of the plan π is always no less than $l(\pi)$, i.e., $L(\pi^*) \geqslant l(\pi)$, if π^* is a continuation of π.*

Proof. Indeed, since the speed $V_A > V_T$, the distance $d = |\mathbf{r}_i(t) - \mathbf{r}_T(t)|$ between the ith attacker and the target T decreases monotonically with increasing time t. Therefore, the loss vector $L(\pi^*)$ of any continuation of the partial plan π will include both the loss vector $L(\pi)$ and the losses for not yet intercepted attackers, which cannot be stopped at a distance greater than the current distance between them and T.

Corollary 2. *If for some partial plan π the loss estimate $l(\pi) \geqslant L(\pi^*)$ exceeds the losses for some complete plan π^*, then none of the continuations of π can be optimal.*

Corollary 3. *If for some partial plan π it is true that the loss estimates $l(\pi')$ of all its continuations of fixed length $\mathrm{len}(\pi') = k$ exceed the estimate of some complete plan π^* (i.e., $\forall \pi' : \mathrm{len}(\pi') = k \to l(\pi') \geqslant L(\pi^*)$), then none of the continuations of π can be optimal.*

These corollaries make it possible to reject obviously nonoptimal search directions when searching for an optimal or suboptimal solution. Moreover, using the fact that the existence of a plan that misses no attacker is known in advance, the vector $l^* = [0, n, 0, 0, 0]$, corresponding to the worst estimate without misses, can be used as the baseline value of the inadmissible estimate $l(\pi)$.

In addition, the interception plan π can be characterized by a set of local properties of its intermediate states. We introduce the definitions.

Definition 6. The complexity of the plan π is understood as the maximum time

$$\theta(\pi) = \max_{k=1\dots n} \theta(t_k)$$

of interception of an individual target within this plan. Here the quantities $\theta(t_k)$ are understood in the sense of (4) with respect to the attacker that was intercepted at time t_k from the vector $\mathbf{t}(\pi)$.

Definition 7. The reliability of the plan π is understood as the minimum advantage

$$A(\pi) = \min_{k=1\ldots n} A(t_k)$$

over the target from the plan π. Here $A(t_k)$ is understood in the sense of Definition 4 with respect to the attacker that was intercepted at time t_k from the vector $\mathbf{t}(\pi)$.

The reliability and complexity of a plan are its global properties, which can be used to study the dependence of the loss function $L(\pi)$ on the characteristics of π.

3. CONSTRUCTION OF AN INTERCEPTION PLAN BY MONTE CARLO TREE SEARCH

This section discusses the algorithm for constructing suboptimal plans by the MCTS method.

3.1. Monte Carlo Tree Search Tree

We introduce the concept of a search tree.

Definition 8. A search tree Ξ is understood as a set of partial plans Π^* equipped with an incidence relation: $\pi \to \pi'$, if π' differs from π by adding exactly one target to the end of the plan. For a partial plan π, the set $C(\pi) = \{\pi' \in \Pi^* : \pi \to \pi'\}$ of all continuations incident to it is called its children. The root of the tree Ξ is the empty plan $\pi_0 = []$.

The MCTS algorithm performs an intelligent search over the tree Ξ, which is guided by the policy (6) and by the q-estimates of its nodes. Each plan π is marked with the number p_π of continuations constructed for this plan and with the average estimate q_π achieved on them. Here, the average q-estimate is understood as the average scalar estimate of the plans constructed from this node (the loss vector $L(\pi)$ is transformed into a scalar estimate using a linear function that assigns the estimate -1.0 to the maximum possible losses and $+1.0$ to the minimum ones). The basic loop of the algorithm consists of two procedures.

Algorithm 1. Tree descent

1. $\pi \leftarrow \pi_0$.
2. While $p_\pi > 0$: $\pi \leftarrow \arg\max\limits_{\pi':\pi\to\pi'} q_{\pi'} + \dfrac{P_Q(\pi')\sqrt{n_\pi}}{1+n_{\pi'}}$.
3. Return π.

For the found node π, a random continuation π^* is constructed according to the local rule (6) and the loss vector $L(\pi^*)$ is computed, on the basis of which the q-estimate of the plan π is formulated and then propagated up to the root.

Algorithm 2. Backpropagation of the estimate.

1. $r \leftarrow L_q(L(\pi^*))$, where $L_q(L(\pi^*))$ is the q-estimate of the loss vector $L(\pi^*)$.
2. While $\pi = [\pi_1 \ldots \pi_k]$ is not π_0: update p_π and q_π on the basis of r (the number of visits is increased by one, and the average q-estimate is recalculated taking into account the addition of r), $\pi \leftarrow \pi' = [\pi_1 \ldots \pi_{k-1}]$.

Thus, this algorithm updates the statistics accumulated over the tree nodes taking into account the new solution π^*, which was constructed from the tree node selected according to the tree descent Algorithm 1. The general scheme of the algorithm is shown in Fig. 1; more details can be found in [9].

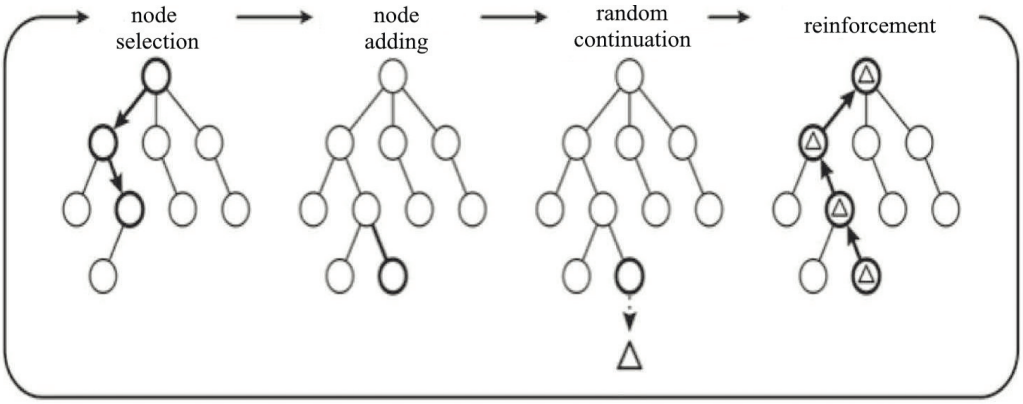


Fig. 1. Principal scheme of the Monte Carlo tree search algorithm. The algorithm consists of alternating two phases: descent through the tree to a new node and backpropagation of the estimate of this node back to the root.

3.2. Construction of Random Continuations

In the previous subsection, it was discussed that the MCTS algorithm relies on random continuations of partial plans to obtain estimates of tree nodes. This subsection describes in detail the methodology for constructing such continuations and the method for computing q-estimates.

Suppose it is required to evaluate a tree node with a partial plan π, to which the intermediate state $S(\pi)$ corresponds. A random continuation is constructed as a solution to the planning problem from the state $S(\pi)$ on the basis of the local rules (5) and (6), where rule (6) is used as a stochastic policy for assigning the defender to intercept attackers.

Algorithm 3. Construction of a random continuation.

1. $\pi^* \leftarrow \pi$.

2. While there are free targets (i.e., not intercepted and not missed): $\pi^* \leftarrow i$, which is chosen according to the probability distribution $P_Q(i|S)$ for the current intermediate state $S(\pi^*)$.

Once a complete plan π^* has been constructed according to this algorithm, the loss vector $L(\pi^*)$ is computed for it. Then $L(\pi^*)$ is transformed into a scalar q-estimate in the range $[-1, 1]$ using the linear function $L_q(L(\pi^*))$, which acts according to the rule

$$L_q([n, 0, 0, 0, 0, 0]) = -1,$$
$$L_q([0, 0, 0, 0, 0, n]) = 1,$$

i.e., assigns the estimate -1 to maximum losses and $+1$ to minimum losses. This estimate is then used in the backpropagation Algorithm 2.

3.3. Pruning of Nonoptimal Nodes and Construction of the Final Plan

In addition, the results of the MCTS search algorithm can be substantially improved on the basis of Theorem 2 and its corollaries. For this purpose, the algorithm stores at each step the best complete plan $\pi_{\max}$ that was constructed during the evaluation of one of the tree nodes over the entire running time, as well as its loss vector $L(\pi_{\max})$. A tree node with a partial plan π is considered nonoptimal if its partial state is nonoptimal with respect to $L(\pi_{\max})$ according to Corollary 2, or if all of its child nodes have been recognized as nonoptimal.

During tree descent in Algorithm 1, the choice is made exclusively among nodes that are not nonoptimal. If a node is reached that must be recognized as nonoptimal, the algorithm returns to its parent node and repeats the search process taking this additional information into account.

When the computational limit is exhausted, the algorithm returns the best plan $\pi_{\max}$ that was constructed during its entire running time.

4. EXPERIMENTAL RESULTS

The Monte Carlo tree search algorithm together with the heuristics for the motion of T and the assignment of targets to D by target danger was used to construct defense plans against attacks by 10 to 30 targets. In this case, the construction of effective defense plans on average requires from 1 to 10 s of algorithm operation depending on the problem size. Examples of characteristic defense plans are given in Figs. 2 and 3.

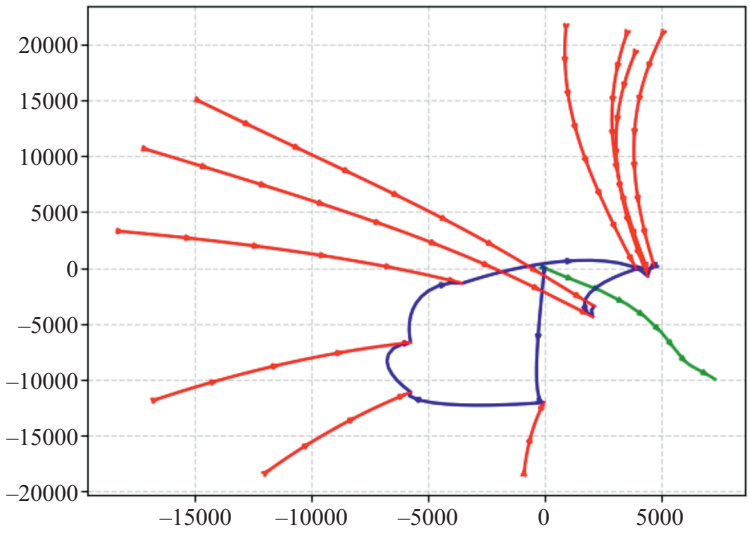


Fig. 2. Defense plan against an attack by 10 attackers. The green line shows the trajectory of the protected target, the blue line that of the interceptor defender, and the red lines those of the attackers. The arrows indicate the direction of motion and the relative arrangement of the objects on equal-time slices.

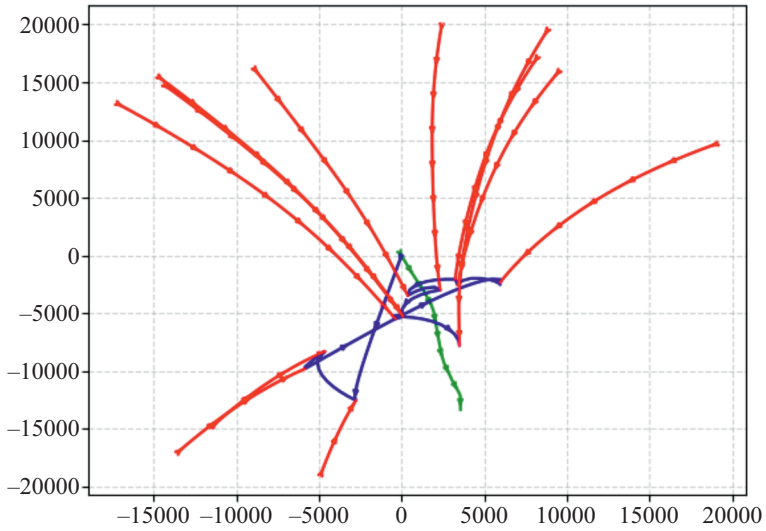


Fig. 3. Defense plan against an attack by 12 attackers. The green line shows the trajectory of the protected target, the blue line that of the interceptor defender, and the red lines those of the attackers. The arrows indicate the direction of motion and the relative arrangement of the objects on equal-time slices.

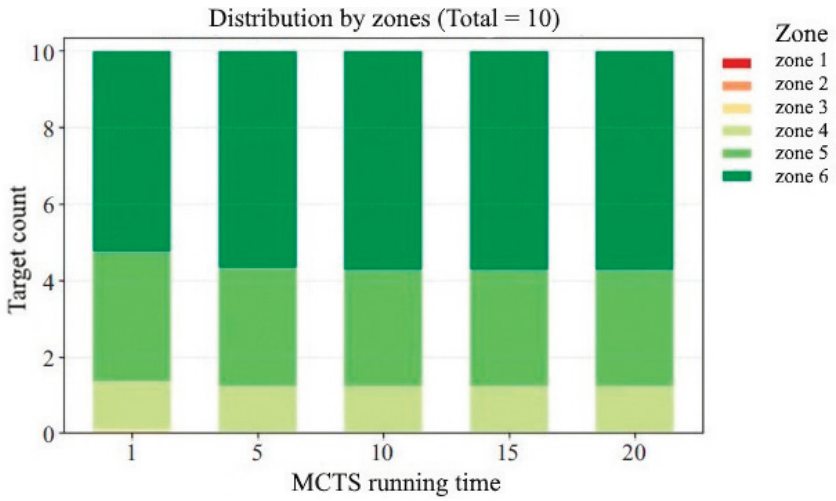


Fig. 4. Distribution of interception points by proximity zones as a function of MCTS running time for problems with $n = 10$ attackers.

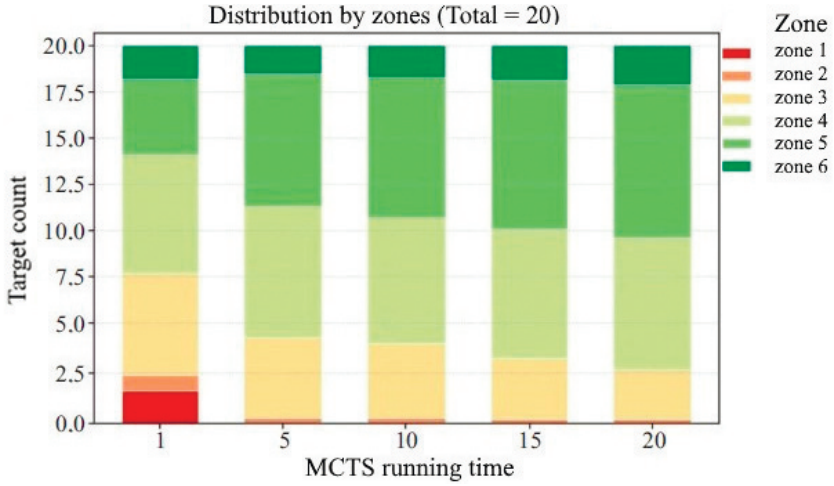


Fig. 5. Distribution of interception points by proximity zones as a function of MCTS running time for problems with $n = 20$ attackers.

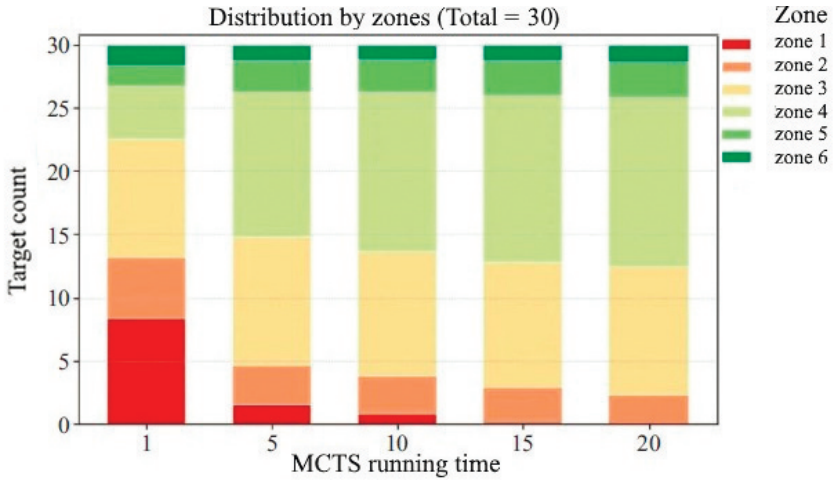


Fig. 6. Distribution of interception points by proximity zones as a function of MCTS running time for problems with $n = 30$ attackers.

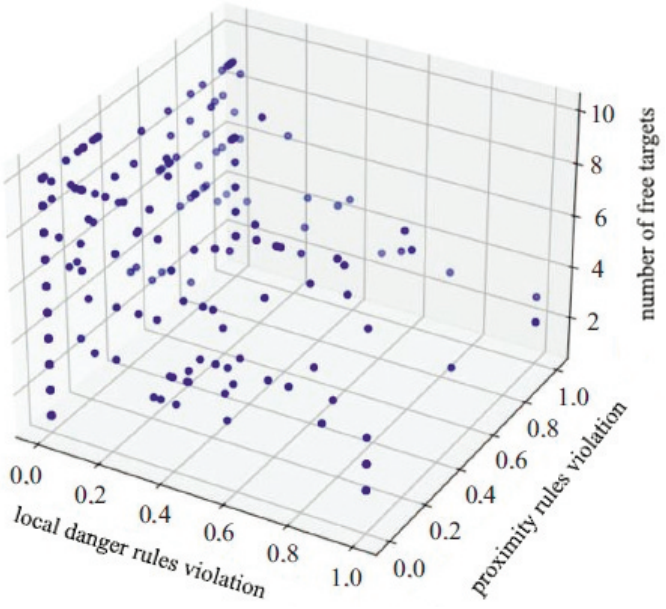


Fig. 7. Violation of the local danger and proximity rules in the optimal plan as a function of the number of free targets in the intermediate state.

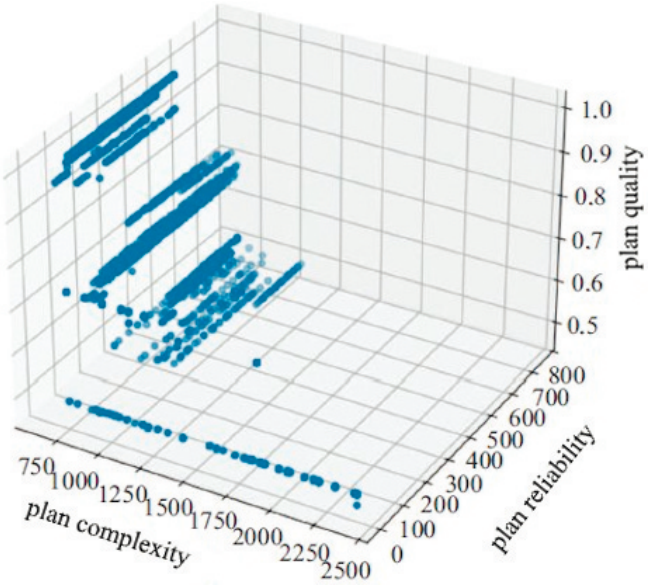


Fig. 8. Dependence of plan quality on their reliability and complexity.

The dependence of the quality of the interception plan on the algorithm running time for different numbers of attackers is given in Figs. 4–6; the data for constructing these figures were collected on a sample of 30 random initial positions with a given number of attackers. Thus, one can see the characteristic times for constructing suboptimal plans for problems of the corresponding size.

In addition, studies were conducted on the quality of the local rules (3) in comparison with the optimal plan, the results of which are given in Fig. 7. Here, the violation of a rule is understood as an estimate that varies linearly from 0 to 1, where an estimate of 0 means that the choice of target in the optimal rule is fully consistent with the heuristic, and 1 means fully inconsistent (i.e., the optimal choice corresponds to the attacker that is least attractive from the heuristic viewpoint).

Also, for completeness, we present the distribution of plans by quality depending on their global properties of complexity and reliability, constructed for the set of all admissible plans for a position with $n = 8$ attackers.

5. CONCLUSION

This work was devoted to the problem of protecting a dynamic target from an attack by multiple attackers using an interceptor defender. The main results of the work can be briefly presented in the following points.

1) The problem of protecting a dynamic target is an NP-hard problem, for which, however, effective algorithms for constructing suboptimal solutions can be proposed.

2) The Monte Carlo tree search algorithm shows good results on input examples with numbers of targets up to $n = 30$ in reasonable computational time.

3) As reference local rules for MCTS, it is logical to use heuristics based on danger and proximity, defined according to the probability distribution 6.

The MCTS algorithm constructs a higher-quality heuristic on the basis of the initial heuristics through intelligent additional search over variants. The distributions of visit numbers of child nodes at the root of the tree can be interpreted as an improved local policy, which can then be used for initial training of neural-network models, opening prospects for applying the results obtained in this work to the construction of large neural-network models for this problem. On the other hand, the possibility of effectively solving the defense problem without involving neural-network training is valuable in itself, since neural-network approaches are not always more effective than simple heuristic solvers [15, 16] in DTSP.

REFERENCES

1. Sikharulidze, G.G., On one generalization of the traveling salesman problem. I, *Autom. Remote Control*, 1971, no. 8, pp. 116–123.

2. Sikharulidze, G.G., On one generalization of the traveling salesman problem. II, *Autom. Remote Control*, 1971, no. 10, pp. 142–147.

3. Picard, J.C. and Queyranne, M., The time-dependent traveling salesman problem and its application to the tardiness problem in one-machine scheduling, *Oper. Res.*, 1978, vol. 26, no. 1, pp. 86–110. https://doi.org/10.1287/opre.26.1.86

4. Helvig, C.S., Robins, G., and Zelikovsky, A., The moving-target traveling salesman problem, *J. Algorithms Comput. Technol.*, 2003, vol. 49, no. 1, pp. 153–174. https://doi.org/10.1016/S0196-6774(03)00075-0

5. Li, C., Yang, M., and Kang, L., A new approach to solving dynamic traveling salesman problems, in *Lect. Notes Comput. Sci.*, 2006, vol. 4247, Berlin, Heidelberg, Springer.

6. Stieber, A., Dealing with time in the multiple traveling salesmen problem with moving targets, *Cottbus Math. Preprints*, 2019, COMP. 4.

7. Maskooki, A. and Kallio, M., A bi-criteria moving-target travelling salesman problem under uncertainty, *Eur. J. Oper. Res.*, 2023, vol. 309, no. 1, pp. 271–285. https://doi.org/10.1016/j.ejor.2023.01.009

8. Garey, M.R. and Johnson, D.S., *Computers and Intractability: A Guide to the Theory of NP-Completeness*, San Francisco, Calif., W. H. Freeman & Co., 1979.

9. Ryabushev, E.A. and Galyaev, A.A., Search for a suboptimal solution of the dynamic traveling salesman problem by the Monte Carlo method, *Autom. Remote Control*, 2024, no. 2, pp. 103–119.

10. Galyaev, A.A., Yakhno, V.P., Berlin, L.M., et al., Optimization of the interception plan for straight-line moving targets, *Autom. Remote Control*, 2023, no. 10, pp. 18–36.

11. Galyaev, A.A. and Ryabushev, E.A., Local dynamic planning for protecting a moving target from an attack by a high-speed interceptor with limited field of view, *Proc. Inst. Math. Mech. Ural Branch RAS*, 2026, vol. 32, no. 3.

12. Silver, D., Huang, A., Maddison, C., et al., Mastering the game of Go with deep neural networks and tree search, *Nature*, 2016, vol. 529, no. 7587, pp. 484–489. https://doi.org/10.1038/nature16961

13. Schadd, M.P.D., Winands, M.H.M., van den Herik, H.J., et al., Single-player Monte-Carlo tree search, in *Comput. Games. CG 2008. Lect. Notes Comput. Sci.*, 2008, vol. 5131, Berlin, Heidelberg, Springer.

14. Marco, K., Beyond games: A systematic review of neural Monte Carlo tree search applications, 2023. https://arxiv.org/abs/2303.08060

15. Xuanhao, P., Chenguang, W., Chaolong, Y., et al., Rethinking the heatmap + Monte Carlo tree search paradigm for solving large scale TSP, 2025. https://openreview.net/forum?id=TMHOHRR0FA

16. Xia, Y., Xianliang, Y., Zichuan, L., et al., Position: Rethinking post-hoc search-based neural approaches for solving large-scale traveling salesman problems, *Proc. 41st Int. Conf. Mach. Learn.*, 2024, vol. 235, pp. 54178–54190. https://proceedings.mlr.press/v235/xia24f.html

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