ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 9, pp. 1041–1064.

Construction of the Time-Optimal Bounded Control for Linear Discrete-Time Systems Based on the Method of Superellipsoidal Approximation

D. N. Ibragimov[*,a] **and V. M. Podgornaya**[*,b]

[*] *Moscow Aviation Institute (National Research University), Moscow, Russia*
e-mail: [a] *rikk.dan@gmail.ru,* [b] *vita1401@outlook.com*

Received April, 2023
Revised June 25, 2023
Accepted July 20, 2023

Abstract—The speed-in-action problem for a linear discrete-time system with bounded control is considered. In the case of superellipsoidal constraints on the control, the optimal control process is constructed explicitly on the basis of the discrete maximum principle. The problem of calculating the initial conditions for an adjoint system is reduced to solving a system of algebraic equations. The algorithm for generating a guaranteeing solution based on the superellipsoidal approximation method is proposed for systems with general convex control constraints. The procedure of superellipsoidal approximation is reduced to solving a number of convex programming problems. Examples are given.

Keywords: linear discrete-time systems, speed-in-action problem, maximum principle, superellipse, ellipsoidal approximations

DOI: 10.25728/arcRAS.2023.98.83.001

1. INTRODUCTION

One of the natural control quality functions is the time spent by the system to achieve a given terminal state. In practice, the resulting optimal control problem is called the speed-in-action problem. It is essential that the speed-in-action problem for linear discrete-time systems has a number of serious differences from a similar problem for continuous systems. While in the case of continuous time, the solution obtained on the basis of the Pontryagin's maximum principle [1] for a linear system guarantees the relay nature of the optimal control in terms of speed, a similar result for a system with discrete time [2, 3] is incorrect.

The direct approach based on minimizing the norm of the terminal state for all control actions turns out to be difficult to apply for high-dimensional systems with a large time horizon and vector control. This is due to the fact that the resulting mathematical programming problem is characterized by a rapid increase in the number of constraints and optimization variables with an increase in the number of steps required for the system to reach origin. At the same time, for almost all initial states, the extremum in the speed-in-action problem is irregular [4], which also complicates the use of known numerical methods.

Consideration of the optimality conditions of the process using various classical approaches leads to two fundamentally different methods for solving the speed problem. Bellman's dynamic programming method [5] makes it possible to construct an optimal control in a positional form. In the case when the set of admissible control values is a polyhedron, the calculation of each control action is reduced to solving a linear programming problem [6]. Also, in [6], a method for forming optimal control in the case of arbitrary convex control constraints based on polyhedral approximation is demonstrated [7]. This approach has a number of disadvantages related to computational difficulties. The accuracy of the guaranteeing solution in the speed-in-action problem is achieved by increasing the number of vertices of the polyhedral approximation, which leads to an exponential increase in the complexity of the corresponding linear programming problems. Due to this fact, such approach, when implemented on standard computing devices, is characterized by either low accuracy of the solution, or a relatively small time horizon, especially for large-dimensional systems.

On the contrary, the combination of optimality conditions in the speed-in-action problem with the discrete maximum principle [1–3] allows optimal program control to be formed [4]. An essential condition for the applicability of these methods is the strict convexity of the set of admissible control values. But the relation for calculating the initial state of a conjugate system in the case of an arbitrary structure of control constraints is difficult to solve. In [8], a special case of an ellipsoidal structure of a set of admissible control values is presented. An analytical solution to the speed-in-action problem based on the necessary and sufficient optimality conditions presented in [4].

A natural approach is to combine the ideas of constructing a guaranteeing solution from [6] on the basis of ellipsoidal approximation of the set of admissible control values in combination with methods of forming program control according to the discrete maximum principle [4, 8]. The technique of ellipsoidal approximation is widely used in the theory of optimal control [9, 10]. However, the class of ellipsoids does not allow achieving arbitrary accuracy of the approximation of the initial set, and consequently, the accuracy of solving the optimal control problem. The article considers a class of superellipsoidal sets (the exact definition is given in Section 2), which allow a higher order of the accuracy while maintaining strict convexity conditions, which guarantees the simplicity of solving the speed-in-action problem in the same way [8].

Superellipses on the plane have been known for a long time as Lame curves [11] and have a large number of different applications in natural science and technical disciplines. They are actively used, for example, in geodesy and mapping tasks [12], in botany for modeling plant growth [13] and describing natural shapes [14], designing waveguides for antenna arrays [15, 16] or for modeling bends of various structures [17]. However, the general study of the properties of these figures is usually limited to the two-dimensional case [12, 18]. This fact makes it relevant to study the properties of this class of geometric bodies in a space of arbitrary dimension in terms of convex analysis: the description of their support function, support point and normal cone, the solution of various approximation problems.

The purpose of this work is to develop a method for generating optimal control explicitly in the case of a superellipsoidal structure of a set of acceptable control values, as well as to describe an approach for constructing a superellipsoidal approximation of an arbitrary convex body with the highest possible accuracy. The fundamental difference from this paper and both classical [19–21] and modern [22, 23] results is the consideration of arbitrary vector control, which is convex constrained, and the lack of restrictions on the dimension of the phase space. It is a more general statement of the problem, expanding the range of possible applications.

The article has the following structure. Section 2 presents non-standard designations and assumptions that are used in the article. In Section 3, the speed-in-action problem is considered, the maximum principle is described as the main tool for its solution, and the formulation of the problem of superellipsoidal approximation of the set of admissible control values is formulated in order to form a guaranteeing process. Section 4 presents an exact solution to the speed-in-action problem in the case of a superellipsoidal structure of a set of admissible control values. Section 5 describes a method for reducing the problem of optimal in the sense of the Lebesgue measure superellipsoidal approximation of a convex body to a number of convex programming problems. Section 6 demonstrates examples of constructing a guaranteeing process in a speed-in-action problem for systems of different dimensions based on the obtained theoretical results. Estimates of the accuracy of the constructed processes in comparison with the optimal solution are given.

2. DESIGNATIONS

We will assume that the phase space is a Euclidean space $\mathbb{R}^n$ with a scalar product defined by the relation

$$(x,y)=\sum_{i=1}^{n} x_i y_i.$$

For any $r \in [1;+\infty)$ define on $\mathbb{R}^n$ norm

$$\|x\|_r=\left(\sum_{i=1}^{n}|x_i|^r\right)^{\frac{1}{r}}.$$

For $r=2$ the norm $\|\cdot\|_2$ is consistent with the scalar product. From the point of view of theory, the value $r=1$ is acceptable, but it will not be considered within the paper, which allows us to define the number $q>1$ as a Helder dual of the number r:

$$\frac{1}{r}+\frac{1}{q}=1.$$

For arbitrary sets $\mathcal{X},\mathcal{U}\subset\mathbb{R}^n$ and the matrix $D\in\mathbb{R}^{n\times n}$ we denote the Minkowski sum by $\mathcal{X}+\mathcal{U}$

$$\mathcal{X}+\mathcal{U}=\{x+u\colon x\in\mathcal{X},\quad u\in\mathcal{U}\},$$

and we denote by $D\mathcal{U}$ the image of the set $\mathcal{U}$ under the mapping D

$$D\mathcal{U}=\{Du\colon u\in\mathcal{U}\}.$$

By $\partial\mathcal{U}$ and $\operatorname{int}\mathcal{U}$ we denote the sets of boundary and interior points of $\mathcal{U}$ respectively. $\operatorname{cone}\{\mathcal{U}\}$ is the conic hull of the set $\mathcal{U}$ [24, § 2 ch. I].

If the set $\mathcal{U}\subset\mathbb{R}^n$ is a convex compact, then for an arbitrary point $u\in\mathcal{U}$ by $\mathcal{N}(u,\mathcal{U})$ we denote the normal cone of the set $\mathcal{U}$ at the point u [24, § 2 ch. I]:

$$\mathcal{N}(u,\mathcal{U})=\left\{p\in\mathbb{R}^n\setminus\{0\}\colon (p,u)=\max_{\tilde u\in\mathcal{U}}(p,\tilde u)\right\}.$$

The elements of the normal cone $\mathcal{N}(u,\mathcal{U})$ are called support vectors to $\mathcal{U}$ at the point u. Note that by construction equality $\mathcal{N}(u,\mathcal{U})=\varnothing$ is valid if and only if the inclusion $u\in\operatorname{int}\mathcal{U}$ is correct. If the inclusion of $0\in\operatorname{int}\mathcal{U}$ is also true, then $\mathcal{U}$ will be called a convex body [25, Section 3 § 1 ch. IV] and for an arbitrary $x\in\mathbb{R}^n$ we introduce the Minkowski functional [25, Section 3 § 2 ch. III] or the calibration function [24, § 4 ch. I]:

$$M(x,\mathcal{U})=\inf\{t>0\colon x\in t\mathcal{U}\}=\inf\left\{t>0\colon \frac{x}{t}\in\mathcal{U}\right\}.$$

The strickly convex set $\mathcal{U} \subset \mathbb{R}^n$ is such set that for any $u^1, u^2 \in \mathcal{U}$, $\lambda \in (0;1)$ the inclusion $\lambda u^1 + (1-\lambda)u^2 \in \operatorname{int} \mathcal{U}$ is correct.

We will call a superellipse or superellipsoidal set for some $a_1 > 0, \ldots, a_n > 0$, $r > 1$ a set of the form

$$\mathcal{E}_r(a_1, \ldots, a_n) = \left\{ x \in \mathbb{R}^n \colon \sum_{i=1}^{n} \left| \frac{x_i}{a_i} \right|^r \leqslant 1 \right\}. \tag{1}$$

We will assume shortering $a = (a_1, \ldots, a_n)^{\mathrm{T}}$ and denote the corresponding superellipse by $\mathcal{E}_r(a)$. By $\operatorname{diag}(a) \in \mathbb{R}^{n \times n}$ we denote a diagonal matrix constructed by the vector $a \in \mathbb{R}^n$:

$$\operatorname{diag}(a) = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_n \end{pmatrix}.$$

3. PROBLEM STATEMENT

The linear discrete-time system with limited control $(A, \mathcal{U})$ is considered:

$$\begin{aligned} &x(k+1) = Ax(k) + u(k), \\ &x(0) = x_0, \quad u(k) \in \mathcal{U}, \quad k \in \mathbb{N} \cup \{0\}, \end{aligned} \tag{2}$$

where $x(k) \in \mathbb{R}^n$ is the state vector of the system, $u(k) \in \mathbb{R}^n$ is the control action, $A \in \mathbb{R}^{n \times n}$ is the matrix of the system, $\mathcal{U} \subset \mathbb{R}^n$ is the set of valid control values. It is assumed that $\det A \neq 0$, $\mathcal{U}$ is a convex compact, $0 \in \operatorname{int} \mathcal{U}$.

For the (2) system, the speed-in-action problem is solved, i.e. it is required to transfer the system $(A, \mathcal{U})$ from a given initial state $x_0 \in \mathbb{R}^n$ to the origin in the minimum number of steps $N_{\min}$:

$$N_{\min} = \min \left\{ N \in \mathbb{N} \cup \{0\} \colon \exists u(0), \ldots, u(N-1) \in \mathcal{U} \colon x(N) = 0 \right\}.$$

The control process $\{x^*(k), u^*(k-1), x^*(0)\}_{k=1}^{N_{\min}}$, satisfying the condition $x^*(N_{\min}) = 0$, we will call optimal. It is assumed that the speed-in-action problem for the system $(A, \mathcal{U})$ is solvable, i.e. $N_{\min} < \infty$. The issues of solvability of the speed-in-action problem for the (2) system are discussed in detail in [26].

The construction of optimal processes is closely related to the apparatus of 0-controllable sets [4, 6]. For an arbitrary $N \in \mathbb{N} \cup \{0\}$ we denote by $\mathcal{X}(N) \subset \mathbb{R}^n$ the 0-controllable set of the system (2) in N steps, i.e. the set of those initial states from which the system (2) can be transferred to 0 in N steps by acceptable control actions:

$$\mathcal{X}(N) = \begin{cases} \{x_0 \in \mathbb{R}^n \colon \exists u(0), \ldots, u(N-1) \in \mathcal{U} \colon x(N) = 0\}, & N \in \mathbb{N}, \\ \{0\}, & N = 0. \end{cases} \tag{3}$$

Then, according to the definition of $N_{\min}$ the following representation is also valid:

$$N_{\min} = \min\{N \in \mathbb{N} \cup \{0\} \colon x_0 \in \mathcal{X}(N)\}. \tag{4}$$

At the same time, the control, as demonstrated in [4, 6], is optimal if and only if for all $k = \overline{0, N_{\min} - 1}$ the inclusion is true

$$x^*(k+1) = Ax^*(k) + u^*(k) \in \mathcal{X}(N_{\min} - k - 1).$$

In [4], a number of results were obtained for the speed-in-action problem, which can be presented in the form of the maximum principle for a strictly convex $\mathcal{U}$.

Theorem 1. *Let $\mathcal{U} \subset \mathbb{R}^n$ be a strictly convex and compact set, $0 \in \operatorname{int} \mathcal{U}$, $\det A \neq 0$, a class of sets $\{\mathcal{X}(N)\}_{N=0}^{\infty}$ is defined according to* (3), *the control process $\{x^*(k), u^*(k-1), x^*(0)\}_{k=1}^{N_{\min}}$ and the trajectory of the conjugate system $\{\psi(k)\}_{k=0}^{N_{\min}-1}$ satisfy the relations*

$$\begin{aligned}
&x^*(k+1) = Ax^*(k) + u^*(k),\\
&u^*(k) = \alpha \arg \max_{u \in \mathcal{U}} \left((A^{-1})^{\mathrm{T}} \psi(k), u \right),\\
&\psi(k+1) = (A^{-1})^{\mathrm{T}} \psi(k),\\
&x^*(0) = x_0,\\
&-\psi(0) \in \mathcal{N}\left(x_0, \alpha \mathcal{X}(N_{\min})\right),\\
&\alpha = M(x_0, \mathcal{X}(N_{\min})).
\end{aligned}$$

Then

1) *$\{x^*(k), u^*(k-1), x^*(0)\}_{k=1}^{N_{\min}}$ is the optimal process for the system $(A,\mathcal{U})$;*
2) *if $\alpha = 1$, then the optimal process is the only one;*
3) *$-\psi(k) \in \mathcal{N}\left(x^*(k), \alpha \mathcal{X}(N_{\min} - k)\right)$, $k = \overline{0, N_{\min} - 1}$.*

From a computational point of view, the question of applying the Theorem 1 comes down to determining α and $\psi(0)$ from the conditions

$$\begin{aligned}
&-\psi(0) \in \mathcal{N}\left(x_0, \alpha \mathcal{X}(N_{\min})\right),\\
&\alpha = M(x_0, \mathcal{X}(N_{\min})).
\end{aligned} \tag{5}$$

This problem in the case of an arbitrary strictly convex body $\mathcal{U}$ can be a nontrivial problem.

The main purpose of this paper is to construct effective methods for solving the conditions (5) with respect to $\psi(0) \in \mathbb{R}^n \setminus \{0\}$ and $\alpha > 0$ for the special case when $\mathcal{U}$ allows the representation

$$\mathcal{U} = B\mathcal{E}_r(a), \quad B \in \mathbb{R}^{n \times n}, \quad \det B \neq 0, \quad a_1, \ldots, a_n > 0, \quad r > 1. \tag{6}$$

Another goal of the paper is to develop a method for approximating an arbitrary convex body $\mathcal{U}$ by a nested set $\hat{\mathcal{U}}$ of the form (6), that minimizes the Lebesgue measure of the difference between two sets $\mu(\mathcal{U} \setminus \hat{\mathcal{U}})$, in order to construct a guaranteed solution in the speed-in-action problem for the system $(A,\mathcal{U})$.

4. THE OPTIMAL PROCESS IN THE CASE OF A SUPERELLIPSOIDAL STRUCTURE OF CONTROL CONSTRAINTS

The conditions (5) can be reduced to an equivalent system of algebraic equations in the case of (6). To do this, we will carry out an analytical description of 0-controllable sets and some properties of strictly convex and superellipsoidal sets.

Lemma 1 [4, Lemma 1]. *Let $\det A \neq 0$, the class of sets $\{\mathcal{X}(N)\}_{N=0}^{\infty}$ be determined by the relations* (3). *Then for any $N \in \mathbb{N}$ the representation is true*

$$\mathcal{X}(N) = -\sum_{k=1}^{N} A^{-k} \mathcal{U}.$$

Lemma 2 [27, Lemma 3]. *Let $\mathcal{U} \subset \mathbb{R}^n$ be a strictly convex compact, $0 \in \operatorname{int} \mathcal{U}$. Then for any different $u^1, u^2 \in \mathcal{U}$ it is true that*

$$\mathcal{N}(u^1,\mathcal{U}) \cap \mathcal{N}(u^2,\mathcal{U}) = \varnothing.$$

The following statement follows from [27, Lemmas 5, 6].

Lemma 3. *Let $\mathcal{U}, \mathcal{X} \subset \mathbb{R}^n$ be convex compacts, $u \in \mathcal{U}$, $x \in \mathcal{X}$, $A \in \mathbb{R}^{n\times n}$, $\det A \neq 0$. Then*

1) $\mathcal{N}(u+x,\mathcal{U}+\mathcal{X}) = \mathcal{N}(u,\mathcal{U}) \cap \mathcal{N}(x,\mathcal{X})$;

2) $\mathcal{N}(Ax, A\mathcal{X}) = (A^{-1})^{\mathrm{T}}\mathcal{N}(x,\mathcal{X})$.

The Lemma 3 defines the transformation of the normal cone of convex sets with non-degenerate linear mapping and Minkowski addition. Taking into account the Lemma 1 this makes it possible to describe an arbitrary normal cone of any 0-controllable set in terms of the normal cones of the set $\mathcal{U}$ or $\mathcal{E}_r(a_1,\ldots,a_n)$ in the case (6). On the other hand, the Lemma 2 establishes a one-to-one correspondence between a boundary point and its normal cone for a strictly convex set. If this dependence is described explicitly, then it is possible to obtain algebraic equations equivalent to the conditions (5).

We introduce for an arbitrary $r > 1$ the bijective operator $I_r\colon \mathbb{R}^n \to \mathbb{R}^n$:

$$I_r(x) = \Big(\operatorname{sgn}(x_1)|x_1|^{r-1},\ldots,\operatorname{sgn}(x_n)|x_n|^{r-1}\Big).$$

Lemma 4. *Let the set $\mathcal{E}_r(a)$ be defined by the relations (1). Then*

1) for any $x \in \partial\mathcal{E}_r(a)$ it is true that

$$\mathcal{N}\left(x,\mathcal{E}_r(a)\right) = \Big\{\gamma \operatorname{diag}(a)^{-1} I_r\Big(\operatorname{diag}(a)^{-1}x\Big) \in \mathbb{R}^n\colon \gamma > 0\Big\};$$

2) for any $p \in \mathbb{R}^n \setminus \{0\}$ there is a unique

$$x^*(p) = \arg \max_{x\in\mathcal{E}_r(a)} (p,x) = \frac{\operatorname{diag}(a) I_q\left(\operatorname{diag}(a)p\right)}{\|\operatorname{diag}(a)p\|_q^{q-1}}.$$

The proof of the Lemma 4 and all other statements is given in the Appendix.

Lemma 5. *Let $\mathcal{U} = D\mathcal{E}_r(a)$, where $\mathcal{E}_r(a)$ is determined by the relations (1), $D \in \mathbb{R}^{n\times n}$, $\det D \neq 0$. Then*

1) for any $u \in \partial\mathcal{U}$

$$\mathcal{N}\left(u,\mathcal{U}\right) = \Big\{\gamma(D^{-1})^{\mathrm{T}}\operatorname{diag}(a)^{-1} I_r\Big(\operatorname{diag}(a)^{-1}D^{-1}u\Big) \in \mathbb{R}^n\colon \gamma > 0\Big\};$$

2) for any $p \in \mathbb{R}^n \setminus \{0\}$ there is only one

$$u^*(p) = \arg \max_{u\in\mathcal{U}} (p,u) = \frac{D\operatorname{diag}(a) I_q\Big(\operatorname{diag}(a)D^{\mathrm{T}}p\Big)}{\|\operatorname{diag}(a)D^{\mathrm{T}}p\|_q^{q-1}}.$$

The Lemma 5, on the one hand, allows us to calculate the optimal control according to the Theorem 1 in the case (6), when we choose $D = B$. On the other hand, the Lemma 5 in combination with Lemmas 1 and 2 connects a point on the boundary of the 0-controllable set with an element of its normal cone, when we choose $D = -A^{-k}B$, which makes it possible to reduce the conditions (5) to equivalent algebraic equations. We formulate this fact in the form of a theorem.

Theorem 2. *Let $\mathcal{U}$ be determined according to* (6)*, $x_0 \neq 0$, $\psi(0) \in \mathbb{R}^n \setminus \{0\}$, $\alpha > 0$. Then $\psi(0)$ and α satisfy the conditions* (5) *if and only if the following equality is true:*

$$-x_0 = \alpha \sum_{k=1}^{N_{\min}} \frac{A^{-k}B\operatorname{diag}(a)I_q\Big(\operatorname{diag}(a)(A^{-k}B)^{\mathrm{T}}\psi(0)\Big)}{\|\operatorname{diag}(a)(A^{-k}B)^{\mathrm{T}}\psi(0)\|_q^{q-1}}.$$

The system of equations presented in the Theorem 2 has not the only solution, since the right part is invariant to the multiplication of the vector $\psi(0)$ by any positive number. To use numerical methods, we can propose a modification of this system, which has a single solution.

Corollary 1. *Let $\mathcal{U}$ be determined according to* (6)*, $\psi(0) \in \mathbb{R}^n \setminus \{0\}$, $\alpha > 0$. Then for any $x_0 \neq 0$ there is a unique solution of the system of equations*

$$\begin{cases} -x_0 = \alpha \displaystyle\sum_{k=1}^{N_{\min}} \frac{A^{-k}B\operatorname{diag}(a)I_q\Big(\operatorname{diag}(a)(A^{-k}B)^{\mathrm{T}}\psi(0)\Big)}{\|\operatorname{diag}(a)(A^{-k}B)^{\mathrm{T}}\psi(0)\|_q^{q-1}}, \\ (\psi(0), \psi(0)) = 1, \end{cases}$$

which also satisfies the conditions (5)*.*

Example 1. Consider the procedure for calculating $\psi(0)$, α, $N_{\min}$ based on the Corollary 1. The parameters of the system (2) have the following values:

$$A = \begin{pmatrix} 3 & 1 \\ 1 & -2 \end{pmatrix}, \quad B = \begin{pmatrix} \dfrac{\sqrt{2}}{2} & \dfrac{\sqrt{2}}{2} \\ -\dfrac{\sqrt{2}}{2} & \dfrac{\sqrt{2}}{2} \end{pmatrix}, \quad a_1 = 2, \quad a_2 = 3,$$

$$r = \frac{4}{3}, \quad q = 4, \quad x_0 = \left(\frac{1}{3}, \frac{4}{3}\right)^{\mathrm{T}}.$$

Suppose that $N_{\min} = 2$, and we make up the system of equations presented in the Theorem 1:

$$\begin{aligned} &\frac{0.20(0.20\psi_1(0) + 0.81\psi_2(0))^3 + 0.91(0.91\psi_1(0) - 0.61\psi_2(0))^3}{\left((0.20\psi_1(0) + 0.81\psi_2(0))^4 + (0.91\psi_1(0) - 0.61\psi_2(0))^4\right)^{\frac{3}{4}}} \\ &+ \frac{0.17(0.17\psi_1(0) - 0.32\psi_2(0))^3 + 0.17(0.17\psi_1(0) + 0.39\psi_2(0))^3}{\left((0.17\psi_1(0) - 0.32\psi_2(0))^4 + (0.17\psi_1(0) + 0.39\psi_2(0))^4\right)^{\frac{3}{4}}} = -\frac{1}{3\alpha}, \\ &\frac{0.81(0.20\psi_1(0) + 0.81\psi_2(0))^3 - 0.61(0.91\psi_1(0) - 0.61\psi_2(0))^3}{\left((0.20\psi_1(0) + 0.81\psi_2(0))^4 + (0.91\psi_1(0) - 0.61\psi_2(0))^4\right)^{\frac{3}{4}}} \\ &+ \frac{-0.32(0.17\psi_1(0) - 0.32\psi_2(0))^3 + 0.39(0.17\psi_1(0) + 0.39\psi_2(0))^3}{\left((0.17\psi_1(0) - 0.32\psi_2(0))^4 + (0.17\psi_1(0) + 0.39\psi_2(0))^4\right)^{\frac{3}{4}}} = -\frac{4}{3\alpha}. \end{aligned}$$

By supplementing this system with the equivalence $\psi_1(0)^2 + \psi_2(0)^2 = 1$ according to the Corollary of 1, we get the following solution:

$$\psi_1(0) = -0.35, \quad \psi_2(0) = -0.94, \quad \alpha = 1.08.$$

Due to (5) it is true that $\alpha = M(x_0, \mathcal{X}(2))$. It is correct that $x_0 \notin \mathcal{X}(2)$ since $\alpha > 1$. We got a contradiction, from which it follows that $N_{\min} > 2$.

Assume that $N_{\min} = 3$, and make up the system of equations presented in the Theorem 1:

$$\frac{0.20\big(0.20\psi_1(0) + 0.81\psi_2(0)\big)^3 + 0.91\big(0.91\psi_1(0) - 0.61\psi_2(0)\big)^3}{\Big(\big(0.20\psi_1(0) + 0.81\psi_2(0)\big)^4 + \big(0.91\psi_1(0) - 0.61\psi_2(0)\big)^4\Big)^{\frac{3}{4}}}$$
$$+ \frac{0.17\big(0.17\psi_1(0) - 0.32\psi_2(0)\big)^3 + 0.17\big(0.17\psi_1(0) + 0.39\psi_2(0)\big)^3}{\Big(\big(0.17\psi_1(0) - 0.32\psi_2(0)\big)^4 + \big(0.17\psi_1(0) + 0.39\psi_2(0)\big)^4\Big)^{\frac{3}{4}}}$$
$$+ \frac{0.004\big(0.004\psi_1(0) + 0.16\psi_2(0)\big)^3 + 0.11\big(0.11\psi_1(0) - 0.14\psi_2(0)\big)^3}{\Big(\big(0.004\psi_1(0) - 0.16\psi_2(0)\big)^4 + \big(0.11\psi_1(0) - 0.14\psi_2(0)\big)^4\Big)^{\frac{3}{4}}} = -\frac{1}{3\alpha},$$

$$\frac{0.81\big(0.20\psi_1(0) + 0.81\psi_2(0)\big)^3 - 0.61\big(0.91\psi_1(0) - 0.61\psi_2(0)\big)^3}{\Big(\big(0.20\psi_1(0) + 0.81\psi_2(0)\big)^4 + \big(0.91\psi_1(0) - 0.61\psi_2(0)\big)^4\Big)^{\frac{3}{4}}}$$
$$+ \frac{-0.32\big(0.17\psi_1(0) - 0.32\psi_2(0)\big)^3 + 0.39\big(0.17\psi_1(0) + 0.39\psi_2(0)\big)^3}{\Big(\big(0.17\psi_1(0) - 0.32\psi_2(0)\big)^4 + \big(0.17\psi_1(0) + 0.39\psi_2(0)\big)^4\Big)^{\frac{3}{4}}}$$
$$+ \frac{0.17\big(0.004\psi_1(0) + 0.16\psi_2(0)\big)^3 - 0.14\big(0.11\psi_1(0) - 0.14\psi_2(0)\big)^3}{\Big(\big(0.004\psi_1(0) + 0.16\psi_2(0)\big)^4 + \big(0.11\psi_1(0) - 0.14\psi_2(0)\big)^4\Big)^{\frac{3}{4}}} = -\frac{4}{3\alpha}.$$

By supplementing this system with the equivalence $\psi_1(0)^2 + \psi_2(0)^2 = 1$ according to the corollary of 1, we get the following solution:

$$\psi_1(0) = -0.50, \quad \psi_2(0) = -0.87, \quad \alpha = 0.96.$$

Then $\alpha = M(x_0, \mathcal{X}(3)) < 1$, i.e. by definition of the Minkowski functional $x_0 \in \mathcal{X}(3)$. Due to (4) it is true that $N_{\min} = 3$.

Remark 1. In the Example 1 and everywhere else, the numerical solution of systems of algebraic equations constructed according to the Corollary 1, is carried out in the Maple software environment by means of built-in procedures based on the Newton method and its modifications [28].

The Theorem 2 and the Corollary 1 in conjunction with the Theorem 1 allow us to completely solve the speed-in-action problem for a linear discrete-time system in the case of a superellipsoidal structure of the set of admissible values of control (6). The solution of the conditions (5) according to the Corollary 1 is equivalent to the numerical solution of a system of algebraic equations. At the same time, the optimal process and the trajectory of the conjugate system can be calculated from the recurrence relations presented in the Theorem 1. Optimal control is explicitly defined by point 2 of Lemma 5.

5. INTERNAL SUPERELLIPSOIDAL APPROXIMATION OF A CONVEX BODY

The case of (6) is quite special. It is often impossible to guarantee even the strict convexity of the set $\mathcal{U}$. In this connection, it turns out to be relevant to carry out an internal approximation of $\mathcal{U}$ by a set $\hat{\mathcal{U}}$ of the form (6). Transition in the speed-in-action problem from the system $(A,\mathcal{U})$ to the auxiliary system $(A,\hat{\mathcal{U}})$ allows us to construct a guaranteeing control in the original system based on the methods presented in the Section 4 in relation to the auxiliary system.

In this case, the error of the guaranteeing solution in comparison with the optimal one will be the smaller, the larger the approximating set $\hat{\mathcal{U}}$ is by inclusion. This fact leads to the need to solve the problem of optimal superellipsoidal approximation of a convex compact body $\mathcal{U} \subset \mathbb{R}^n$ by a set of the form (6). As an approximation quality criterion, we consider the Lebesgue measure of the n-dimensional set $\mu(\cdot)$ [25, Section 1 § 3 ch. V]. The resulting optimization problem will take the form

$$\begin{aligned}
&\mu(\mathcal{U} \setminus B\mathcal{E}_r(a_1,\dots,a_n)) \to \min_{a_1,\dots,a_n,r,B},\\
&a_i > 0,\ i = \overline{1,n},\\
&r > 1,\\
&B \in \mathbb{R}^{n\times n},\quad \det B \neq 0,\\
&\mathcal{E}_r(a_1,\dots,a_n) \subset \mathcal{U}.
\end{aligned}$$

This problem can be divided into two separate stages: the first stage is the selection of the orientation matrix of the superellipse $B \in R^{n\times n}$and the second stage is the selection of the numbers $a_1,\dots,a_n > 0$, $r > 1$, parametrizing the set (1).

5.1. Selection of the Orientation Matrix of a Superellipsoidal Set

In general, the search for the optimal value of the matrix B can be a complex optimization problem, the solvability conditions of which are unknown due to its non-convexity. We propose a heuristic method for choosing B in the form of an orthogonal matrix. Since the rotation transformation preserves the Lebesgue measure, then the following equalities are valid:

$$\mu(\mathcal{U} \setminus B\mathcal{E}_r(a)) = \mu(B^{-1}(\mathcal{U} \setminus B\mathcal{E}_r(a))) = \mu(B^{-1}\mathcal{U} \setminus \mathcal{E}_r(a)).$$

They make it possible to reduce the original approximation problem to the problem of optimal internal approximation of an arbitrary convex compact body $B^{-1}\mathcal{U} \subset \mathbb{R}^n$ by the superellipse $\mathcal{E}_r(a)$. Due to the symmetry of the set $\mathcal{E}_r(a)$, it is acceptable to assume that B^{-1} should provide such a rotation of the set $\mathcal{U}$, so that the coordinate axes coincide with any axes of "symmetry" of $\mathcal{U}$, for example, with the main axes of inertia of a convex bodie $\mathcal{U}$ [29, § 32 ch. VI].

In this case, B must satisfy the condition

$$I_{\mathcal{U}} = B\mathrm{diag}(\lambda_1,\dots,\lambda_n)B^{-1},$$

where $I_{\mathcal{U}} \in \mathbb{R}^{n\times n}$ is the inertia tensor of a convex body $\mathcal{U} \subset \mathbb{R}^n$:

$$I_{\mathcal{U}} = \begin{pmatrix} I_{11} & \dots & I_{1n}\\ \vdots & \ddots & \vdots\\ I_{n1} & \dots & I_{nn}\end{pmatrix},\quad I_{ij} = \begin{cases} \displaystyle\int\limits_{\mathcal{U}} \sum_{\substack{k=1\\k\neq i}}^{n} x_k^2 dx_1\dots dx_n, & i = j,\\ \displaystyle -\int\limits_{\mathcal{U}} x_i x_j dx_1\dots dx_n, & i \neq j.\end{cases}$$

Then according to [30, Theorem 3.1.11] B is determined in a unique way up to the permutation of its columns and its construction is reduced to calculating the eigenvectors of the matrix $I_{\mathcal{U}}$.

Example 2. Let us calculate the matrix B for the polyhedron $\mathcal{U} \subset \mathbb{R}^2$:

$$\mathcal{U} = \mathrm{conv}\left\{\begin{pmatrix}4\\4\end{pmatrix}, \begin{pmatrix}2\\4\end{pmatrix}, \begin{pmatrix}-2\\2\end{pmatrix}, \begin{pmatrix}-4\\-4\end{pmatrix}, \begin{pmatrix}-2\\4\end{pmatrix}, \begin{pmatrix}2\\-2\end{pmatrix}\right\}.$$

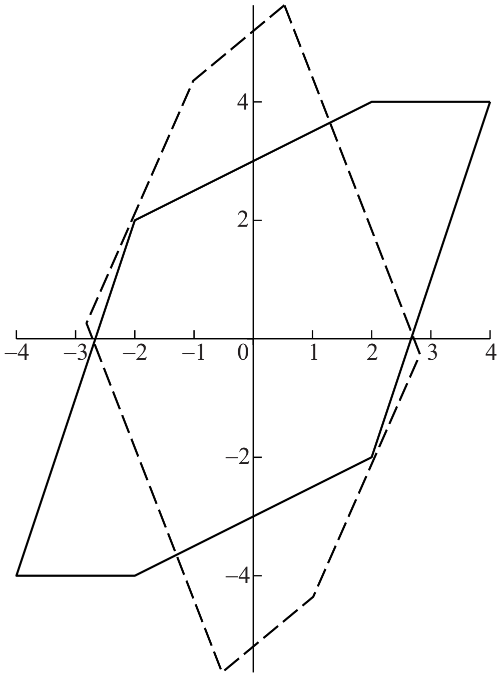

Fig. 1. The original set $\mathcal{U}$ (solid line) and the set $B^{-1}\mathcal{U}$ oriented along the axes of inertia (dotted line).

The inertia tensor $I_{\mathcal{U}}$ and the matrix B have the following numerical values:

$$I_{\mathcal{U}} = \begin{pmatrix} 153.28 & -85.03 \\ -85.03 & 121.20 \end{pmatrix}, \quad B = \begin{pmatrix} 0.64 & -0.77 \\ 0.77 & 0.64 \end{pmatrix}.$$

Then the oriented set $B^{-1}\mathcal{U}$, for which it is necessary to carry out a further superellipsoidal approximation, has the following form:

$$B^{-1}\mathcal{U} = \operatorname{conv}\left\{\begin{pmatrix} 0.53 \\ 5.63 \end{pmatrix}, \begin{pmatrix} -1.01 \\ 4.36 \end{pmatrix}, \begin{pmatrix} -2.82 \\ 0.26 \end{pmatrix}, \begin{pmatrix} -0.53 \\ -5.63 \end{pmatrix}, \begin{pmatrix} 1.01 \\ -4.36 \end{pmatrix}, \begin{pmatrix} 2.82 \\ -0.26 \end{pmatrix}\right\}.$$

The initial set $\mathcal{U}$ and the oriented set $B^{-1}\mathcal{U}$ are shown in Fig. 1.

5.2. Selection of Parameters of a Superellipsoidal Set

Next, we will assume that the matrix B of the orientation of the superellipse is selected in the form of a rotation matrix. Then the initial approximation problem is reduced to the following optimization problem:

$$\begin{aligned} &\mu(\mathcal{U} \setminus \mathcal{E}_r(a_1, \ldots, a_n)) \to \min_{a_1,\ldots,a_n,r}, \\ &a_i > 0, \quad i = \overline{1, n}, \\ &r > 1, \\ &\mathcal{E}_r(a_1, \ldots, a_n) \subset \mathcal{U}. \end{aligned} \tag{7}$$

We formulate a number of statements that allow us to reduce the problem (7) to an equivalent convex programming problem that can be solved numerically.

Lemma 6. *Let $\mathcal{E}_r(a)$ be defined by the relations* (1). *Then equality*

$$\mu(\mathcal{E}_r(a)) = a_1 \cdot \ldots \cdot a_n \frac{\left(2\Gamma\left(\frac{1}{r}+1\right)\right)^n}{\Gamma\left(\frac{n}{r}+1\right)}$$

is correct.

Lemma 7. *Let $\mathcal{E}_r(a)$ be defined by the relations* (1), *$\mathcal{U}$ is the convex body.*

Then the inclusion of $\mathcal{E}_r(a) \subset \mathcal{U}$ is valid if and only if the inequality

$$\left(\sum_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}} \geqslant M(x,\mathcal{U})$$

is true for any $x \in \mathbb{R}^n$.

Based on the Lemmas 6 and 7 we present the problem (7) in an equivalent form.

Theorem 3. *Let $\mathcal{E}_r(a)$ be defined by the relations* (1), *$\mathcal{U}$ is a convex body. Then the optimization problem* (7) *is equivalent to the following problem:*

$$\begin{aligned}
&a_1 \cdot \ldots \cdot a_n \frac{\left(2\Gamma\left(\frac{1}{r}+1\right)\right)^n}{\Gamma\left(\frac{n}{r}+1\right)} \to \max_{a_1,\ldots,a_n,r},\\
&\left(\sum_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}} \geqslant M(x,\mathcal{U}),\ \textit{for any } x \in \mathbb{R}^n,\\
&a_i > 0, \quad i = \overline{1,n},\\
&r > 1.
\end{aligned} \tag{8}$$

Generally speaking, (8) is not a convex programming problem, which means that in general it cannot be solved by standard optimization methods [31]. We will carry out a number of transformations that will allow us to solve (8) numerically. We will also separately consider the case when $\mathcal{U}$ is a polyhedron, which will allow us to explicitly construct the Minkowski functional $M(x,\mathcal{U})$.

Lemma 8. *Let $\mathcal{E}_r(a)$ be defined by the relations* (1), *$\mathcal{U}$ is a bounded polyhedron, i.e. there are such $K \in \mathbb{N}$, $p^1, \ldots, p^K \in \mathbb{R}^n \setminus \{0\}$, $\alpha_1, \ldots, \alpha_n > 0$, which provide representation*

$$\mathcal{U} = \bigcap_{k=1}^{K}\{x \in \mathbb{R}^n \colon (p^k, x) \leqslant \alpha_k\}.$$

Then the inclusion of $\mathcal{E}_r(a) \subset \mathcal{U}$ is equivalent to the condition

$$\left\|\operatorname{diag}(a)p^k\right\|_q \leqslant \alpha_k, \quad k = \overline{1,K}.$$

The complexity of solving the problem (8) lies in the fact that the set of acceptable values of the vector of optimization variables $(r, a_1, \ldots, a_n)^{\mathrm{T}}$ is not convex in $\mathbb{R}^{n+1}$. Nevertheless, for a fixed value of $r > 1$ the corresponding set of valid values of the vector $(a_1, \ldots, a_n)^{\mathrm{T}}$ is already convex. We formulate this fact in the form of a lemma.

Lemma 9. *Let $\mathcal{E}_r(a)$ be defined by the relations* (1), *$\mathcal{U}$ is a convex and compact body, for arbitrary $r > 1$ by $\mathcal{P}_r(\mathcal{U}) = \{a \in \mathbb{R}^n \colon \mathcal{E}_r(a) \subset \mathcal{U}, a_i > 0,\ i = \overline{1,n}\}$ we denote the set of all valid values of $a_1, \ldots, a_n$ in the problems* (7) *and* (8).

Then $\mathcal{P}_r(\mathcal{U})$ is a convex and compact set.

The Lemma 9 allows us to approximate the equivalent problems (7) and (8) with a similar optimization problem in which the domain of the parameter r is narrowed to a finite set:

$$r \in \{r_1, \ldots, r_M\} \subset (1; +\infty).$$

Then the approximation problem reduces to solving N convex programming problems of the following form:

$$\begin{aligned} &a_1 \cdot \ldots \cdot a_n \to \max_{a_1, \ldots, a_n}, \\ &(a_1, \ldots, a_n)^{\mathrm{T}} \in \mathcal{P}_r(\mathcal{U}). \end{aligned} \tag{9}$$

The choice of the resulting superellipsoidal approximation corresponding to a specific value of $r^* \in \{r_1, \ldots, r_M\}$ may be made in accordance with the Lemma 6 and the idea of maximizing the measure of the nested superellipse:

$$r^* = \arg \max_{r \in \{r_1, \ldots, r_M\}} \mu(\mathcal{E}_r(a^*(r))), \tag{10}$$

where $a^*(r) \in \mathbb{R}^n$ is the maximum point in the problem (9).

Example 3. Let's construct a superellipsoidal approximation for the set $B^{-1}\mathcal{U}$, calculated in the Example 2. To use the Lemma 8 we represent $B^{-1}\mathcal{U}$ as a bounded polyhedron:

$$B^{-1}\mathcal{U} = \bigcap_{k=1}^{6} \left\{ x \in \mathbb{R}^2 \colon (p^k, x) \leqslant \alpha_k \right\},$$

$$\left(p^1, \ldots, p^6\right) = \begin{pmatrix} -1.28 & -4.09 & -5.90 & -1.28 & -4.09 & -5.90 \\ 1.54 & 1.80 & -2.29 & -1.54 & -1.80 & 2.29 \end{pmatrix},$$

$$(\alpha_1, \ldots, \alpha_6) = (8, 12, 16, 8, 12, 16).$$

We describe the set $\mathcal{P}_r(\mathcal{U})$ for $r \in \left\{\frac{4}{3}, 2, 4\right\}$ and solve the corresponding optimization problems (9).

$$\begin{aligned} \mathcal{P}_{\frac{4}{3}}(\mathcal{U})\colon\ &\left(2.65a_1^4 + 5.62a_2^4\right)^{\frac{1}{4}} \leqslant 8, \\ &\left(280.53a_1^4 + 10.57a_2^4\right)^{\frac{1}{4}} \leqslant 12, \\ &\left(1208.13a_1^4 + 27.48a_2^4\right)^{\frac{1}{4}} \leqslant 16, \end{aligned} \qquad \begin{cases} a_1^*\left(\dfrac{3}{4}\right) = 2.48, \\ a_2^*\left(\dfrac{3}{4}\right) = 5.16. \end{cases}$$

$$\begin{aligned} \mathcal{P}_2(\mathcal{U})\colon\ &\left(1.63a_1^2 + 2.37a_2^2\right)^{\frac{1}{2}} \leqslant 8, \\ &\left(16.75a_1^2 + 3.25a_2^2\right)^{\frac{1}{2}} \leqslant 12, \\ &\left(34.76a_1^2 + 5.24a_2^2\right)^{\frac{1}{2}} \leqslant 16, \end{aligned} \qquad \begin{cases} a_1^*(2) = 1.92, \\ a_2^*(2) = 4.94. \end{cases}$$

$$\begin{aligned} \mathcal{P}_4(\mathcal{U})\colon\ &\left(\sqrt[3]{2.65a_1^4} + \sqrt[3]{5.62a_2^4}\right)^{\frac{3}{4}} \leqslant 8, \\ &\left(\sqrt[3]{280.53a_1^4} + \sqrt[3]{10.57a_2^4}\right)^{\frac{3}{4}} \leqslant 12, \\ &\left(\sqrt[3]{1208.13a_1^4} + \sqrt[3]{27.48a_2^4}\right)^{\frac{3}{4}} \leqslant 16, \end{aligned} \qquad \begin{cases} a_1^*(4) = 1.61, \\ a_2^*(4) = 4.16. \end{cases}$$

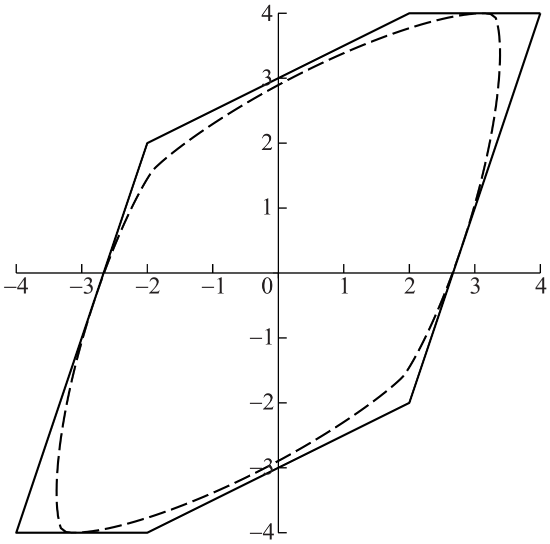

Fig. 2. The original set $\mathcal{U}$ (solid line) and its superellipsoidal approximation $B\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)$ (dotted line).

Let us compare the obtained solutions in the sense of the Lebesgue measure of the approximating superellipse in accordance with the Lemma 6:

$$\mu\left(\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)\right) = 32.60, \quad \mu\left(\mathcal{E}_2\left(a^*\left(2\right)\right)\right) = 29.79, \quad \mu\left(\mathcal{E}_4\left(a^*\left(4\right)\right)\right) = 24.86.$$

It follows that the best approximation of $B^{-1}\mathcal{U}$ is $\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)$. Therefore, for the initial set $\mathcal{U}$ the most qualitative approximation of the considered ones is the set $B\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)$. The results of the approximation are shown in Fig. 2.

6. EXAMPLES OF OPTIMAL CONTROL FORMATION

We will construct a solution to the speed problem for (2) systems of various dimensions based on the developed methods. To do this, we will use the following

Algorithm 1.

1. For a given set $\mathcal{U} \subset \mathbb{R}^n$ construct the inertia tensor $I_{\mathcal{U}}$ and calculate the orientation matrix of the superellipsoidal set $B \in \mathbb{R}^{n\times n}$ according to the Subsection 5.1.

2. Select the set of values of the superellipsoidal approximation parameter $\{r_1, \ldots, r_M\} \subset (1; +\infty)$.

3. For all $r \in \{r_1, \ldots, r_M\}$ construct optimization problems (9) for the set $B^{-1}\mathcal{U}$ and calculate the corresponding maximum points $a^*(r)$.

4. Using the Lemma 10 to determine the optimal parameter of the superellipsoidal approximation r^* according to (10).

5. For a given initial state $x_0 \in \mathbb{R}^n$ and various $N \in \mathbb{N}$ construct the systems of equations presented in the Corollary 1.

6. Determine the value of $N_{\min}$ as the smallest value of $N \in \mathbb{N}$, at which the solution of the system of equations constructed at step 5 satisfies the condition $\alpha \leqslant 1$.

7. For the value $N_{\min}$ calculated at step 6 and the corresponding $\alpha > 0$ and $\psi(0) \in \mathbb{R}^n \setminus \{0\}$ construct the optimal control $\{u^*(k)\}_{k=0}^{N_{\min}-1}$ for the system $(A, B\mathcal{E}_{r^*}(a^*(r^*)))$ according to the Theorem 1 and the Lemma 5.

Table 1. Optimal control process for a two-dimensional system

k	0	1	2	3	4	5	6	7	8	9	10
$x_1(k)$	-4.5	2.19	-3.86	3.01	-3.36	2.95	-2.82	2.54	-1.94	1.97	0
$x_2(k)$	8	-8.51	7.96	-7.86	7.28	-6.76	5.96	-4.95	3.70	-1.83	0
$u_1(k)$	3.19	0.27	2.77	-1.51	2.38	-1.95	2.21	-2.07	2.15	-2.11	–
$u_2(k)$	3.99	-2.74	3.96	-3.59	3.88	-3.75	3.83	-3.79	3.81	-3.80	–

Table 2. Results of superellipsoidal approximation for a three-dimensional system

r	$\frac{6}{5}$	$\frac{4}{3}$	2	4	6
$\mu\left(\mathcal{E}_r\left(a^*(r)\right)\right)$	57,58	61,11	57,64	41,91	35,90
$a_1^*(r)$	5,06	5,04	4,53	3,71	3,45
$a_2^*(r)$	2,48	2,24	1,58	1,20	1,10
$a_3^*(r)$	2,29	2,22	1,98	1,45	1,32

Example 4. Let $n = 2$. As $\mathcal{U}$ we choose the polyhedron considered in the Examples 2 and 3, we define the matrix of the system and the initial state as follows:

$$A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}, \quad x_0 = \begin{pmatrix} -4.5 \\ 8 \end{pmatrix}.$$

We can assume that the set $\mathcal{U}$ is approximated by $B\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)$ according to the Example 3. Then the solution of the system of equations presented in the Corollary 1, for $N = 9$ has the form

$$\alpha = 1.019, \quad \psi_1(0) = 0.775, \quad \psi_2(0) = -0.632.$$

The solution obtained for $N = 10$, has the form

$$\alpha = 0.998, \quad \psi_1(0) = 0.792, \quad \psi_2(0) = -0.610.$$

From where it follows that for the auxiliary system $\left(A, B\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)\right)$ due to (4) the equation $N_{\min} = 10$ is correct.

The optimal trajectory of the system and optimal control, calculated on the basis of the Theorem 1, are presented in Table 1.

Based on the exact methods described in [6], value $N_{\min} = 9$ was calculated for the original system $(A,\mathcal{U})$. Thus, from the point of view of control quality, the error of the guaranteeing solution is one step.

Example 5. Let $n = 3$. The set of acceptable control values, the matrix of the system and the initial state are defined as follows:

$$\mathcal{U} = \operatorname{conv}\left\{\begin{pmatrix} 4 \\ 4 \\ -3 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ -3 \end{pmatrix} \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix} \begin{pmatrix} -4 \\ -4 \\ 3 \end{pmatrix} \begin{pmatrix} -2 \\ -4 \\ 3 \end{pmatrix} \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 4 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ -4 \end{pmatrix}\right\},$$

$$A = \begin{pmatrix} 0.486 & -0.315 & 0.689 \\ -0.757 & -0.202 & 0.442 \\ 0 & -0.818 & -0.375 \end{pmatrix}, \quad x_0 = \begin{pmatrix} 26 \\ 24 \\ 30 \end{pmatrix}.$$

The inertia tensor $I_{\mathcal{U}}$ and the orientation matrix B have the form

$$I_{\mathcal{U}} = \begin{pmatrix} 526.73 & -135.75 & 132.41 \\ -135.75 & 474.79 & 164.87 \\ 132.41 & 164.87 & 439.35 \end{pmatrix}, \quad B = \begin{pmatrix} 0.49 & 0.73 & 0.48 \\ 0.59 & -0.68 & 0.43 \\ -0.64 & -0.08 & 0.76 \end{pmatrix}.$$

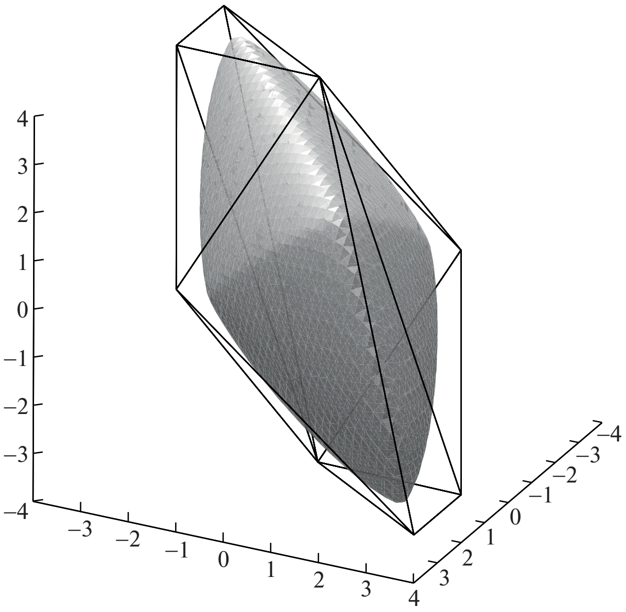

Fig. 3. The original set $\mathcal{U}$ and its super ellipsoidal approximation $B\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)$.

The superellipsoidal approximation of the set $B^{-1}\mathcal{U}$ is carried out for $r \in \left\{\frac{6}{5}, \frac{4}{3}, 2, 4, 6\right\}$. Solutions to problems of the form (9) are presented in Table 2. It follows that the best approximation is $\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)$. Graphically, the result of the superellipsoidal approximation of $\mathcal{U}$ by the set $B\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)$ is shown in Fig. 3. The solution of the system of equations presented in the Corollary 1 for $N = 9$ has the form

$$\alpha = 1.038, \quad \psi_1(0) = -0.827, \quad \psi_2(0) = -0.012, \quad \psi_3(0) = -0.563.$$

The solution obtained for $N = 10$ has the form

$$\alpha = 0.890, \quad \psi_1(0) = -0.805, \quad \psi_2(0) = -0.075, \quad \psi_3(0) = -0.589.$$

It follows that for the auxiliary system $\left(A, B\mathcal{E}_{\frac{4}{3}}\left(a^*\left(\frac{4}{3}\right)\right)\right)$ due to (4) the equality $N_{\min} = 10$ is correct.

The optimal trajectory of the system and optimal control, calculated on the basis of the Theorem 1, are presented in Table 3.

Based on the exact methods described in [6], value $N_{\min} = 8$ was calculated for the original system $(A,\mathcal{U})$. Thus, from the point of view of control quality, the error of the guaranteeing solution is 2 steps.

Table 3. Optimal control process for a three-dimensional system

k	0	1	2	3	4	5	6	7	8	9	10
$x_1(k)$	26	23.64	−3.16	17.65	11.14	−0.59	9.51	3.71	1.01	2.11	0
$x_2(k)$	24	−11.76	−25.76	14.85	−10.19	−10.80	6.71	−5.75	−1.86	−0.03	0
$x_3(k)$	30	−29.28	17.93	13.24	−16.11	11.64	4.12	−6.56	4.39	0.65	0
$u_1(k)$	−2.10	1.80	−1.28	−1.88	1.88	−1.62	−1.64	1.91	−1.99	−1.48	–
$u_2(k)$	−0.48	2.70	−0.68	0.33	2.70	−1.06	0.99	2.68	−1.59	1.30	–
$u_3(k)$	1.60	−2.67	−1.13	1.01	−2.73	−0.36	0.48	−2.77	0.77	0.21	–

7. CONCLUSION

The article considers the solution of the speed-in-action problem for linear discrete-time systems with limited control. It is assumed that the set of acceptable control values is a convex compact body containing the origin, the matrix of the system is nondegenerate. For the case of strictly convex control constraints, sufficient conditions for the optimality of the control process are formulated in the form of a discrete maximum principle. At the same time, from a practical point of view, the procedure for constructing optimal control is reduced to calculating the initial conditions of the conjugate system.

A class of superellipsoidal sets, which are a generalization of the ellipsoids for normalized space, is studied in detail. In particular, the dependence of the normal cone on the support point is explicitly described, the Lebesgue measure of the superellipse in n-dimensional space is calculated. In the case when the set of admissible values of the controls of the system is a superellipsoidal set, the definition of the initial conditions of the conjugate system in the maximum principle is reduced to a system of algebraic equations with a single solution. It is essential that the dimensionality and, consequently, the complexity of the solution of this system does not depend on the optimal value of the objective function in the speed-in-action problem, but is determined only by the number of phase variables, which ensures the effectiveness of such a method in comparison with other approaches to the solution.

For systems with a general set of admissible values of controls, a superellipsoidal approximation method has been developed, which consists in constructing a superellipse of maximum measure inscribed in original convex body. The approximation procedure is divided into two stages: the selection of the orientation matrix of the superellipse and the calculation of the parameters of the superellipsoidal set. The first stage consists in calculating the inertia tensor of the approximated body, the second stage can be reduced to solving a number of convex programming problems.

The developed technique makes it possible to build optimal control processes for various discrete-time systems. Due to the general formulation of the superellipsoidal approximation problem, it is possible to generalize the discrete maximum principle, including systems with control constraints that are not strictly convex initially, for example, systems with linear constraints.

The obtained theoretical results are tested on numerical examples.

FUNDING

The work was supported by the Russian Scientific Foundation, project no. 23-21-00293.

APPENDIX

Proof of Lemma 4. Since the Minkowski functional of set (1) is a smooth function on all $\mathbb{R}^n$:

$$M(x, \mathcal{E}_r(a) = \left(\sum_{i=1}^{n} \left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}},$$

then according to [24, Theorem 26.1] for an arbitrary $x \in \partial\mathcal{E}_r(a)$ the representation is correct

$$\begin{aligned}\mathcal{N}(x, \mathcal{E}_r(a)) &= \mathrm{cone}\{\nabla_x M(x, \mathcal{E}_r(a))\} \setminus \{0\} \\ &= \mathrm{cone}\left\{\frac{1}{r}\left(\sum_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}-1}\left(\frac{r|x_1|^{r-1}\mathrm{sgn}(x_1)}{|a_1|^r}, \ldots, \frac{r|x_n|^{r-1}\mathrm{sgn}(x_n)}{|a_n|^r}\right)^{\mathrm{T}}\right\} \setminus \{0\} \\ &= \mathrm{cone}\left\{\left(\frac{|x_1|^{r-1}\mathrm{sgn}(x_1)}{|a_1|^r}, \ldots, \frac{|x_n|^{r-1}\mathrm{sgn}(x_n)}{|a_n|^r}\right)^{\mathrm{T}}\right\} \setminus \{0\}.\end{aligned}$$

Hence follows point 1 of the Lemma 4.

According to the definition of a normal cone, the following inclusion is true:

$$p \in \mathcal{N}(x^*(p), \mathcal{E}_r(a)).$$

Then, taking into account point 1 of Lemma 4 there will be $\alpha > 0$ such that

$$p = \alpha \left(\frac{|x_1^*(p)|^{r-1}\mathrm{sgn}(x_1^*(p))}{|a_1|^r}, \ldots, \frac{|x_n^*(p)|^{r-1}\mathrm{sgn}(x_n^*(p))}{|a_n|^r} \right)^{\mathrm{T}},$$

$$\begin{aligned} x^*(p) &= \frac{1}{\alpha^{\frac{1}{r-1}}} \left(|p_1 a_1^r|^{\frac{1}{r-1}} \mathrm{sgn}(p_1), \ldots, |p_n a_n^r|^{\frac{1}{r-1}} \mathrm{sgn}(p_n) \right)^{\mathrm{T}} \\ &= \frac{1}{\alpha^{\frac{1}{r-1}}} \left(|p_1|^{q-1} a_1^q \mathrm{sgn}(p_1), \ldots, |p_n|^{q-1} a_n^q \mathrm{sgn}(p_n) \right)^{\mathrm{T}} \\ &= \frac{1}{\alpha^{\frac{1}{r-1}}} \mathrm{diag}(a) I_q(\mathrm{diag}(a)p). \end{aligned}$$

The value of α can be calculated from the condition $x^*(p) \in \partial\mathcal{E}_r(a)$, which is equivalent to the equality

$$\left(\sum_{i=1}^n \left| \frac{x_i^*(p)}{a_i} \right|^r \right)^{\frac{1}{r}} = 1,$$

$$1 = \frac{1}{\alpha^{\frac{1}{r-1}}} \left(\sum_{i=1}^n \left| \frac{|p_i|^{q-1} a_i^q}{a_i} \right|^r \right)^{\frac{1}{r}} = \frac{1}{\alpha^{\frac{1}{r-1}}} \left(\sum_{i=1}^n |p_i a_i|^q \right)^{\frac{1}{r}},$$

$$\alpha^{\frac{1}{r-1}} = \left(\sum_{i=1}^n |p_i a_i|^q \right)^{\frac{1}{r}} = \|\mathrm{diag}(a)p\|_q^{q-1}.$$

The second point of the Lemma 4 is proved.

Proof of Lemma 5. Point 1 follows from point 1 of Lemma 4, point 2 of Lemma 3 and the representation

$$\mathcal{N}(u, \mathcal{U}) = \mathcal{N}(DD^{-1}u, D\mathcal{E}_r(a)).$$

Point 2 follows from point 2 of the Lemma 4 and the chain of equalities

$$\arg \max_{u \in \mathcal{U}} (p, u) = D\arg \max_{x \in \mathcal{E}_r(a)} (p, Dx) = D\arg \max_{x \in \mathcal{E}_r(a)} (D^{\mathrm{T}}p, x).$$

Lemma 5 is proved.

Proof of Theorem 2. Since $x_0 \neq 0$, then according to the definitions of the Minkowski functional and the normal cone, the conditions (5) are equivalent to the conditions

$$-\psi(0) \in \mathcal{N}\left(\frac{x_0}{\alpha}, \mathcal{X}(N_{\min}) \right), \tag{A.1}$$

$$\frac{x_0}{\alpha} \in \partial\mathcal{X}(N_{\min}). \tag{A.2}$$

The inclusion of (A.2) due to the Lemma 1 and the representation (6) is equivalent to the condition

$$\frac{x_0}{\alpha} \in \partial \left(-\sum_{k=1}^{N_{\min}} A^{-k}\mathcal{U} \right) = \partial \sum_{k=1}^{N_{\min}} A^{-k} B\mathcal{E}_r(a).$$

Then, taking into account point 1 of the Lemma 3 and the definition of the algebraic sum of sets inclusion (A.1) is equivalent to the fact that there are $x^1 \in A^{-1}B\mathcal{E}_r(a), \ldots, x^{N_{\min}} \in A^{-N_{\min}}B\mathcal{E}_r(a)$, for which the following relations are true:

$$\frac{x_0}{\alpha} = \sum_{k=1}^{N_{\min}} x^k,$$

$$-\psi(0) \in \mathcal{N}\left(\frac{x_0}{\alpha}, \mathcal{X}(N_{\min})\right) = \mathcal{N}\left(\sum_{k=1}^{N_{\min}} x^k, \sum_{k=1}^{N_{\min}} A^{-k}B\mathcal{E}_r(a)\right) = \bigcap_{k=1}^{N_{\min}} \mathcal{N}\left(x^k, A^{-k}B\mathcal{E}_r(a)\right).$$

Due to point 2 of Lemma 5 it is possible if and only if the condition

$$x^k = \frac{A^{-k}B\mathrm{diag}(a)I_q\left(-\mathrm{diag}(a)(A^{-k}B)^{\mathrm{T}}\psi(0)\right)}{\|\mathrm{diag}(a)(A^{-k}B)^{\mathrm{T}}\psi(0)\|_q^{q-1}}$$

is correct. Since $I_q(-x) = -I_q(x)$ for any $x \in \mathbb{R}^n$, we obtain equivalent relations

$$\frac{x_0}{\alpha} = \sum_{k=1}^{N_{\min}} x^k = -\sum_{k=1}^{N_{\min}} \frac{A^{-k}B\mathrm{diag}(a)I_q\left(\mathrm{diag}(a)(A^{-k}B)^{\mathrm{T}}\psi(0)\right)}{\|\mathrm{diag}(a)(A^{-k}B)^{\mathrm{T}}\psi(0)\|_q^{q-1}}.$$

That is, the conditions (5) are equivalent to the equality specified in the condition of the Theorem 2.

Proof of Corollary 1. Due to the Theorem 2 the solution of the system exists and satisfies the conditions (5). Then, due to the Lemma 1 and the symmetry of sets of the form (1) there will be such $x^1 \in \alpha A^{-1}B\mathcal{E}_r(a), \ldots, x^{N_{\min}} \in \alpha A^{-N_{\min}}B\mathcal{E}_r(a)$, which make true equality $x_0 = x^1 + \ldots x^{N_{\min}}$. From where, by point 1 of Lemma 3 it follows that any solution $(\psi(0), \alpha)$ satisfies inclusion

$$-\psi(0) \in \mathcal{N}\left(x_0, \alpha\mathcal{X}(N_{\min})\right) = \bigcap_{k=1}^{N_{\min}} \mathcal{N}\left(x^k, A^{-k}B\mathcal{E}_r(a)\right).$$

But according to point 1 of Lemma 5 for all $k = \overline{1, N_{\min}}$ sets $\mathcal{N}\left(x^k, A^{-k}B\mathcal{E}_r(a)\right)$ are one-dimensional rays with starting at 0, i.e. they contain a single vector $-\psi(0)$, satisfying the equality $(\psi(0), \psi(0)) = 1$. The uniqueness of the value $\alpha > 0$ follows from the definition of the Minkowski functional and the conditions (5).

The consequence 1 is proved.

Lemma 10. *Let $\mathcal{E}_r(a)$ be defined by the relations* (1). *Then*

$$\mu(\mathcal{E}_r(a)) = a_1 \cdot \ldots \cdot a_n \mu(\mathcal{E}_r(1, \ldots, 1)).$$

Proof of Lemma 10. Consider replacement of variables

$$\begin{cases} x_1 = a_1 y_1, \\ \vdots \\ x_n = a_n y_n, \end{cases}$$

the Jacobian of which has the form $J = a_1 \cdot \ldots \cdot a_n$. Then

$$\mu(\mathcal{E}_r(a)) = \int\limits_{\sum_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r \leqslant 1} 1dx = \int\limits_{\sum_{i=1}^{n}|y_i|^r \leqslant 1} |J|dy = a_1 \cdot \ldots \cdot a_n \mu(\mathcal{E}_r(1, \ldots, 1)).$$

The Lemma 10 is proved.

Proof of Lemma 6. In the part of the space $x_i \geqslant 0$, $i = \overline{1,n}$ consider the replacement of variables

$$\begin{cases} x_1 = R(\cos\phi_2 \cdot \cos\phi_3 \ldots \cdot \cos\phi_n)^{\frac{2}{r}}, \\ x_2 = R(\sin\phi_2 \cdot \cos\phi_3 \ldots \cdot \cos\phi_n)^{\frac{2}{r}}, \\ x_3 = R(\sin\phi_3 \cdot \cos\phi_4 \ldots \cdot \cos\phi_n)^{\frac{2}{r}}, \\ \vdots \\ x_n = R(\sin\phi_n)^{\frac{2}{r}}. \end{cases} \tag{A.3}$$
$$R \geqslant 0, \quad \phi_j \in \left(0; \frac{\pi}{2}\right), \quad j = \overline{2,n}.$$

Construct a replacement Jacobian (A.3).

$$\frac{\partial x_i}{\partial R} = \frac{x_i}{R}, \quad i = \overline{1,n}, \qquad \frac{\partial x_i}{\partial \phi_j} = \begin{cases} \dfrac{2}{r}\dfrac{\cos\phi_j}{\sin\phi_j}x_i, \quad i = \overline{2,n}, \quad j = i, \\ -\dfrac{2}{r}\dfrac{\sin\phi_j}{\cos\phi_j}x_i, \quad i = \overline{1,n-1}, \quad j = \overline{i+1,n}, \\ 0, \quad i = \overline{3,n}, \quad j = \overline{2,i-1}, \end{cases}$$

$$J = \begin{vmatrix} \dfrac{x_1}{R} & \dfrac{x_2}{R} & \dfrac{x_3}{R} & \ldots & \dfrac{x_n}{R} \\ -\dfrac{2x_1}{r}\tan\phi_2 & \dfrac{2x_2}{r}\mathrm{cot}\phi_2 & 0 & \ldots & 0 \\ -\dfrac{2x_1}{r}\tan\phi_3 & -\dfrac{2x_2}{r}\tan\phi_2 & \dfrac{2x_3}{r}\mathrm{cot}\phi_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\dfrac{2x_1}{r}\tan\phi_n & -\dfrac{2x_2}{r}\tan\phi_n & -\dfrac{2x_3}{r}\tan\phi_n & \ldots & \dfrac{2x_n}{r}\mathrm{cot}\phi_n \end{vmatrix}$$

$$= \frac{1}{R}\left(\prod_{i=1}^{n} x_i\right)\left(\prod_{j=2}^{n}\tan\phi_j\right)\left(\frac{2}{r}\right)^{n-1} \begin{vmatrix} 1 & 1 & 1 & \ldots & 1 \\ -1 & \cot^2\phi_2 & 0 & \ldots & 0 \\ -1 & -1 & \cot^2\phi_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \ldots & \cot^2\phi_n \end{vmatrix}$$

$$= \frac{1}{R}\left(\prod_{i=1}^{n} x_i\right)\left(\prod_{j=2}^{n}\tan\phi_j\right)\left(\frac{2}{r}\right)^{n-1} \begin{vmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & \cot^2\phi_2 + 1 & 1 & \ldots & 1 \\ 0 & 0 & \cot^2\phi_3 + 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \cot^2\phi_n + 1 \end{vmatrix}$$

$$= \frac{1}{R}\left(\prod_{i=1}^{n} x_i\right)\left(\prod_{j=2}^{n}(\tan\phi_j + \mathrm{cot}\phi_j)\right)\left(\frac{2}{r}\right)^{n-1} = \frac{1}{R}\left(\prod_{i=1}^{n} x_i\right)\left(\prod_{j=2}^{n}\frac{1}{\sin\phi_j\cos\phi_j}\right)\left(\frac{2}{r}\right)^{n-1}$$

$$= R^{n-1}\left(\frac{2}{r}\right)^{n-1}\prod_{j=2}^{n}(\sin\phi_j)^{\frac{2}{r}-1}(\cos\phi_j)^{\frac{2}{r}(j-1)-1}.$$

Then we can calculate the Lebesgue measure of the superellipse $\mathcal{E}_r(1,\dots,1)$ via the Lebesgue integral:

$$\mu(\mathcal{E}_r(1,\dots,1)) = \int\limits_{\sum\limits_{i=1}^{n}|x_i|^r\leqslant 1} 1dx = 2^n\int\limits_0^1 R^{n-1}\left(\frac{2}{r}\right)^{n-1}dR\prod_{j=2}^{n}\int\limits_0^{\frac{\pi}{2}}(\sin\phi_j)^{\frac{2}{r}-1}(\cos\phi_j)^{\frac{2}{r}(j-1)-1}d\phi_j.$$

For each $j=\overline{2,n}$ we calculate auxiliary integrals:

$$\int\limits_0^{\frac{\pi}{2}}(\sin\phi_j)^{\frac{2}{r}-1}(\cos\phi_j)^{\frac{2}{r}(j-1)-1}d\phi_j = \int\limits_0^{\frac{\pi}{2}}(\sin\phi_j)^{\frac{2}{r}-1}(\cos\phi_j)^{\frac{2}{r}(j-1)-2}d\sin\phi_j$$

$$= \int\limits_0^{\frac{\pi}{2}}(\sin\phi_j)^{\frac{2}{r}-1}\left(1-\sin^2\phi_j\right)^{\frac{j-1}{r}-1}d\sin\phi_j = \frac{1}{2}\int\limits_0^{\frac{\pi}{2}}\left(\sin^2\phi_j\right)^{\frac{1}{r}-1}\left(1-\sin^2\phi_j\right)^{\frac{j-1}{r}-1}d\sin^2\phi_j$$

$$= \frac{1}{2}\int\limits_0^1 t^{\frac{1}{r}-1}(1-t)^{\frac{j-1}{r}-1}dt = \frac{1}{2}\mathrm{B}\left(\frac{1}{r},\frac{j-1}{r}\right),$$

where $\mathrm{B}(x,y)$ denotes the Euler beta function.

Then the original integral has the form

$$\mu(\mathcal{E}_r(1,\dots,1)) = \frac{2^n}{n}\left(\frac{2}{r}\right)^{n-1}\prod_{j=2}^{n}\left(\frac{1}{2}\mathrm{B}\left(\frac{1}{r},\frac{j-1}{r}\right)\right) = \frac{2}{n}\left(\frac{2}{r}\right)^{n-1}\prod_{j=2}^{n}\frac{\Gamma\left(\frac{1}{r}\right)\Gamma\left(\frac{j-1}{r}\right)}{\Gamma\left(\frac{1}{r}+\frac{j-1}{r}\right)}$$

$$= \frac{2}{n}\left(\frac{2}{r}\Gamma\left(\frac{1}{r}\right)\right)^{n-1}\prod_{j=1}^{n-1}\frac{\Gamma\left(\frac{j}{r}\right)}{\Gamma\left(\frac{j+1}{r}\right)} = \frac{2}{n}\left(\frac{2}{r}\Gamma\left(\frac{1}{r}\right)\right)^{n-1}\frac{\Gamma\left(\frac{1}{r}\right)}{\Gamma\left(\frac{n}{r}\right)} = \frac{\left(\frac{2}{r}\Gamma\left(\frac{1}{r}\right)\right)^n}{\frac{n}{r}\Gamma\left(\frac{n}{r}\right)} = \frac{\left(2\Gamma\left(\frac{1}{r}+1\right)\right)^n}{\Gamma\left(\frac{n}{r}+1\right)}.$$

Taking into account the Lemma 10 we finally obtain the equality

$$\mu(\mathcal{E}_r(a)) = a_1\cdot\ldots\cdot a_n\frac{\left(2\Gamma\left(\frac{1}{r}+1\right)\right)^n}{\Gamma\left(\frac{n}{r}+1\right)}.$$

Lemma 6 is proved.

Lemma 11. *Let $\mathcal{U}_1,\mathcal{U}_2\subset\mathbb{R}^n$ be convex and compact bodies containing* 0 *as an internal point. In this case, the inclusion $\mathcal{U}_1\subset\mathcal{U}_2$ is true if and only if the following inequality is correct for any $x\in\mathbb{R}^n$:*

$$M(x,\mathcal{U}_1)\geqslant M(x,\mathcal{U}_2).$$

Proof of Lemma 11. Let $\mathcal{U}_1\subset\mathcal{U}_2$, $x\in\mathbb{R}^n$. Then by definition of the Minkowski functional

$$x\in M(x,\mathcal{U}_1)\mathcal{U}_1\subset M(x,\mathcal{U}_1)\mathcal{U}_2,$$
$$M(x,\mathcal{U}_1)\geqslant\inf\{t>0\colon x\in t\mathcal{U}_2\} = M(x,\mathcal{U}_2).$$

Let for all $x\in\mathbb{R}^n$ inequality be fair

$$M(x,\mathcal{U}_1)\geqslant M(x,\mathcal{U}_2).$$

Then by definition of the Minkowski functional

$$\mathcal{U}_1 = \{x \in \mathbb{R}^n\colon M(x,\mathcal{U}_1) \leqslant 1\} \subset \{x \in \mathbb{R}^n\colon M(x,\mathcal{U}_2) \leqslant 1\} = \mathcal{U}_2.$$

The Lemma 11 is proved.

Proof of Lemma 7. Lemma 7 follows directly from Lemma 11 and the fact that

$$M(x, \mathcal{E}_r(a)) = \left(\sum_{i=1}^{n} \left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}}.$$

Lemma 7 is proved.

Proof of Theorem 3. Due to the Lemma 7 the inclusion $\mathcal{E}_r(a) \subset \mathcal{U}$ is equivalent to the condition

$$\left(\sum_{i=1}^{n} \left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}} \geqslant M(x,\mathcal{U}) \text{ for any } x \in \mathbb{R}^n.$$

Also, due to this limitation, it is true that

$$\mu(\mathcal{U} \setminus \mathcal{E}_r(a)) = \mu(\mathcal{U}) - \mu(\mathcal{E}_r(a)).$$

Hence, taking into account the fact that the value of $\mu(\mathcal{U})$ does not depend on optimization variables, the statement of the Theorem 3 follows.

Lemma 12. *Let there be $p^1, \ldots, p^K \in \mathbb{R}^n \setminus \{0\}$ and $\alpha_1, \ldots, \alpha_K > 0$ such that*

$$\mathcal{U} = \bigcap_{k=1}^{K} \left\{u \in \mathbb{R}^n\colon (p^k, u) \leqslant \alpha_k\right\}, \quad 0 \in \operatorname{int} \mathcal{U}.$$

Then

$$M(x,\mathcal{U}) = \max_{k=\overline{1,K}} \frac{(p^k, x)}{\alpha_k}.$$

Proof of Lemma 12. Since for any $t > 0$

$$\begin{aligned} t\mathcal{U} &= \left\{u \in \mathbb{R}^n\colon u = tx, \ \ x \in \mathcal{U}\right\} = \left\{u \in \mathbb{R}^n\colon u = tx, \ \ (p^k, x) \leqslant \alpha_k, \ \ k = \overline{1, K}\right\} \\ &= \left\{u \in \mathbb{R}^n\colon \ \left(p^k, \frac{u}{t}\right) \leqslant \alpha_k, \ \ k = \overline{1, K}\right\} = \left\{u \in \mathbb{R}^n\colon \ (p^k, u) \leqslant t\alpha_k, \ \ k = \overline{1, K}\right\} \\ &= \bigcap_{k=1}^{K} \left\{u \in \mathbb{R}^n\colon (p^k, u) \leqslant t\alpha_k\right\}, \end{aligned}$$

then according to the definition of the Minkowski functional

$$\begin{aligned} M(x,\mathcal{U}) &= \inf\{t > 0\colon \ x \in t\mathcal{U}\} = \inf\left\{t > 0\colon \ (p^k, x) \leqslant t\alpha_k, \ \ k = \overline{1, K}\right\} \\ &= \inf\left\{t > 0\colon \ t \geqslant \frac{(p^k, x)}{\alpha_k}, \ \ k = \overline{1, K}\right\} = \max_{k=\overline{1,K}} \frac{(p^k, x)}{\alpha_k}. \end{aligned}$$

The Lemma 12 is proved.

Proof of Lemma 8. According to Lemmas 7 and 12, the inclusion of $\mathcal{E}_r(a) \subset \mathcal{U}$ is equivalent to the fact that for all $x \in \mathbb{R}^n$ the following inequality is valid:

$$\left(\sum_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}} \geqslant \max_{k=\overline{1,K}} \frac{(p^k, x)}{\alpha_k}.$$

For $x = 0$, this inequality holds trivially. Consider the case of $x \neq 0$ and move on to equivalent inequalities. For all $k = \overline{1,K}$

$$\left(\sum_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}} \geqslant \frac{(p^k, x)}{\alpha_k},$$

$$\alpha_k \geqslant \frac{(p^k, x)}{\left(\sum\limits_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}}}.$$

Since these inequalities must be satisfied for any $x \in \mathbb{R}^n \setminus \{0\}$, it is possible, taking into account the Lemma 4, to proceed to the equivalent relation

$$\alpha_k \geqslant \max_{x\in\mathbb{R}^n\setminus\{0\}} \frac{(p^k, x)}{\left(\sum\limits_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}}} = \max_{x\in\mathbb{R}^n\setminus\{0\}} \left(p^k, \frac{x}{\left(\sum\limits_{i=1}^{n}\left|\frac{x_i}{a_i}\right|^r\right)^{\frac{1}{r}}}\right)$$

$$= \max_{y\in\partial\mathcal{E}_r(a)} \left(p^k, y\right) = \left(p^k, x^*(p^k)\right) = \frac{\left(p^k, \operatorname{diag}(a) I_q\left(\operatorname{diag}(a)p^k\right)\right)}{\|\operatorname{diag}(a)p^k\|_q^{q-1}} = \left\|\operatorname{diag}(a)p^k\right\|_q.$$

The Lemma 8 is fully proved.

Proof of Lemma 9. Denote for arbitrary convex sets $\mathcal{U}$ and $p \in \mathbb{R}^n \setminus \{0\}$ via $s(p,\mathcal{U})$ support function $\mathcal{U}$:

$$s(p,\mathcal{U}) = \sup_{x\in\mathcal{U}}(p, x).$$

As demonstrated in [24, Theorem 11.5], an arbitrary convex compact set $\mathcal{U}$ is the intersection of all support half-spaces:

$$\mathcal{U} = \bigcap_{p\in\mathbb{R}^n\setminus\{0\}} \left\{x \in \mathbb{R}^n \colon (p, x) \leqslant s(p,\mathcal{U})\right\}.$$

Then the inclusion $\mathcal{E}_r(a) \subset \mathcal{U}$ is equivalent to the fact that for every $p \in \mathbb{R}^n \setminus \{0\}$ the following inequality will be satisfied

$$s(p, \mathcal{E}_r(a)) \leqslant s(p,\mathcal{U}). \tag{A.4}$$

Let $a, b \in \mathcal{P}_r(\mathcal{U})$, $\lambda \in (0; 1)$, $p \in \mathbb{R}^n \setminus \{0\}$. Then, due to point 2 of the Lemma 4 and the Minkowski inequality [25, section 1 §1 ch. II] folowing relations are correct:

$$s(p, \mathcal{E}_r(\lambda a + (1-\lambda)b)) = \max_{x\in\mathcal{E}_r(\lambda a+(1-\lambda)b)} (p, x) = \left(\sum_{i=1}^{n} |(\lambda a_i + (1-\lambda)b_i)p_i|^q\right)^{\frac{1}{q}}$$

$$\leqslant \lambda\left(\sum_{i=1}^{n} |a_i p_i|^q\right)^{\frac{1}{q}} + (1-\lambda)\left(\sum_{i=1}^{n} |b_i p_i|^q\right)^{\frac{1}{q}} = \lambda s(p, \mathcal{E}_r(a)) + (1-\lambda)s(p, \mathcal{E}_r(b)) \leqslant s(p,\mathcal{U}).$$

Then the condition $\mathcal{E}_r(\lambda a + (1-\lambda)b) \subset \mathcal{U}$ is correct, which by definition is equivalent to inclusion $\lambda a + (1-\lambda)b \in \mathcal{P}_r(\mathcal{U})$. This implies the convexity of $\mathcal{P}_r(\mathcal{U})$.

Choose as $p \in \mathbb{R}^n \setminus \{0\}$ the i-th coordinate vector:

$$p = (\underbrace{0,\ldots,0}_{i-1},1,0,\ldots,0)^{\mathrm{T}}.$$

Then by construction it is correct that

$$s(\pm p, \mathcal{E}_r(a)) = a_i.$$

Taking into account the condition (A.4), we obtain that for any $a \in \mathcal{P}_r(\mathcal{U})$ the following inequality is correct:

$$0 \leqslant a_i \leqslant \min\{s(p,\mathcal{U}), s(-p,\mathcal{U})\}.$$

Since $\mathcal{U}$ is limited, then for any $p \in \mathbb{R}^n \setminus \{0\}$ the value of the support function $s(p,\mathcal{U})$ is finite. Then $\mathcal{P}_r(\mathcal{U})$ is limited.

The closeness of $\mathcal{P}_r(\mathcal{U})$ follows from the closeness of $\mathcal{U}$.

The Lemma 9 is proved.

REFERENCES

1. Pontryagin, L.S., Boltyansky, V.G., Gamkrelidze, R.V., and Mishchenko, B.F., *Matematicheskaya teoriya optimal'nykh protsessov* (Mathematical Theory of Optimal Processes), Moscow: Nauka, 1969.

2. Boltyanskii, V.G., *Optimal'noe upravlenie diskretnymi sistemami* (Optimal Control of Discrete Systems), Moscow: Nauka, 1973.

3. Propoi, A.I., *Elementy teorii optimal'nykh diskretnykh protsessov* (Elements of the Theory of Optimal Discrete Processes), Moscow: Nauka, 1973.

4. Ibragimov, D.N. and Sirotin, A.N., On the Problem of Operation Speed for the Class of Linear Infinite-Dimensional Discrete-Time Systems with Bounded Control, *Autom. Remote Control*, 2017, vol. 78, no. 10, pp. 1731–1756. https://doi.org/10.1134/S0005117917100010

5. Bellman, R., *Dinamicheskoe programmirovanie* (Dynamic Programming), Moscow: Inostrannaya Literatura, 1960.

6. Ibragimov, D.N., Novozhilin, N.M., and Portseva, E.Yu., On Sufficient Optimality Conditions for a Guaranteed Control in the Speed Problem for a Linear Time-Varying Discrete-Time System with Bounded Control, *Autom. Remote Control*, 2021, vol. 82, no. 12, pp. 2076–2096. https://doi.org/10.1134/S000511792112002X

7. Kamenev, G.K., *Chislennoe issledovanie effektivnosti metodov poliedral'noi approksimatsii vypuklykh tel* (Numerical Study of the Efficiency of Polyhedral Approximation Methods for Convex Bodies), Moscow: Vychisl. Tsentr Ross. Akad. Nauk, 2010.

8. Ibragimov, D.N., The minimum-time correction of the satellite's orbit, *Elektron. Zh. Tr. MAI*, 2017, no. 94. http://trudymai.ru/published.php

9. Kurzhanskiy A. and Varaiya P., Ellipsoidal Techniques for Reachability Analysis of Discrete-Time Linear Systems, *IEEE Transactions on Automatic Control*, 2007, vol. 52, no. 1, pp. 26–38. https://doi.org/10.1109/TAC.2006.887900

10. Chernous'ko, F.L., *Otsenivanie fazovogo sostoyaniya dinamicheskikh sistem. Metod ellipsoidov* (Phase State Estimation of Dynamic Systems. The Ellipsoid Method), Moscow: Nauka, 1988.

11. Gridgeman, N.T., Lame Ovals, *The Mathematical Gazette*, 1970, vol. 54, no. 387, pp. 31–37. https://doi.org/10.2307/3613154

12. Tobler, W.R., The Hyperelliptical and Other New Pseudo Cylindrical Equal Area Map Projections, *J. Geophys. Res.*, 1973, vol. 78, no. 11, pp. 1753–1759. https://doi.org/10.1029/JB078i011p01753

13. Shi, P.J., Huang, J.G., Hui, C., Grissino-Mayer, H.D., Tardif, J.C., Zhai, L.H., Wang, F.S., and Li, B.L., Capturing Spiral Radial Growth of Conifers Using the Superellipse to Model Tree-Ring Geometric Shape, *Frontiers in Plant Science*, 2015, vol. 6, no. 856, pp. 1–13. https://doi.org/10.3389/fpls.2015.00856

14. Gielis, J., A Generic Geometric Transformation That Unifies a Wide Range of Natural and Abstract Shapes, *Amer. J. Botany*, 2003, vol. 90, no. 3, pp. 333–338. https://doi.org/10.3732/ajb.90.3.333

15. Maximidis, R.T., Caratelli, D., Toso, G., and Smolders, B.A., Analysis of a Novel Class of Waveguiding Structures Suitable for Reactively Loaded Antenna Array Design, *Doklady TUSUR*, 2017, no. 1, pp. 10–13. https://doi.org/10.21293/1818-0442-2017-20-1-09-13

16. Zolotenkova, M.K. and Egorov, V.V., Development and Analysis of Ultrasound Registrating and Performing Rodent Vocalization Device, *IEEE-EDM*, 2022, pp. 506–509. https://doi.org/10.1109/EDM55285.2022.9855056

17. Sadowski, A.J., Geometric Properties for the Design of Unusual Member Cross-Sections in Bending, *Engineering Structures*, 2011, vol. 33, no. 5, pp. 1850–1854. https://doi.org/10.1016/j.engstruct.2011.01.026

18. Tobler, W.R., Superquadrics and Angle-Preserving Transformations, *IEEE-CGA*, 1981, vol. 1, no. 1, pp. 11–23. https://doi.org/10.1109/MCG.1981.1673799

19. Desoer, C.A. and Wing, J., The Minimal Time Regulator Problem for Linear Sampled-Data Systems: General Theory, *J. Franklin Inst.*, 1961, vol. 272, no. 3, pp. 208–228. https://doi.org/10.1016/0016-0032(61)90784-0

20. Lin, W.-S., Time-Optimal Control Strategy for Saturating Linear Discrete Systems, *Int. J. Control*, 1986, vol. 43, no. 5, pp. 1343–1351. https://doi.org/10.1080/00207178608933543

21. Moroz, A.I., Synthesis of Time-Optimal Control for Linear Discrete Objects of the Third Order, *Autom. Remote Control*, 1965, vol. 25, no. 9, pp. 193–206.

22. Krasnoshchechenko, V.I., Simplex Method for Solving the Brachistochroneproblem at State and Control Constraints, *Inzhenernyi zhurnal: nauka i innovatsii*, 2014, no. 6. http://engjournal.ru/catalog/it/asu/1252.html

23. Cazanova, L.A., Stability of optimal synthesis in the time-optimality problem, *Izvestiya vuzov, Matematika*, 2002, no. 2, pp. 46–57.

24. Rockafellar, R., *Vypuklyi analiz* (Convex Analysis), Moscow: Mir, 1973.

25. Kolmogorov, A.N. and Fomin, S.V., *Elementy teorii funktsii i funktsional'nogo analiza* (Elements of the Theory of Functions and Functional Analysis), Moscow: Fizmatlit, 2012.

26. 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

27. Ibragimov, D.N., On the Optimal Speed Problem for the Class of Linear Autonomous Infinite-Dimensional Discrete-Time Systems with Bounded Control and Degenerate Operator, *Autom. Remote Control*, 2019, vol. 80, no. 3, pp. 393–412. https://doi.org/10.1134/S0005117919030019

28. Ostrowski, A.M., *Reshenie uravnenii i sistem uravnenii* (Solution of equations and systems of equations), Moscow: Izdatel'stvo inostrannoi literatury, 1963.

29. Landau, L.D. and Lifshitz, E.M., *Mekhanika* (Mechanics), Moscow: Nauka, 1988.

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

31. Ashmanov, S.A. and Timohov, S.V., *Teoriya optimizatsii v zadachakh i uprazhneniyakh* (Optimization theory in problems and exercises), Moscow: Nauka, 1991.

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