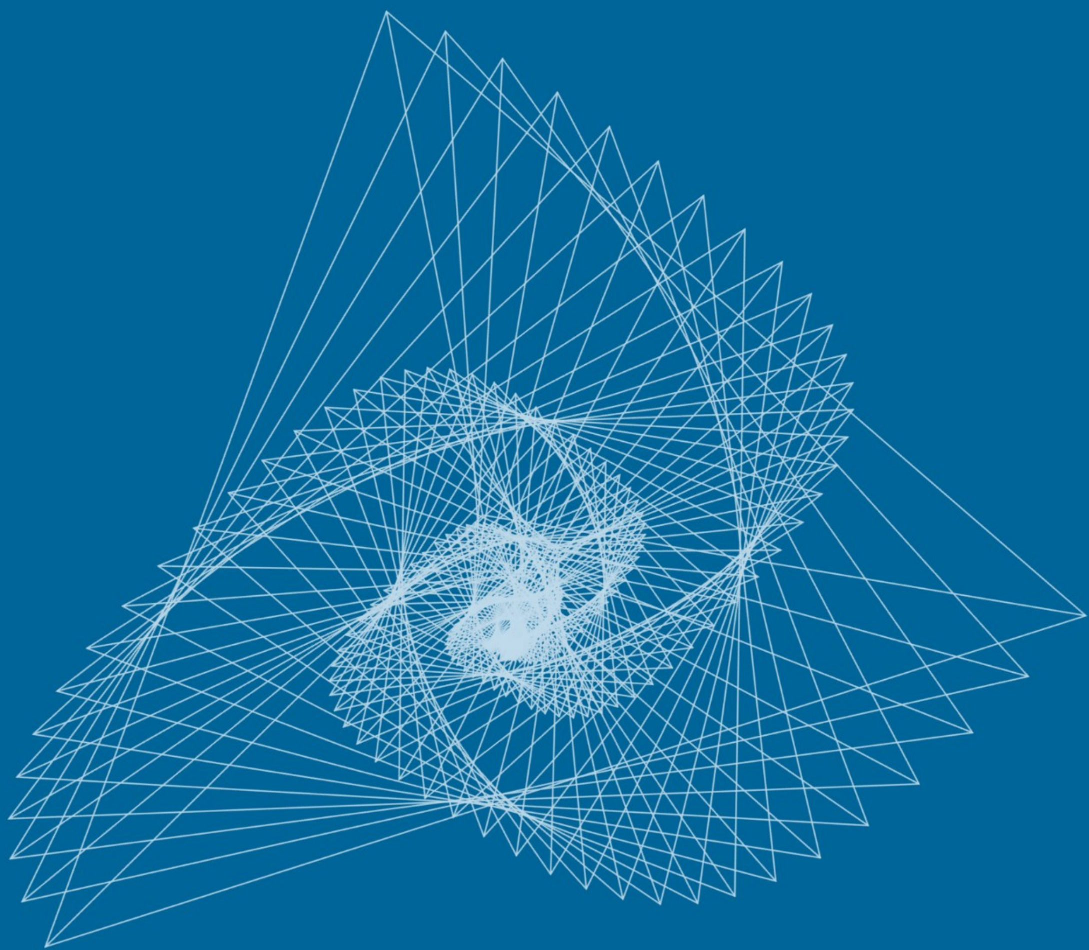
VOL. 8, NO. 1

# URAL MATHEMATICAL JOURNAL

N.N. Krasovskii Institute of Mathematics and Mechanics of
the Ural Branch of Russian Academy of Sciences and
Ural Federal University named after the first President of Russia B.N.Yeltsin

ISSN: 2414-3952



UMJURAN.RU
UMJ

2022

ISSN 2414-3952
(Online)

# Ural Mathematical Journal

## Vol. 8, No. 1, 2022

https://umjuran.ru

*Electronic Periodical Scientific Journal*
*Founded in 2015*

*The Journal is registered by the Federal Service for Supervision in the Sphere of Communication, Information Technologies and Mass Communications*
***Certificate of Registration of the Mass Media Эл № ФС77-61719 of 07.05.2015***

**Founders**

N.N. Krasovskii Institute of Mathematics and Mechanics of the Ural Branch of Russian Academy of Sciences

Ural Federal University named after the first President of Russia B.N. Yeltsin

**Contact Information**

16 S. Kovalevskaya str., Ekaterinburg, Russia, 620990
Phone: +7 (343) 375-34-73 Fax: +7 (343) 374-25-81
Email: secretary@umjuran.ru
Web-site: https://umjuran.ru

# TABLE OF CONTENTS

URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 3–12

DOI: 10.15826/umj.2022.1.001

# FIXED POINT THEOREM FOR MULTIVALUED NON-SELF MAPPINGS SATISFYING JS-CONTRACTION WITH AN APPLICATION

**David Aron**[†], **Santosh Kumar**[††]

Department of Mathematics, College of Natural and Applied Sciences,
University of Dar es Salaam, Tanzania

[†]d.aron.da6@gmail.com, [††]drsengar2002@gmail.com

**Abstract:** In this paper, we present some fixed point results for multivalued non-self mappings. We generalize the fixed point theorem due to Altun and Minak [2] by using Jleli and Sameti [9] $\vartheta$-contraction. To validate the results proved here, we provide an appropriate application of our main result.

**Keywords:** JS-contraction mapping, Multivalued mapping, Metric space, Non-self mapping, Fixed point.

## 1. Introduction and preliminaries

In 1922, in Banach's PhD thesis a remarkable fixed point theorem well known as the Banach contraction principal was initiated. It's simplicity, usefulness and application made it a supreme tool in finding the existence and uniqueness of solution in numerous branches of mathematical analysis and applied sciences. Following the Banach contraction principal, some authors, Nadler [13], Assad and Kirk [4], Itoh [8] and several others have extended and generalized this theorem in several ways. In fact, Nadler [13] introduced the concept of using Hausdorff metric on multi-valued contraction of self mappings in the study of fixed points. Assad and Kirk [4] proved the Banach contraction mapping theorem for multi-valued contraction of non-self mappings and Itoh [8] generalized the theorems due to Assad and Kirk, and many other researchers have made significant contributions in this area (see [3, 7, 11]). In 2013, Alghamdi et al. [1] proved fixed point results for multivalued nonself almost contractions on convex metric spaces. Recently, Altun and Minak [2] introduced a new approach to Assad and Kirk fixed point theorem and a new real generalization of it, by using $\vartheta-$ contractiveness of a multivalued mapping. Jleli and Samet [9] introduced $\vartheta-$ contraction and established a new fixed point theorem for such mappings in the setting of generalized metric spaces. Following the notion $\vartheta$, Hussain et al. [6] supposed that $\Theta$ is the set of all functions $\vartheta : [0,\infty) \to [1,\infty)$ satisfying the following conditions:

($\vartheta_1$) $\vartheta$ is nondecreasing and $\vartheta(t) = 1$ if and only if $t = 0$;
($\vartheta_2$) for each sequence $\{t_n\} \subseteq (0,\infty)$, $\lim\limits_{n\to\infty} \vartheta(t_n) = 1$ if and only if the limit of $\lim\limits_{n\to\infty} t_n = 0$;
($\vartheta_3$) there exists $r \in (0, 1)$ and $l \in (0,\infty]$ such that $\lim\limits_{t\to 0^+} \dfrac{\vartheta(t) - 1}{t^r} = l$;
($\vartheta_4$) $\vartheta(a + b) \le \vartheta(a)\vartheta(b)$ for all $a, b > 0$.

Throughout this paper we shall denote by $\Theta$ the set of all functions $\vartheta$ satisfying $(\vartheta_1) - (\vartheta_4)$.

Next, we present some definitions and preliminaries that are required to prove the main result of this paper.

Since we are dealing with multivalued mapping it is important to state a brief description of the Hausdorff metric. The Hausdorff metric measures the distance between subsets of a metric space. One among many interesting properties of this metric space, which will be our focus in this paper is that the Hausdorff induced metric space is complete if our original metric space is complete. Now, we define the Hausdorff metric as follows:

**Definition 1.** [5] *Let $(M, \varrho)$ be a metric space. Denote by $CB(M)$ the collection of non-empty closed bounded subsets of $M$. For $A, B \in CB(M)$ and $u \in M$, define*

$$\rho(u, A) = \inf_{a \in A} \varrho(u, a)$$

*and*

$$\mathcal{H}(A, B) = \max\big\{ \sup_{a \in A} \rho(a, B), \sup_{b \in B} \rho(b, A) \big\}.$$

*It is seen that $H$ is a metric on $CB(M)$. $H$ is called the Hausdorff metric induced by $\varrho$. The completion of $(M, \varrho)$ implies that $(CB(M), H)$ is a complete metric space.*

Before proceeding further, it is important to know that we will need an extra condition, call it $(\vartheta_5)$, a very useful part of our tool to help us to prove our main results in multivalued mapping and $\Theta$ satisfies $(\vartheta_5)$.

$(\vartheta_5)$ $\vartheta(\inf A) = \inf \vartheta(A)$ for all $A \subset (0, \infty)$ with $\inf A > 0$.

The following definition is important for future work in this paper.

**Definition 2.** [10].

(i) *A sequence $\{u_n\}$ in a metric space $(M, \varrho)$ is said to converge or to be convergent if there is $u \in M$ such that*

$$\lim_{n \to \infty} \varrho(u_n, u) = 0.$$

(ii) *A sequence $\{u_n\}$ in a metric space $(M, \varrho)$ is said to be Cauchy sequence if for every $\epsilon > 0$ there is a number $N = N(\epsilon)$ such that*

$$\varrho(u_n, u_m) < \epsilon$$

*for every $m, n > N$.*

(iii) *A metric space $(M, \varrho)$ is said to be complete if every Cauchy sequence in $M$ converges to an element of $M$.*

The following is the description of a metrically convex metric space and some of its properties are stated.

The following definition is due to Assad and Kirk [4].

**Definition 3.** [4] *A metric space $(M, \varrho)$ is said to be metrically convex if for any $u, v \in M$ with $u \neq v$, there exists a point $z \in M$,$(u \neq z \neq v)$ such that*

$$\varrho(u, v) = \varrho(u, z) + \varrho(z, v).$$

The following result is taken from Assad [4] where $\partial K$ denotes the boundary of $K$.

**Lemma 1.** [4] *If $K$ is a closed subset of the complete and convex metric space $M$ and if $u \in K$, $v \notin K$, then there exists a point $z \in \partial K$, such that*

$$\varrho(u,v) = \varrho(u,z) + \varrho(z,v),$$

*where $\partial K$ denotes the boundary of $K$.*

Assad and Kirk [4] proved the following fixed point theorem.

**Theorem 1.** [4] *Let $(M,\varrho)$ be a complete and metrically convex metric space, $K$ be a nonempty closed subset of $M$, $T : K \to CB(M)$ be a mapping such that, for all $u, v \in K$,*

$$\rho(Tu,Tv) \leq k\varrho(u,v),$$

*for some $k \in (0,1)$. If $Tu \subseteq K$ for each $u \in \partial K$, then $T$ has a fixed point in $K$.*

In 2014, Jleli and Sameti [9] gave a new generalization of Banach contraction mapping theorem in the setting of Banciari metric spaces as follows:

**Theorem 2.** [9] *Let $(M,\varrho)$ be a complete generalized metric space and $T : M \to M$ be a mapping. Suppose that there exist $\vartheta \in \Theta$ and $k \in (0,1)$ such that for $u, v \in M$,*

$$\varrho(Tu,Tv) \neq 0 \Longrightarrow \vartheta(\varrho(Tu,Tv)) \leq [\vartheta(\varrho(u,v))]^k.$$

*Then $T$ has a unique fixed point.*

Recently, Altun and Minak [2] obtained a new approach to Theorem 2 and a new generalization of it, by using $\vartheta$-contraction as follows:

**Theorem 3.** [2] *Let $(M,\varrho)$ be a complete and metrically convex metric space, $K$ be a nonempty closed subset of $M$, $T : K \to CB(M)$ be a mapping such that for all $u, v \in K$ with $\mathcal{H}(Tu,Tv) > 0$,*

$$\vartheta(\mathcal{H}(Tu,Tv)) \leq [\vartheta(\varrho(u,v))]^k$$

*for some $k \in (0,1)$ and $\vartheta \in \Theta$. If $Tu \subseteq K$ for each $u \in \partial K$, then $T$ has a fixed point in $K$.*

Suppose we want to use a different $\vartheta$-contraction in the above theorem. Hussain et al. [6] introduced a new concept that we can apply in the proof of the above theorem and obtain a new result.

**Definition 4.** [6] *Let $(M,\varrho)$ be a metric space and let $T : M \to M$ be a mapping. $T$ is said to be a JS-contraction whenever there is a function $\vartheta \in \Theta$ and positive real numbers $\tau_1, \tau_2, \tau_3, \tau_4$ with $0 \leq \tau_1 + \tau_2 + \tau_3 + 2\tau_4 < 1$ such that*

$$\vartheta(\varrho(Tu,Tv)) \leq [\vartheta(\varrho(u,v))]^{\tau_1}[\vartheta(\varrho(u,Tu))]^{\tau_2}[\vartheta(\varrho(v,Tv))]^{\tau_3}[\vartheta(\varrho(u,Tv)+\varrho(v,Tu))]^{\tau_4}$$

*for all $u, v \in M$.*

## 2. JS-contraction fixed point theorem

We give now a definition of a generalized multivalued JS-contraction mapping.

**Definition 5.** *Let $(M, \varrho)$ be a metric space and $K$ be a nonempty closed subset of $M$. Let $T$ be a mapping of $K$ into $CB(M)$. Then $T$ is said to be a generalized JS-contraction mapping whenever there is a function $\vartheta \in \Theta$ and nonnegative real numbers $\tau_1, \tau_2, \tau_3, \tau_4$ with*

$$0 \leq \tau_1 + \tau_2 + \tau_3 + 2\tau_4 < 1$$

*such that*

$$\vartheta(\mathcal{H}(Tu, Tv)) \leq [\vartheta(\varrho(u, v))]^{\tau_1} [\vartheta(\varrho(u, Tu))]^{\tau_2} [\vartheta(\varrho(v, Tv))]^{\tau_3} [\vartheta(\varrho(u, Tv) + \varrho(v, Tu))]^{\tau_4}. \tag{2.1}$$

*for all $u, v \in K$.*

We now present an extended version of Theorem 3.

**Theorem 4.** *Let $(M, \varrho)$ be a complete and metrically convex metric space, $K$ be a nonempty closed subset of $M$. Let $T : K \to CB(M)$ be a generalized multivalued JS-contraction mapping. If for any $u \in \partial K$, $Tu \subseteq K$ and*

$$\frac{(1 + \tau_1 + \tau_2 + \tau_4)(\tau_1 + \tau_2 + \tau_4)}{(1 - \tau_3 - \tau_4)^2} < 1,$$

*then there is $z \in K$ such that $z \in T(z)$.*

P r o o f. We construct two sequences $\{u_n\}$ and $\{v_n\}$ in $K$ in the following way: let $u_0 \in K$ and $v_1 \in Tu_0$. If $v_1 \in K$, let $u_1 = v_1$. If $v_1 \notin K$, then from Lemma 1, there exists $u_1 \in \partial K$ such that

$$\varrho(u_0, u_1) + \varrho(u_1, v_1) = \varrho(u_0, v_1).$$

Thus, $u_1 \in K$. Now, we claim that $\varrho(v_1, Tu_1) \geq 0$. Suppose $\varrho(v_1, Tu_1) = 0$. If $v_1 \in K$, then $u_1$ is a fixed point of $T$, which is a contradiction. If $v_1 \notin K$, then $u_1 \in \partial K$ and so $Tu_1 \subseteq K$. Therefore, $v_1 \notin Tu_1$, which is a contradiction. Thus, $\varrho(v_1, Tu_1) \geq 0$. Now, since $\varrho(v_1, Tu_1) \leq \mathcal{H}(Tu_0, Tu_1)$, then we have

$$\begin{aligned} \vartheta(\varrho(v_1, Tu_1)) &\leq \vartheta(\mathcal{H}(Tu_0, Tu_1)) \\ &\leq [\vartheta(\varrho(u_0, u_1))]^{\tau_1} [\vartheta(\varrho(u_0, Tu_0))]^{\tau_2} [\vartheta(\varrho(u_1, Tu_1))]^{\tau_3} \\ &\quad \times [\vartheta(\varrho(u_0, Tu_1) + \varrho(u_1, Tu_0))]^{\tau_4}. \end{aligned} \tag{2.2}$$

On the other hand, from $\vartheta_5$ we get

$$\vartheta(\varrho(v_1, Tu_1)) = \vartheta(\inf\{\varrho(v_1, m) : m \in Tu_1\}) = \inf\{\vartheta(\varrho(v_1, m)) : m \in Tu_1\}$$

and so from condition (2.2) we get

$$\begin{aligned} \inf\{\vartheta(\varrho(v_1, m)) : m \in Tu_1\} \leq\ & [\vartheta(\varrho(u_0, u_1))]^{\tau_1} [\vartheta(\varrho(u_0, Tu_0))]^{\tau_2} [\vartheta(\varrho(u_1, Tu_1))]^{\tau_3} \\ & \times [\vartheta(\varrho(u_0, Tu_1) + \varrho(u_1, Tu_0))]^{\tau_4}. \end{aligned}$$

Thus, there exists $v_2 \in Tu_1$ such that

$$\vartheta(\varrho(v_1, v_2)) \leq [\vartheta(\varrho(u_0, u_1))]^{\gamma_1} [\vartheta(\varrho(u_0, Tu_0))]^{\gamma_2} [\vartheta(\varrho(u_1, Tu_1))]^{\gamma_3} [\vartheta(\varrho(u_0, Tu_1) + \varrho(u_1, Tu_0))]^{\gamma_4},$$

where

$$0 \le \tau_1 + \tau_2 + \tau_3 + 2\tau_4 < \gamma_1 + \gamma_2 + \gamma_3 + 2\gamma_4 < 1.$$

If $v_2 \in K$ let $u_2 = v_2$. If $v_2 \notin K$, select a point $u_2 \in \partial K$ such that

$$\varrho(u_1, u_2) + \varrho(u_2, v_2) = \varrho(u_1, v_2).$$

Thus, $u_2 \in K$. We can show that $\varrho(v_2, Tu_2) > 0$. As above, we can find a point $v_3 \in Tu_2$ such that

$$\vartheta(\varrho(v_2, v_3)) \le [\vartheta(\varrho(u_1, u_2))]^{\gamma_1} [\vartheta(\varrho(u_1, Tu_1))]^{\gamma_2} [\vartheta(\varrho(u_2, Tu_2))]^{\gamma_3} [\vartheta(\varrho(u_1, Tu_2) + \varrho(u_2, Tu_1))]^{\gamma_4}.$$

Continuing the arguments, two sequences $\{u_n\}$ and $\{v_n\}$ are obtained such that for $n \in N$ we have

(i) $v_{n+1} \in Tu_n$,

(ii) $$\vartheta(\varrho(v_n, v_{n+1})) \le [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_1} [\vartheta(\varrho(u_{n-1}, Tu_{n-1}))]^{\gamma_2} [\vartheta(\varrho(u_n, Tu_n))]^{\gamma_3} \times [\vartheta(\varrho(u_{n-1}, Tu_n) + \varrho(u_n, Tu_{n-1}))]^{\gamma_4},$$

where $v_{n+1} = u_{n+1}$ if $v_{n+1} \in K$ or

$$\varrho(u_n, u_{n+1}) + \varrho(u_{n+1}, v_{n+1}) = \varrho(u_n, v_{n+1}) \tag{2.3}$$

if $v_{n+1} \notin K$ and $u_{n+1} \in \partial K$.

Now, we consider sets

$$P = \{u_\xi \in \{u_n\} : u_\xi = v_\xi,\ \xi \in \mathbb{N}\}, \quad Q = \{u_\xi \in \{u_n\} : u_\xi \neq v_\xi,\ \xi \in \mathbb{N}\}.$$

Observe that if $u_\xi \in Q$ for some $\xi$, then $u_{\xi+1} \in P$. Here, the intention is to estimate the distance $\varrho(u_n, u_{n+1})$ for $n \ge 2$. Note that $\varrho(u_n, u_{n+1}) > 0$, otherwise, $T$ has a fixed point. For this, three cases have to be considered:

**Case 1.** If $u_n \in P$ and $u_{n+1} \in P$, then, we get

$$\begin{aligned}
\vartheta(\varrho(u_n, u_{n+1})) &= \vartheta(\varrho(v_n, v_{n+1})) \\
&\le [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_1} [\vartheta(\varrho(u_{n-1}, Tu_{n-1}))]^{\gamma_2} [\vartheta(\varrho(u_n, Tu_n))]^{\gamma_3} \\
&\quad \times [\vartheta(\varrho(u_{n-1}, Tu_n) + \varrho(u_n, Tu_{n-1}))]^{\gamma_4} \\
&= [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_1} [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_2} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3} \\
&\quad \times [\vartheta(\varrho(u_{n-1}, u_{n+1}) + \varrho(u_n, u_n))]^{\gamma_4} \\
&= [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_1+\gamma_2} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3} [\vartheta(\varrho(u_{n-1}, u_n) + \varrho(u_n, u_{n+1}))]^{\gamma_4} \\
&\le [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_1+\gamma_2} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3} [\vartheta(\varrho(u_{n-1}, u_n)]^{\gamma_4} [\varrho(u_n, u_{n+1}))]^{\gamma_4} \\
&= [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_1+\gamma_2+\gamma_4} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3+\gamma_4}.
\end{aligned}$$

It follows that

$$\vartheta(\varrho(u_n, u_{n+1})) \le [\vartheta(\varrho(u_{n-1}, u_n))]^{\frac{\gamma_1+\gamma_2+\gamma_4}{1-\gamma_3-\gamma_4}}.$$

**Case 2.** If $u_n \in P$ and $u_{n+1} \in Q$, then, from condition (2.3), we get

$$\vartheta(\varrho(u_n, u_{n+1})) \le \vartheta(\varrho(u_n, u_{n+1}) + \varrho(u_{n+1}, v_{n+1})) = \vartheta(\varrho(v_n, v_{n+1})) \le [\vartheta(\varrho(u_{n-1}, u_n))]^{\frac{\gamma_1+\gamma_2+\gamma_4}{1-\gamma_3-\gamma_4}}.$$

**Case 3.** If $u_n \in Q$ and $u_{n+1} \in P$, then, since

$$\begin{aligned}\vartheta(\varrho(v_n, v_{n+1})) \leq [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_1} [\vartheta(\varrho(u_{n-1}, Tu_{n-1}))]^{\gamma_2} [\vartheta(\varrho(u_n, Tu_n))]^{\gamma_3} \\ \times [\vartheta(\varrho(u_{n-1}, Tu_n) + \varrho(u_n, Tu_{n-1}))]^{\gamma_4},\end{aligned}$$

$\Rightarrow$

$$\varrho(v_n, v_{n+1}) < (\varrho(u_{n-1}, u_n))(\varrho(u_{n-1}, Tu_{n-1}))(\varrho(u_n, Tu_n))(\varrho(u_{n-1}, Tu_n) + \varrho(u_n, Tu_{n-1})).$$

In our case, if we simplify we get

$$\begin{aligned}\vartheta(\varrho(v_n, v_{n+1})) &\leq [\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma_1} [\vartheta(\varrho(u_{n-1}, v_n))]^{\gamma_2} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3} \\ &\quad \times [\vartheta(\varrho(u_{n-1}, u_{n+1}) + \varrho(u_n, v_n))]^{\gamma_4} \\ &\leq [\vartheta(\varrho(u_{n-1}, u_n) + \varrho(u_n, v_n))]^{\gamma_1} [\vartheta(\varrho(u_{n-1}, v_n))]^{\gamma_2} \\ &\quad \times [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3} [\vartheta(\varrho(u_{n-1}, u_n) + \varrho(u_n, u_{n+1}) + \varrho(u_n, v_n))]^{\gamma_4} \\ &\leq [\vartheta(\varrho(u_{n-1}, v_n))]^{\gamma_1} [\vartheta(\varrho(u_{n-1}, v_n))]^{\gamma_2} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3} \\ &\quad \times [\vartheta(\varrho(u_{n-1}, v_n) + \varrho(u_n, u_{n+1}))]^{\gamma_4} \\ &\leq [\vartheta(\varrho(u_{n-1}, v_n))]^{\gamma_1+\gamma_2} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3} \\ &\quad \times [\vartheta(\varrho(u_{n-1}, v_n)]^{\gamma_4} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_4} \\ &= [\vartheta(\varrho(u_{n-1}, v_n))]^{\gamma_1+\gamma_2+\gamma_4} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3+\gamma_4}.\end{aligned}$$

It follows that

$$\begin{aligned}\vartheta(\varrho(u_n, u_{n+1})) &\leq \vartheta(\varrho(u_n, v_n) + \varrho(v_n, u_{n+1})) \\ &< \vartheta(\varrho(u_{n-1}, u_n) + \varrho(u_n, v_n) + \varrho(v_n, v_{n+1})) \\ &= \vartheta(\varrho(u_{n-1}, v_n) + \varrho(v_n, v_{n+1})) \\ &\leq \vartheta(\varrho(u_{n-1}, v_n))\vartheta(\varrho(v_n, v_{n+1})) \\ &\leq \vartheta(\varrho(u_{n-1}, v_n))[\vartheta(\varrho(u_{n-1}, v_n))]^{\gamma_1+\gamma_2+\gamma_4} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3+\gamma_4} \\ &= [\vartheta(\varrho(u_{n-1}, v_n))]^{1+\gamma_1+\gamma_2+\gamma_4} [\vartheta(\varrho(u_n, u_{n+1}))]^{\gamma_3+\gamma_4}.\end{aligned}$$

Hence

$$[\vartheta(\varrho(u_n, u_{n+1}))]^{1-\gamma_3-\gamma_4} \leq [\vartheta(\varrho(u_{n-1}, v_n))]^{1+\gamma_1+\gamma_2+\gamma_4}$$

by Case 2, since $u_n \in Q$ implies $u_{n-1} \in P$ we have

$$\vartheta(\varrho(u_{n-1}, v_n)) \leq [\vartheta(\varrho(u_{n-2}, u_{n-1}))]^{\frac{\gamma_1+\gamma_2+\gamma_4}{1-\gamma_3-\gamma_4}}.$$

Therefore,

$$\vartheta(\varrho(u_n, u_{n+1})) \leq [\vartheta(\varrho(u_{n-2}, u_{n-1}))]^{\frac{(1+\gamma_1+\gamma_2+\gamma_4)(\gamma_1+\gamma_2+\gamma_4)}{(1-\gamma_3-\gamma_4)^2}}.$$

The case that $u_n \in Q$ and $u_{n+1} \in Q$ does not occur.

Since

$$\frac{\gamma_1+\gamma_2+\gamma_4}{1-\gamma_3-\gamma_4} \leq \frac{(1+\gamma_1+\gamma_2+\gamma_4)(\gamma_1+\gamma_2+\gamma_4)}{(1-\gamma_3-\gamma_4)^2},$$

for $n \geq 2$ we have

$$\vartheta(\varrho(u_n, u_{n+1})) \leq \begin{cases}[\vartheta(\varrho(u_{n-1}, u_n))]^{\gamma}, \\ [\vartheta(\varrho(u_{n-2}, u_{n-1}))]^{\gamma}.\end{cases}$$

Now we claim that

$$\vartheta(\varrho(u_n, u_{n+1})) \leq \delta^{\left(\gamma^{\frac{n-1}{2}}\right)} \tag{2.4}$$

for all $n \in \mathbb{N}$, where

$$\delta = \max\{\vartheta(\varrho(u_0, u_1)), \vartheta(\varrho(u_1, u_2))\}.$$

Using (2.4) we obtain

$$\lim_{n\to\infty} \vartheta(\varrho(u_n, u_{n+1})) = 1.$$

From $(\vartheta_2)$, $\lim_{n\to\infty} \varrho(u_n, u_{n+1}) = 0$ and so from $(\vartheta_3)$ there exists $r \in (0, 1)$ and $l \in (0, \infty]$ such that

$$\lim_{n\to\infty} \frac{\vartheta(\varrho(u_n, u_{n+1})) - 1}{[\varrho(u_n, u_{n+1})]^r} = l.$$

Suppose that $l < \infty$. In this case, let $\Psi = l/2 > 0$. Recall from the definition of the limit, there exists $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,

$$\left| \frac{\vartheta(\varrho(u_n, u_{n+1})) - 1}{[\varrho(u_n, u_{n+1})]^r} - l \right| \leq \Psi.$$

This implies that, for all $n \geq n_0$,

$$\frac{\vartheta(\varrho(u_n, u_{n+1})) - 1}{[\varrho(u_n, u_{n+1})]^r} \geq l - \Psi = \Psi.$$

Then, for all $n \geq n_0$,

$$n[\varrho(u_n, u_{n+1})]^r \leq \Phi n[\vartheta(\varrho(u_n, u_{n+1})) - 1],$$

where $\Phi = 1/\vartheta$. Thus, in all cases, there exist $\Phi > 0$ and $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,

$$n[\varrho(u_n, u_{n+1})]^r \leq \Phi n[\vartheta(\varrho(u_n, u_{n+1})) - 1].$$

Using (2.4), we obtain, for all $n \geq n_0$,

$$n[\varrho(u_n, u_{n+1})]^r \leq \Phi n[\delta^{\left(\gamma^{\frac{n-1}{2}}\right)} - 1].$$

Letting $n \to \infty$ in the above inequality, we obtain

$$\lim_{n\to\infty} n[\varrho(u_n, u_{n+1})]^r = 0.$$

Thus, there exists $n_1 \in \mathbb{N}$ such that $n[\varrho(u_n, u_{n+1})]^r \leq 1$ for all $n \geq n_1$. So, we have, for all $n \geq n_0$

$$\varrho(u_n, u_{n+1}) \leq \frac{1}{n^{1/r}}. \tag{2.5}$$

In order to demonstrate that $\{x_n\}$ is a Cauchy sequence consider $m, n \in \mathbb{N}$ such that $m > n \geq n_1$. Now, applying the metric triangle inequality and from condition (2.5), we have

$$\begin{aligned}\varrho(u_n, u_m) \leq& \varrho(u_n, u_{n+1}) + \varrho(u_{n+1}, u_{n+2}) + \cdots + \varrho(u_{m-1}, u_m) \\ =& \sum_{\xi=n}^{m-1} \varrho(u_\xi, u_{\xi+1}) \leq \sum_{\xi=n}^{\infty} \varrho(u_\xi, u_{\xi+1}) \leq \sum_{\xi=n}^{\infty} \frac{1}{\xi^{1/r}}.\end{aligned}$$

Since the series $\sum_{\xi=1}^{\infty} \frac{1}{\xi^{1/r}}$ converges, then passing to limit with $n, m \to \infty$, we get $\varrho(u_n, u_m) \to 0$.

It is an obvious implication that the sequence $\{u_n\}$ is a Cauchy sequence in $K$. Since $K$ is closed, the sequence $\{u_n\}$ converges to some point $z \in K$. By our choice of $\{u_n\}$, there exists a subsequence $\{u_{n_k}\}$ of $\{u_n\}$ such that $\{u_{n_k}\} \in P$ that is, $\{u_{n_k}\} = \{v_{n_k}\}$, $k \in \mathbb{N}$. Note that

$\{u_{n_k}\} \in \{u_{n_k-1}\}$ for $k \in \mathbb{N}$ and $\{u_{n_k}\} \to z$ as $k \to \infty$. Also note that from condition (2.1) and $(\vartheta_1)$ we get

$$\mathcal{H}(Tu, Tv) \leq (\varrho(u, v))(\varrho(u, Tu))(\varrho(v, Tv))(\varrho(v, Tu) + \varrho(u, Tv))$$

for all $u, v \in K$ and so, we have

$$\begin{aligned} \varrho(u_{n_k}, Tz) &\leq \mathcal{H}(Tu_{n_k-1}, Tz) \\ &\leq \varrho(u_{n_k-1}, z)\varrho(u_{n_k-1}, Tu_{n_k-1})\varrho(z, Tz)\{\varrho(z, Tu_{n_k-1}) + \varrho(u_{n_k-1}, Tz)\} \end{aligned}$$

which on letting $k \to \infty$ implies that $\varrho(z, Tz) = 0$, which is a contradiction. Therefore, $T$ has a fixed point $z \in K$. □

*Remark 1.* If $\tau_2 = \tau_3 = \tau_4 = 0$ and $\tau_1 = \tau$ in Theorem 4 we obtain Theorem 3 of Altun [2].

For particular function $\vartheta$ selections, some significant results are obtained. First, by setting $\vartheta(\mu) = e^{\sqrt{\mu}}$ in Theorem 4, the following corollary is obtained:

**Corollary 1.** *Let $K$ be a nonempty closed subset of a complete and metrically convex metric space $M$. Let $T : K \to CB(M)$ be a mapping such that the following condition holds:*

$$\sqrt{\mathcal{H}(Tu, Tv)} \leq \tau_1\sqrt{\varrho(u, v)} + \tau_2\sqrt{\varrho(u, Tu)} + \tau_3\sqrt{\varrho(v, Tv)} + \tau_4\sqrt{\varrho(u, Tv) + \varrho(v, Tu)}$$

*for all $u, v \in K$, $\vartheta \in \Theta$ and $\tau_1, \tau_2, \tau_3, \tau_4 \geq 0$ with $0 \leq \tau_1 + \tau_2 + \tau_3 + 2\tau_4 < 1$. Then $T$ has a unique fixed point.*

And, by putting $\vartheta(\mu) = e^{\sqrt[n]{\mu}}$ in Theorem 4, the following corollary is obtained:

**Corollary 2.** *Let $K$ be a nonempty closed subset of a complete and metrically convex metric space $M$. Let $T : K \to CB(M)$ be a mapping such that the following condition holds*:

$$\sqrt[n]{\mathcal{H}(Tu, Tv)} \leq \tau_1\sqrt[n]{\varrho(u, v)} + \tau_2\sqrt[n]{\varrho(u, Tu)} + \tau_3\sqrt[n]{\varrho(v, Tv)} + \tau_4\sqrt[n]{\varrho(u, Tv) + \varrho(v, Tu)}$$

*for all $u, v \in K$, $\vartheta \in \Theta$ and $\tau_1, \tau_2, \tau_3, \tau_4 \geq 0$ with $0 \leq \tau_1 + \tau_2 + \tau_3 + 2\tau_4 < 1$. Then $T$ has a unique fixed point.*

## 3. Application to nonlinear integral equations

Nonlinear integral equations can be solved using a variety of numerical approaches. The integral equation is usually transformed into a system of nonlinear algebraic equations. Solving these systems is difficult, or the solution may be impossible to find. Therefore, in this section, we describe how the fixed point approach may be used to solve Volterra–Hammerstein integral equations. This approach does not result in a system of nonlinear algebraic equations.

Now, consider the nonlinear integral equation below:

$$u(t) = g(t) + \int_a^b k(t, \tau)\mathcal{H}(\tau, u(\tau))d\tau, \tag{3.1}$$

where $t, \tau \in [a, b]$, $a, b \in \mathbb{R}$, $u \in C[a, b]$, $g : [a, b] \to \mathbb{R}$, $\mathcal{H} \in C[a, b] \times \mathbb{R} \to \mathbb{R}$ and $k \in C^2[a, b]^2$ such that $k(t, \tau) > 0$ are given functions.

Maleknejad [12] established some conditions which ensure the uniqueness of the solution and how the fixed point method approximates this solution.

Referring from Maleknejad [12] we are going to establish the following fixed point theorem.

**Theorem 5.** *Let $M = C[a,b]$ be a metric space endowed with the metric*

$$\varrho(u,v) = \sup_{t\in[a,b]} |u(t) - v(t)|.$$

*Define the mapping $T : K \to CB(M)$ by*

$$T(u)(t) = g(t) + \int_a^b k(t,\tau)\mathcal{H}(\tau, u(\tau))d\tau.$$

*Let $u, v \in K$ and $t \in [a,b]$. Assume that $g \in C[a,b]$, $k \in C^2[a,b]^2$, i.e. there exists a constant $M > 0$ where*

$$\left(\int_a^b k^2(t,\tau)d\tau\right)^{\frac{1}{2}} \le M < \infty,$$

*and $\mathcal{H} : [a,b] \times \mathbb{R} \to \mathbb{R}$ is continuous and there is $\vartheta \in \Theta$ so that $\vartheta(\sup f(t)) = \sup \vartheta(f(t))$ for arbitrary function $f$ with*

$$\vartheta\bigg(\int_a^b |\mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau))|d\tau\bigg) \le \int_a^b \vartheta(|\mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau))|)d\tau,$$

*there is $\tau_i \in (0,1)$ where $i = 1,2,3,4$ such that*

$$\begin{aligned}\vartheta\big(\big|\mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau))\big|\big) \le \bigg\{&\big[\vartheta(|u(t)-v(t)|)\big]^{\tau_1}\Big[\vartheta(|u(t) - \int_a^b k(t,\tau)\mathcal{H}(\tau,u(\tau))d\tau|)\Big]^{\tau_2}\Big[\vartheta(|v(t)\\ &- \int_a^b k(t,\tau)\mathcal{H}(\tau,v(\tau))d\tau|)\Big]^{\tau_3}\Big[\vartheta(|u(t) - \int_a^b k(t,\tau)\mathcal{H}(\tau,v(\tau))d\tau| + |v(t)\\ &- \int_a^b k(t,\tau)\mathcal{H}(\tau,u(\tau))d\tau|)\Big]^{\tau_4}\bigg\}\Big/ M(b-a).\end{aligned}$$

*Then equation (3.1) has a unique solution.*

P r o o f. We begin our proof by deriving the following relation where Cauchy–Schwartz inequality is used:

$$\begin{aligned}|Tu(t) - Tv(t)| &= \bigg|\int_a^b k(t,\tau)(\mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau)))d\tau\bigg|\\ &\le \int_a^b |k(t,\tau)||\mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau)))d\tau|\\ &\le \bigg(\int_a^b k^2(t,\tau)d\tau\bigg)^{1/2}\bigg(\int_a^b \mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau)))d\tau\bigg)^{1/2}\\ &\le M\bigg(\int_a^b \mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau)))d\tau\bigg)^{1/2}\\ &\le M\int_a^b \big|\mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau)))\big|d\tau.\end{aligned}$$

Now, we have

$$\begin{aligned}\vartheta(|Tu(t) - Tv(t)|) &= \vartheta\bigg(M\int_a^b |\mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau))|d\tau\bigg)\\ &\le M\int_a^b \vartheta\big(|\mathcal{H}(\tau,u(\tau)) - \mathcal{H}(\tau,v(\tau))|\big)d\tau\end{aligned}$$

$$
\begin{aligned}
&\leq \Big\{ M\big[\vartheta(|u(t)-v(t)|)\big]^{\tau_1}\big[\vartheta(|u(t)-\int_a^b k(t,\tau)\mathcal{H}(\tau,u(\tau))d\tau|)\big]^{\tau_2} \\
&\quad \big[\vartheta(|v(t)-\int_a^b k(t,\tau)\mathcal{H}(\tau,v(\tau))d\tau|)\big]^{\tau_3}\big[\vartheta(|u(t)-\int_a^b k(t,\tau)\mathcal{H}(\tau,v(\tau))d\tau| \\
&\qquad + |v(t)-\int_a^b k(t,\tau)\mathcal{H}(\tau,u(\tau))d\tau|)\big]^{\tau_4}\Big\} \Big/ M(b-a) \\
&\leq \frac{1}{b-a}\int_a^b \big[\vartheta(\varrho(u,v))\big]^{\tau_1}\big[\vartheta(\varrho(u,Tu))\big]^{\tau_2}\big[\vartheta(\varrho(v,Tv))\big]^{\tau_3}\big[\vartheta(\varrho(u,Tv)+\varrho(v,Tu))\big]^{\tau_4} \\
&= \big[\vartheta(\varrho(u,v))\big]^{\tau_1}\big[\vartheta(\varrho(u,Tu))\big]^{\tau_2}\big[\vartheta(\varrho(v,Tv))\big]^{\tau_3}\big[\vartheta(\varrho(u,Tv)+\varrho(v,Tu))\big]^{\tau_4}.
\end{aligned}
$$

Thus, all the conditions of Theorem 4 are satisfied. Hence the integral equation (3.1) has a solution. □

## 4. Conclusion

The main contribution of this study is Definition 5 and Theorem 4. This theorem is proved for multivalued non-self mappings in complete and metrically convex space. This theorem generalizes the fixed point theorem due to Altun and Minak [2] by using $\vartheta$-contraction due to Jleli and Sameti [9]. To validate the results proved here, we provide an appropriate application of our main result.

### REFERENCES


1. Alghamdi M. A., Berinde V., Shahzad N. Fixed points of multivalued nonself almost contractions. *J. Appl. Math*, 2013. Vol. 2013. Art. ID 621614. 6 p. DOI: 10.1155/2013/621614
2. Altun I., Minak G. An extension of Assad-Kirk's fixed point theorem for multivalued nonself mappings. *Carpathian J. Math.*, 2016. Vol. 32, No. 2. P. 147–155. URL: https://www.jstor.org/stable/44000101
3. Aron D., Kumar S. Fixed point theorem for a sequence of multivalued nonself mappings in metrically convex metric spaces. *Topol. Algebra Appl.*, 2022. Vol. 10 P. 1–12. DOI: 10.1515/taa-2020-0108
4. Assad N. A., Kirk W. A. Fixed point theorems for set-valued mappings of contractive type. *Pacific J. Math.*, 1972. Vol. 43. P. 553–562. DOI: 10.2140/pjm.1972.43.553
5. Damjanović B., Samet B., Vetro C. Common fixed point theorems for multi-valued maps. *Acta Math. Sci. Ser. B Engl. Ed.*, 2012. Vol. 32, No. 2. P. 818–824. DOI: 10.1016/S0252-9602(12)60063-0
6. Hussain N., Parvaneh V., Samet B., Vetro C. Some fixed point theorems for generalized contractive mappings in complete metric spaces. *Fixed Point Theory Appl.*, 2015. Vol. 2015. Art. no. 185. 17 p.DOI: 10.1186/s13663-015-0433-z
7. Imdad M., Kumar S. Rhoades-type fixed-point theorems for a pair of nonself mappings. *Comput. Math. Appl.*, 2003. Vol. 46, No. 5–6. P. 919–927. DOI: 10.1016/S0898-1221(03)90153-2
8. Itoh S. Multivalued generalized contractions and fixed point theorems. *Comment. Math. Univ. Carolin.*, 1977. Vol. 018, No. 2. P. 247–258.
9. Jleli M., Samet B. A new generalization of the Banach contraction principle. *J. Inequal. Appl.*, 2014. Vol. 2014. Art. no. 38. 8 p. DOI: 10.1186/1029-242X-2014-38
10. Kreyszig E. *Introductory Functional Analysis with Applications.* NY: John Wiley & Sons. Inc., 1978. 688 p.
11. Kumar S., Rugumisa T., Imdad M. Common fixed points in metrically convex partial metric spaces. *Konuralp J. Math.*, 2017. Vol. 5, No. 2. P. 56–71.
12. Maleknejad K., Torabi P. Application of fixed point method for solving nonlinear Volterra–Hammerstein integral equation. *U.P.B. Sci. Bull. Ser. A*, 2012. Vol. 74, No. 1. P. 45–56.
13. Nadler S. B., Jr. Multi-valued contraction mappings. *Pacific. J. Math.*, 1969. Vol. 30, No. 2. P. 475–488. DOI: 10.2140/pjm.1969.30.475

URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 13–22
DOI: 10.15826/umj.2022.1.002

# ON $A^{\mathcal{I}^{\mathcal{K}}}$–SUMMABILITY

**Chiranjib Choudhury**[†], **Shyamal Debnath**[††]

Tripura University (A Central University),
Suryamaninagar-799022, Agartala, India

[†]chiranjibchoudhury123@gmail.com, chiranjib.mathematics@tripurauniv.in
[††]shyamalnitamath@gmail.com

**Abstract:** In this paper, we introduce and investigate the concept of $A^{\mathcal{I}^{\mathcal{K}}}$-summability as an extension of $A^{\mathcal{I}^*}$-summability which was recently (2021) introduced by O.H.H. Edely, where $A=(a_{nk})_{n,k=1}^{\infty}$ is a non-negative regular matrix and $\mathcal{I}$ and $\mathcal{K}$ represent two non-trivial admissible ideals in $\mathbb{N}$. We study some of its fundamental properties as well as a few inclusion relationships with some other known summability methods. We prove that $A^{\mathcal{K}}$-summability always implies $A^{\mathcal{I}^{\mathcal{K}}}$-summability whereas $A^{\mathcal{I}}$-summability not necessarily implies $A^{\mathcal{I}^{\mathcal{K}}}$-summability. Finally, we give a condition namely $AP(\mathcal{I},\mathcal{K})$ (which is a natural generalization of the condition $AP$) under which $A^{\mathcal{I}}$-summability implies $A^{\mathcal{I}^{\mathcal{K}}}$-summability.

**Keywords:** Ideal, Filter, $\mathcal{I}$-convergence, $\mathcal{I}^{\mathcal{K}}$-convergence, $A^{\mathcal{I}}$-summa-bility, $A^{\mathcal{I}^{\mathcal{K}}}$-summability.

## 1. Introduction

In 2000, Kostrkyo and Salat [12] introduced the notion of ideal convergence. They studied several fundamental properties of $\mathcal{I}$ and $\mathcal{I}^*$-convergence and showed that their idea was the extended version of so many known convergence methods. Based on $\mathcal{I}$-convergence several generalizations were made by researchers and several analytical and topological properties have been investigated (see [1, 9, 11, 15–19, 21, 22] where many more references can be found) and this area becomes one of the most focused areas of research.

In 2011, M. Macaj and M. Sleziak [13] generalized the idea of $\mathcal{I}^*$-convergence to $\mathcal{I}^{\mathcal{K}}$-convergence by involving two ideals $\mathcal{I}$ and $\mathcal{K}$. In the case of $\mathcal{I}^{\mathcal{K}}$-convergence, the convergence along the large set is taken with regard to another ideal $\mathcal{K}$ instead of considering ordinary convergence. So from that point of view the concept of $\mathcal{I}^{\mathcal{K}}$-convergence being an extension of $\mathcal{I}^*$-convergence shows a strong analogy for further investigation. Recent developments in the direction of $\mathcal{I}^{\mathcal{K}}$-convergence from topological aspects can be found from the works of Das et al. [4, 5], Banerjee and Paul [2, 3] and many others.

If $x=(x_k)$ be a real-valued sequence and $A=(a_{nk})_{n,k=1}^{\infty}$ be an infinite matrix, then $Ax$ is the sequence having $n^{th}$ term $A_n(x)=\sum_{k=1}^{\infty}a_{nk}x_k$. A sequence $x=(x_k)$ is said to be $A$-summable to $L$, if $\lim\limits_{n\to\infty}A_n(x)=L$. A matrix $A=(a_{nk})_{n,k=1}^{\infty}$ is said to be regular if it maps a convergent sequence into a convergent sequence keeping the same limit i.e., $A\in(c,c)_{reg}$ if $A\in(c,c)$ and $\lim\limits_{n\to\infty}A_n(x)=\lim\limits_{k\to\infty}x_k$. Here $c$, $(c,c)$, and $(c,c)_{reg}$ denote the collection of all real-valued convergent sequences, collection of all matrices which maps an element of $c$ to an element of $c$, and the collection of all regular matrices which maps an element of $c$ to an element of $c$, respectively. The necessary and sufficient Silverman–Toeplitz conditions for an infinite matrix $A=(a_{nk})_{n,k=1}^{\infty}$ to be regular are as follows:

(i) $\sup\limits_{n}\sum_{k=1}^{\infty}|a_{nk}|<\infty$;

(ii) For any $k \in \mathbb{N}$, $\lim_{n\to\infty} a_{nk} = 0$;

(iii) $\lim_{n\to\infty} \sum_{k=1}^{\infty} a_{nk} = 1$.

In 2008, Edely and Mursaleen [7] generalized the notion of $A$-summability to statistical $A$-summability by using the concept of natural density. Recently, Edely [6] further extended the notion of statistical $A$-summability to $A^{\mathcal{I}}$-summability, where $\mathcal{I}$ represents an ideal in $\mathbb{N}$. In this paper we intend to introduce the notion of $A^{\mathcal{I}^{\mathcal{K}}}$-summability which is a natural generalization of $A^{\mathcal{I}^*}$-summability. For more details regarding summability theory, one may refer to [8, 10, 14, 20].

Throughout the paper, we will use $(y_n)$ to denote the image $(A_n(x))$ of the sequence $x = (x_k)$ under the transformation of the non-negative regular infinite matrix $A$.

## 2. Definitions and preliminaries

**Definition 1.** *A collection $\mathcal{I}$ containing subsets of a nonempty set $X$ is called an ideal in $X$ if and only if* (i) $\emptyset \in \mathcal{I}$, (ii) $P, Q \in \mathcal{I}$ *implies* $P \cup Q \in \mathcal{I}$ *(Additive), and* (iii) $P \in \mathcal{I}, Q \subset P$ *implies* $Q \in \mathcal{I}$ *(Hereditary).*

If for any $x \in X$ $\{\{x\}\} \subset \mathcal{I}$ then it is said that $\mathcal{I}$ satisfies the admissibility property or simply is called admissible. Also $\mathcal{I}$ is called non-trivial if $X \notin \mathcal{I}$ and $\mathcal{I} \neq \{\emptyset\}$.

Some standard examples of ideal are given below:

(i) The set $\mathcal{I}_f$ consisting of all subsets of $\mathbb{N}$ having finite cardinality is an admissible ideal in $\mathbb{N}$.

(ii) The set $\mathcal{I}_d$ of all subsets of natural numbers having natural density 0 is an ideal in $\mathbb{N}$ which is also admissible.

(iii) The set $\mathcal{I}_c = \{A \subseteq \mathbb{N} : \sum_{a\in A} a^{-1} < \infty\}$ is an ideal in $\mathbb{N}$ which also has the so called admissibility property.

(iv) Suppose $\mathbb{N} = \bigcup\limits_{p=1}^{\infty} D_p$, where $D_p \subset \mathbb{N}$ for any $p \in \mathbb{N}$ and for $i \neq j$, $D_i \cap D_j = \emptyset$. Then, the set $\mathcal{I}$ of all subsets of $\mathbb{N}$ which intersects finitely many $D_p$'s forms an ideal in $\mathbb{N}$.

More important examples can be found in [9] and [11].

**Definition 2.** *A collection $\mathcal{F}$ containing subsets of a nonempty set $X$ is called a filter in $X$ if and only if (i) $\emptyset \notin \mathcal{F}$ (ii) $M, N \in \mathcal{F}$ implies $M \cap N \in \mathcal{F}$ and (iii) $M \in \mathcal{F}, N \supset M$ implies $N \in \mathcal{F}$.*

If $\mathcal{I}$ is a proper non-trivial ideal in $X$, then the collection $\mathcal{F}(\mathcal{I}) = \{M \subset X : \exists\, P \in \mathcal{I}\, such\, that\, M = X \setminus P\}$ forms a filter in $X$. It is known as the filter associated with the ideal $\mathcal{I}$.

**Definition 3** [12]. *Let $\mathcal{I}$ be an ideal in $\mathbb{N}$ which satisfies the admissibility property. A real-valued sequence $x = (x_k)$ is called $\mathcal{I}$-convergent to $l$ if for every $\varepsilon > 0$ the set $\{k \in \mathbb{N} : |x_k - l| \geq \varepsilon\}$ is contained in $\mathcal{I}$. The number $l$ is called the $\mathcal{I}$-limit of the sequence $x = (x_k)$. Symbolically, $\mathcal{I} - \lim x = l$.*

**Definition 4** [12]. *Let $\mathcal{I}$ be an ideal in $\mathbb{N}$ which satisfies the admissibility property. A sequence $x = (x_k)$ is called $\mathcal{I}^*$-convergent to $l$, if there exists a set $M = \{m_1 < m_2 < ... < m_k < ...\}$ in the associated filter $\mathcal{F}(\mathcal{I})$, for which $\lim_k x_{m_k} = l$ holds.*

**Definition 5** [13]. *Let $\mathcal{I}, \mathcal{K}$ denote two ideals in $\mathbb{N}$. A sequence $x = (x_k)$ is called $\mathcal{I}^{\mathcal{K}}$-convergent to $l$ if the associated filter $\mathcal{F}(\mathcal{I})$ contains a set $M$ such that the sequence $y = (y_k)$ defined by*

$$y_k = \begin{cases} x_k, & k \in M, \\ l, & k \notin M \end{cases}$$

*is $\mathcal{K}$-convergent to $l$.*

If we consider $\mathcal{K} = \mathcal{I}_f$ then $\mathcal{I}^{\mathcal{K}}$-convergence concept coincides with $\mathcal{I}^*$-convergence [12].

**Definition 6** [13]. *Let $\mathcal{K}$ be an ideal in $\mathbb{N}$. Then, $P \subset_{\mathcal{K}} Q$ denotes the property $P \setminus Q \in \mathcal{K}$. Also $P \subset_{\mathcal{K}} Q$ and $Q \subset_{\mathcal{K}} P$ together implies $P \sim_{\mathcal{K}} Q$. Thus $P \sim_{\mathcal{K}} Q$ if and only if $P \triangle Q \in \mathcal{K}$. A set $P$ is said to be $\mathcal{K}$-pseudointersection of a system $\{P_i : i \in \mathbb{N}\}$ if for every $i \in \mathbb{N}$ $P \subset_{\mathcal{K}} P_i$ holds.*

**Definition 7** [13]. *Let $\mathcal{I}$ and $\mathcal{K}$ be two ideals on $\mathbb{N}$. Then $\mathcal{I}$ is said to have the additive property with respect to $\mathcal{K}$ or the condition $AP(\mathcal{I}, \mathcal{K})$ holds if every sequence $(F_n)_{n\in\mathbb{N}}$ of sets from $\mathcal{F}(\mathcal{I})$ has $\mathcal{K}$-pseudointersection in $\mathcal{F}(\mathcal{I})$.*

**Definition 8** [6]. *A real-valued sequence $x = (x_k)$ is said to be $A^{\mathcal{I}}$-summable to a real number $L$, if the transformed sequence $(A_n(x))$ is $\mathcal{I}$-convergent to $L$. Symbolically, it is written as $A^{\mathcal{I}} - \lim x_k = L$.*

**Definition 9** [6]. *A real-valued sequence $x = (x_k)$ is said to be $A^{\mathcal{I}^*}$-summable to a real number $L$, if there exists a set $M = \{m_1 < m_2 < ... < m_i < ...\} \in \mathcal{F}(\mathcal{I})$ such that*

$$\lim_{i\to\infty} \sum_k a_{m_i k} x_k = \lim_{i\to\infty} y_{m_i} = L.$$

## 3. Main results

Throughout the section, for a sequence $x = (x_k)$ we will use $y = (y_n)$ to denote the transformed sequence $(A_n(x))$ where $A_n(x) = \sum_{k=1}^{\infty} a_{nk} x_k$.

**Definition 10.** *Let $A = (a_{nk})_{n,k=1}^{\infty}$ be a non-negative regular matrix and suppose $\mathcal{I}, \mathcal{K}$ be two admissible ideals in $\mathbb{N}$. A real-valued sequence $x = (x_k)$ is said to be $A^{\mathcal{I}^{\mathcal{K}}}$-summable to $L \in \mathbb{R}$, if there exists a set $M \in \mathcal{F}(\mathcal{I})$ such that the sequence $z = (z_k)$ defined by*

$$z_k = \begin{cases} y_k, & k \in M, \\ L, & k \notin M \end{cases}$$

*is $\mathcal{K}$-convergent to $L$, where the sequence $y = (y_n)$ is defined as*

$$y_n = A_n(x) = \sum_{k=1}^{\infty} a_{nk} x_k.$$

In this case we write, $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$.

*Example 1.* Consider the decomposition of $\mathbb{N}$ given by

$$\mathbb{N} = \bigcup_{i=1}^{\infty} D_i, \quad D_i = \{2^{i-1}(2s-1) : \ s = 1, 2, 3, ...\}.$$

Let $\mathcal{I}$ denotes the ideal consisting of all subsets of $\mathbb{N}$ which intersects finitely many of $D_i$'s. Consider the sequence $x = (x_k)$ defined by $x_k = 1/i$ if $k \in D_i$ and the infinite matrix $A = (a_{nk})_{n,k=1}^{\infty}$ as

$$a_{nk} = \begin{cases} 1, & k = n+2, \\ 0, & otherwise. \end{cases}$$

Then, the sequence is $A^{\mathcal{I}^{\mathcal{K}}}$-summable to 0 for $\mathcal{K} = \mathcal{I}$.

*Justification*: Clearly,

$$y_n = \sum_{k=1}^{\infty} a_{nk} x_k = \frac{1}{i}, \quad n+2 \in D_i.$$

Let $M = \mathbb{N} \setminus D_1$. Then, $M \in \mathcal{F}(\mathcal{I})$ and it is easy to verify that the sequence $z = (z_k)$ defined by

$$z_k = \begin{cases} y_k, & k \in M, \\ 0, & k \notin M \end{cases}$$

is $\mathcal{I}$-convergent to 0. Hence, $A^{\mathcal{I}^{\mathcal{I}}} - \lim x_k = 0$.

**Theorem 1.** *Let $A^{\mathcal{I}^*} - \lim x_k = L$ then $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$.*

P r o o f. Let $A^{\mathcal{I}^*} - \lim x_k = L$. Then, there exists a set

$$M = \{m_1 < m_2 < ... < m_k < ...\} \in \mathcal{F}(\mathcal{I})$$

such that $\lim_i y_{m_i} = L$. This implies that the sequence $z = (z_k)$ defined as

$$z_k = \begin{cases} y_k, & k \in M, \\ L, & k \notin M \end{cases}$$

is usual convergent to $L$. Now by Theorem 2.1 of [11], we can say that for any ideal $\mathcal{K}$, the sequence $z = (z_k)$ is $\mathcal{K}$-convergent to $L$. Hence, $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$. □

**Theorem 2.** *Let $A^{\mathcal{K}} - \lim x_k = L$ then $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$.*

P r o o f. Since $A^{\mathcal{K}} - \lim x = L$, so for every $\varepsilon > 0$,

$$\{k \in \mathbb{N} : |y_k - L| \geq \varepsilon\} \in \mathcal{K}. \tag{3.1}$$

Choose $M = \mathbb{N}$ from $\mathcal{F}(\mathcal{I})$. Consider the sequence $z = (z_k)$ defined by $z_k = y_k,\ k \in M$. Then, using (3.1), we get for every $\varepsilon > 0$,

$$\{k \in \mathbb{N} : |z_k - L| \geq \varepsilon\} \in \mathcal{K}$$

i.e. $z = (z_k)$ is $\mathcal{K}$-convergent to $L$. Hence $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$. □

*Remark 1.* Converse of the above theorem is not necessarily true.

*Example 2.* Consider the ideals

$$\mathcal{I}_c = \{B \subseteq \mathbb{N} : \sum_{b \in B} b^{-1} < \infty\}, \quad \mathcal{I}_d = \{B \subseteq \mathbb{N} : d(B) = 0\}$$

and the infinite matrix $A = (a_{nk})_{n,k=1}^{\infty}$ defined by

$$a_{nk} = \begin{cases} 1, & k = n, \\ 0, & \textit{otherwise}. \end{cases}$$

Let $x = (x_k)$ be the sequence defined as

$$x_k = \begin{cases} 1, & k \text{ is prime}, \\ 0, & k \text{ is not prime}. \end{cases}$$

Then, there exists set $M$ of all non prime numbers $\in \mathcal{F}(\mathcal{I}_{\mathrm{d}})$ such that the sequence $z = (z_k)$ defined as

$$z_k = \begin{cases} y_k, & k \in M, \\ 0, & k \notin M \end{cases}$$

is $\mathcal{I}_c$-converegnt to 0. Hence, $A^{\mathcal{I}_d{}^{\mathcal{I}_c}} - \lim x_k = 0$. But we claim that $A^{\mathcal{I}_c} - \lim x_k \neq 0$. Because if $A^{\mathcal{I}_c} - \lim x_k = 0$, then for any particular $\varepsilon$ with $0 < \varepsilon < 1$, we have the set

$$\{k \in \mathbb{N} : |y_k - 0| \geq \varepsilon\} = \text{set of all prime numbers} \in \mathcal{I}_c,$$

it is a contradiction.

The next theorem gives the condition under which $A^{\mathcal{I}^{\mathcal{K}}}$-summability implies $A^{\mathcal{K}}$-summability.

**Theorem 3.** *Let $\mathcal{I}$ and $\mathcal{K}$ be two admissible ideals in $\mathbb{N}$. If $\mathcal{I} \subseteq \mathcal{K}$ then $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$ implies $A^{\mathcal{K}} - \lim x_k = L$.*

P r o o f. Let $\mathcal{I} \subseteq \mathcal{K}$. Then, $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$ gives the assurance of the existence of a set $M \in \mathcal{F}(\mathcal{I})$ such that the sequence $z = (z_k)$ defined as

$$z_k = \begin{cases} y_k, & k \in M, \\ L, & k \notin M \end{cases}$$

is $\mathcal{K}$-convergent to $L$ and subsequently, we have

$$\forall \varepsilon > 0, \quad \{k \in M : |y_k - L| \geq \varepsilon\} \in \mathcal{K}. \tag{3.2}$$

Now as the inclusion

$$\{k \in \mathbb{N} : |y_k - L| \geq \varepsilon\} \subseteq \{k \in M : |y_k - L| \geq \varepsilon\} \cup (\mathbb{N} \setminus M)$$

holds and by our assumption, $\mathbb{N} \setminus M \in \mathcal{I} \subseteq \mathcal{K}$, from (3.2) we have

$$\{k \in \mathbb{N} : |y_k - L| \geq \varepsilon\} \in \mathcal{K}.$$

Hence, $A^{\mathcal{K}} - \lim x_k = L$. □

**Theorem 4.** *If every subsequence of $x = (x_k)$ is $A^{\mathcal{I}^{\mathcal{K}}}$-summable to $L$, then $x$ is $A^{\mathcal{I}^{\mathcal{K}}}$-summable to $L$.*

P r o o f. If possible let us assume the contrary. Then, for every $M \in \mathcal{F}(\mathcal{I})$, the sequence $z = (z_k)$ defined as

$$z_k = \begin{cases} y_k, & k \in M, \\ L, & k \notin M \end{cases}$$

is not $\mathcal{K}$-convergent to $L$. In other words, for any $M \in \mathcal{F}(\mathcal{I})$, there exists an $\varepsilon_M > 0$ such that

$$B = M \cap \{k \in \mathbb{N} : |y_k - L| \geq \varepsilon_M\} \notin \mathcal{K}.$$

Since $\mathcal{K}$ is admissible, so $B$ is infinite. Let $B = \{b_1 < b_2 < ... < b_k < ...\}$. Construct a subsequence $w = (w_k)$ defined as $w_k = y_{b_k}$ for $k \in \mathbb{N}$. Then, $A^{\mathcal{I}^{\mathcal{K}}} - \lim w_k \neq L$, we get a contradiction to the hypothesis. □

**Theorem 5.** *Let $x = (x_k)$ be a sequence such that $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$. Then, every subsequence of $x$ is $A^{\mathcal{I}^{\mathcal{K}}}$-summable to $L$ if and only if both $\mathcal{I}$ and $\mathcal{K}$ does not contain infinite sets.*

P r o o f. There are two possible cases.

*Case I.* Let $\mathcal{K}$ contain an infinite set. Suppose $C$ be an infinite set and $C \in \mathcal{K}$. Then, $\mathbb{N} \setminus C \in \mathcal{F}(\mathcal{K})$ and $\mathbb{N} \setminus C$ is also infinite. Let $\varepsilon > 0$ be arbitrary. Choose $L_1 \in \mathbb{R}$ such that $L_1 \neq L$. Consider the infinite matrix $A = (a_{nk})_{n,k=1}^{\infty}$, defined as

$$a_{nk} = \begin{cases} 1, & k = n, \\ 0, & \textit{otherwise}, \end{cases}$$

and the sequence $x = (x_k)$ such that

$$x_k = \begin{cases} L_1, & k \in C, \\ L, & k \in \mathbb{N} \setminus C. \end{cases}$$

Then,

$$\{k \in \mathbb{N} : |y_k - L| \geq \varepsilon\} \subseteq C \in \mathcal{K}.$$

This means that $x$ is $A^{\mathcal{K}}$-summable to $L$. Therefore by Theorem 2, $x$ is $A^{\mathcal{I}^{\mathcal{K}}}$-summable to $L$. But clearly the subsequence $(x_k)_{k \in C}$ of $x$ is $A^{\mathcal{I}^{\mathcal{K}}}$-summable to $L_1$ and not to $L$.

*Case II.* Let $\mathcal{K}$ does not contain an infinite set. Then $\mathcal{K} = \mathcal{I}_f$ and $A^{\mathcal{I}^{\mathcal{K}}}$-summability concept coincides with $A^{\mathcal{I}^*}$-summability.

*Subcase I: if $\mathcal{I}$ contains an infinite set.* Let $B$ be any infinite set such that $B \in \mathcal{I}$. Then, $\mathbb{N} \setminus B \in \mathcal{F}(\mathcal{I})$ and $\mathbb{N} \setminus B$ is also infinite. Define a sequence $x = (x_k)$ as

$$x_k = \begin{cases} \xi, & k \in B, \\ L, & k \in \mathbb{N} \setminus B, \end{cases}$$

where $\xi (\neq L) \in \mathbb{R}$. Clearly $x$ is $A^{\mathcal{I}^*}$-summable to $L$ for the infinite matrix considered in Case I. But clearly the subsequence $(x_k)_{k \in B}$ of $x$ is not $A^{\mathcal{I}^*}$-summable to $L$.

*Subcase II: if $\mathcal{I}$ does not contain an infinite set.* In this subcase, we have $\mathcal{I} = \mathcal{K} = \mathcal{I}_f$ and therefore $A^{\mathcal{I}^{\mathcal{K}}}$-summability concept coincides with ordinary summability ([10]) so any subsequence of $x$ is ordinary summable to $L$. □

*Remark 2.* If a sequence is $A^{\mathcal{I}^{\mathcal{K}}}$-summable then it may not be $A^{\mathcal{I}}$-summable.

*Example 3.* Let us consider the ideal $\mathcal{I}$ which is defined in Example 1 and the ideal

$$\mathcal{I}_c = \{A \subseteq \mathbb{N} : \sum_{a \in A} a^{-1} < \infty\}.$$

Let $M = \{k \in \mathbb{N} : k = 2^p$ for some non-negative integer p$\}$. Then, for the regular matrix $A = (a_{nk})_{n,k=1}^{\infty}$ defined as

$$a_{nk} = \begin{cases} 1, & k = n, \\ 0, & otherwise, \end{cases}$$

the sequence $x = (x_k)$ defined by

$$x_k = \begin{cases} 1, & k \in M, \\ 0, & k \notin M \end{cases}$$

is $A^{\mathcal{I}^{\mathcal{I}_c}}$-summable to 0 but $x$ is not $A^{\mathcal{I}}$-summable to 0.

**Theorem 6.** *Let $\mathcal{I}$ and $\mathcal{K}$ be two ideals in $\mathbb{N}$. Let $x = (x_k)$ be any real-valued sequence. Then, $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$ implies $A^{\mathcal{I}} - \lim x_k = L$ if and only if $\mathcal{K} \subseteq \mathcal{I}$.*

P r o o f. Let $\mathcal{K} \subseteq \mathcal{I}$ and suppose $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$. Then, the result follows directly from the following inclusion

$$\{k \in \mathbb{N} : |y_k - L| \geq \varepsilon\} \subseteq \{k \in M : |y_k - L| \geq \varepsilon\} \cup (\mathbb{N} \setminus M).$$

For the converse part, we assume the contrary. Then, there exists a set say $C \in \mathcal{K} \setminus \mathcal{I}$. Let $L_1$ and $L_2$ be two real numbers such that $L_1 \neq L_2$. Define a sequence $x = (x_k)$ as

$$x_k = \begin{cases} L_1, & k \in C, \\ L_2, & k \in \mathbb{N} \setminus C \end{cases}$$

and the regular matrix $A = (a_{nk})_{n,k=1}^{\infty}$ as

$$a_{nk} = \begin{cases} 1, & k = n, \\ 0, & otherwise. \end{cases}$$

Then, for any $\varepsilon > 0$ we have,

$$\{k \in \mathbb{N} : |y_k - L_2| \geq \varepsilon\} \subseteq C \in \mathcal{K}$$

which means that $x$ is $A^{\mathcal{K}}$-summable to $L_2$. Therefore by Theorem 2, $x$ is $A^{\mathcal{I}^{\mathcal{K}}}$-summable to $L_2$. By hypothesis $x$ is $A^{\mathcal{I}}$-summable to $L_2$. Therefore for $\varepsilon = |L_1 - L_2|$,

$$\{k \in \mathbb{N} : |y_k - L_2| \geq |L_1 - L_2|\} = C \in \mathcal{I},$$

it is a contradiction. Hence we must have $\mathcal{K} \subseteq \mathcal{I}$. $\square$

*Remark 3.* If a sequence is $A^{\mathcal{I}}$-summable then it may not be $A^{\mathcal{I}^{\mathcal{K}}}$-summable. Consider the ideal $\mathcal{I}$ and the sequence $x = (x_k)$ defined in Example 1. Then, proceeding as Example 1 of [6], we can prove that $A^{\mathcal{I}^{\mathcal{I}_f}} - \lim x_k \neq 0$ although $A^{\mathcal{I}} - \lim x_k = 0$.

**Theorem 7.** *Let $\mathcal{I}$ and $\mathcal{K}$ be two admissible ideals of $\mathbb{N}$ such that the condition $AP(\mathcal{I},\mathcal{K})$ holds. Then, for a sequence $x=(x_k)$, $A^{\mathcal{I}}$-summability implies $A^{\mathcal{I}^{\mathcal{K}}}$-summability to the same limit.*

P r o o f. Let $A^{\mathcal{I}} - \lim x_k = L$. Choose a sequence of rationales $(\varepsilon_i)_{i\in\mathbb{N}}$. Then, for every $i$,

$$M_i = \{k \in \mathbb{N} : |y_k - L| < \varepsilon_i\} \in \mathcal{F}(\mathcal{I}).$$

Thus by Definition 7, there exists a set $M \in \mathcal{F}(\mathcal{I})$ such that for any $i \in \mathbb{N}$, $M \setminus M_i \in \mathcal{K}$. Consider the sequence $z=(z_k)_{k\in\mathbb{N}}$ defined by

$$z_k = \begin{cases} y_k, & k \in M, \\ L, & k \notin M. \end{cases}$$

To complete the proof, it is sufficient to show that the sequence $z=(z_k)$ is $\mathcal{K}$-convergent to $L$. Now,

$$\begin{aligned}\{k \in \mathbb{N} : |z_k - L| < \varepsilon_i\} &= \{k \in M : |z_k - L| < \varepsilon_i\} \cup \{k \in \mathbb{N}\setminus M : |z_k - L| < \varepsilon_i\} \\ &= (\mathbb{N}\setminus M) \cup \{k \in M : |z_k - L| < \varepsilon_i\} \\ &= (\mathbb{N}\setminus M) \cup (M_i \cap M) \\ &= \mathbb{N} \setminus (M \setminus M_i).\end{aligned}$$

Now as $M \setminus M_i \in \mathcal{K}$, so $\mathbb{N}\setminus(M\setminus M_i) \in \mathcal{F}(\mathcal{K})$ and consequently we have

$$\{k \in \mathbb{N} : |z_k - L| < \varepsilon_i\} \in \mathcal{F}(\mathcal{K})$$

i.e. $\mathcal{K} - \lim z_k = L$. Hence, $A^{\mathcal{I}^{\mathcal{K}}} - \lim x_k = L$. This completes the proof. □

**Theorem 8.** *Let $\mathcal{I},\mathcal{I}_1,\mathcal{I}_2,\mathcal{K},\mathcal{K}_1,\mathcal{K}_2$ be admissible ideals in $\mathbb{N}$ satisfying $\mathcal{I}_1 \subseteq \mathcal{I}_2$ and $\mathcal{K}_1 \subseteq \mathcal{K}_2$. Then,*

(i) $A^{\mathcal{I}_1^{\mathcal{K}}} - \lim x_k = L$ *implies* $A^{\mathcal{I}_2^{\mathcal{K}}} - \lim x_k = L$;

(ii) $A^{\mathcal{I}^{\mathcal{K}_1}} - \lim x_k = L$ *implies* $A^{\mathcal{I}^{\mathcal{K}_2}} - \lim x_k = L$.

P r o o f. (i) Suppose $A^{\mathcal{I}_1^{\mathcal{K}}} - \lim x_k = L$. Then, by Definition 10, there exists $M \in \mathcal{F}(\mathcal{I}_1)$ such that the sequence $z=(z_k)$ defined as

$$z_k = \begin{cases} y_k, & k \in M, \\ L, & k \notin M \end{cases}$$

is $\mathcal{K}$-convergent to $L$. Now since $M \in \mathcal{F}(\mathcal{I}_1)$, we have $\mathbb{N}\setminus M \in \mathcal{I}_1$ and therefore by hypothesis $\mathbb{N}\setminus M \in \mathcal{I}_2$, which again implies $M \in \mathcal{F}(\mathcal{I}_2)$. Hence we must have that $A^{\mathcal{I}_2^{\mathcal{K}}} - \lim x_k = L$.

(ii) Suppose $A^{\mathcal{I}^{\mathcal{K}_1}} - \lim x_k = L$. Then, by Definition 10, there exists $M \in \mathcal{F}(\mathcal{I}_1)$ such that the sequence $z=(z_k)$ defined as,

$$z_k = \begin{cases} y_k, & k \in M, \\ l, & k \notin M \end{cases}$$

satisfies the following property $\forall \varepsilon > 0$,

$$\{k \in \mathbb{N} : |z_k - l| \geq \varepsilon\} \in \mathcal{K}_1.$$

Now by hypothesis the inclusion $\mathcal{K}_1 \subseteq \mathcal{K}_2$ holds, so we must have for $\forall \varepsilon > 0$,

$$\{k \in \mathbb{N} : |z_k - l| \geq \varepsilon\} \in \mathcal{K}_2.$$

Hence $A^{\mathcal{I}^{\mathcal{K}_2}} - \lim x_k = L$. □

## 4. Conclusion

Summability plays an important role in mathematics, particularly in mathematical analysis. In this paper, we introduce and investigate a few properties of $A^{\mathcal{I}^{\mathcal{K}}}$-summability. We generate a few examples and counterexamples in order to study some inclusion relationships with some known methods of summability. But the main focus was to link $A^{\mathcal{I}}$ and $A^{\mathcal{I}^*}$-summability with $A^{\mathcal{I}^{\mathcal{K}}}$-summability. We prove that the condition $AP(\mathcal{I},\mathcal{K})$ plays a crucial role in this regard. In the future, this idea can be utilized by the researchers to develop some other forms of summability.

## Acknowledgements

The authors thank the anonymous referees for their constructive suggestions to improve the quality of the paper. The first author is grateful to the University Grants Commission, India for their fellowships funding under the UGC-JRF scheme (F. No. 16-6(DEC. 2018)/2019(NET/CSIR)) during the preparation of this paper.

**REFERENCES**

1. Altinok M., Küçükaslan M. Ideal limit superior-inferior. *Gazi Univ. J. Sci.*, 2017. Vol. 30, No. 1. P. 401–411.
2. Banerjee A. K., Paul M. *Weak and Weak* $I^K$-convergence in Normed Spaces.* 2018. 10 p. arXiv:1811.06707v1 [math.GN]
3. Banerjee A. K., Paul M. *A Note on $I^K$ and $I^{K^*}$-convergence in Topological Spaces.* 2018. 10 p. arXiv:1807.11772v1 [math.GN]
4. Das P., Sleziak M., Toma V. $I^K$-Cauchy functions. *Topology Appl.*, 2014. Vol. 173, P. 9–27. DOI: 10.1016/j.topol.2014.05.008
5. Das P., Sengupta S., Supina J. $I^K$-convergence of sequence of functions. *Math. Slovaca.*, 2019. Vol. 69, No. 5. P. 1137–1148. DOI: 10.1515/ms-2017-0296
6. Edely O. H. H. On some properties of $A^I$-summability and $A^{I^*}$-summability. *Azerb. J. Math.*, 2021. Vol. 11, No. 1. P. 189–200.
7. Edely O. H. H., Mursaleen M. On statistical $A$-summability. *Math. Comput. Model*, 2009. Vol. 49, No. 3. P. 672–680. DOI: 10.1016/j.mcm.2008.05.053
8. Freedman A. R., Sember J. J. Densities and summability. *Pacific J. Math.*, 1981. Vol. 95, No. 2. P. 293–305.
9. Gogola J., Mačaj M., Visnyai T. On $\mathcal{I}_c^{(q)}$-convergence. *Ann. Math. Inform.*, 2011. Vol. 38, P. 27–36.
10. Jarrah A. M., Malkowsky E. Ordinary, absolute and strong summability and matrix transformations. *Filomat*, 2003. No. 17. P. 59–78. DOI: 10.2298/FIL0317059J
11. Kostyrko P., Mačaj M., Šalát T., Sleziak M. $\mathcal{I}$-convergence and extremal $\mathcal{I}$-limit points. *Math. Slovaca*, 2005. Vol. 55, No. 4. P. 443–464.
12. Kostyrko P., Šalát T., Wilczyński W. $I$-convergence. *Real Anal. Exch.*, 2000–2001. Vol. 26, No. 2. P. 669–686.
13. Mačaj M., Sleziak M. $\mathcal{I}^K$-convergence. *Real Anal. Exch.*, 2010–2011. Vol. 36, No. 1. P. 177–194.
14. Mursaleen M. On some new invariant matrix methods of summability. *Q. J. Math.*, 1983. Vol. 34, No. 1. P. 77–86. DOI: 10.1093/qmath/34.1.77
15. Mursaleen M., Alotaibi A. On $\mathcal{I}$-convergence in random 2-normed spaces. *Math. Slovaca*, 2011. Vol. 61, No. 6. P. 933–940. DOI: 10.2478/s12175-011-0059-5
16. Mursaleen M., Mohiuddine S. A. On ideal convergence in probabilistic normed spaces. *Math. Slovaca*, 2012. Vol. 62, No. 1. P. 49–62. DOI: 10.2478/s12175-011-0071-9
17. Nabiev A., Pehlivan S., Gürdal M. On $\mathcal{I}$-Cauchy sequences. *Taiwanese J. Math.*, 2007. Vol. 11, No. 2. P. 569–576.
18. Šalát T., Tripathy B. C., Ziman M. On some properties of $\mathcal{I}$-convergence. *Tatra Mt. Math. Publ.*, 2004. Vol. 28, No. 2. P. 274–286.

19. Savaş E. Generalized asymptotically $\mathcal{I}$-lacunary statistical equivalent of order $\alpha$ for sequences of sets. *Filomat*, 2017. Vol. 31, No. 6. P. 1507–1514. DOI: 10.2298/FIL1706507S
20. Savaş E. General inclusion relations for absolute summability. *Math. Inequal. Appl.*, 2005. Vol. 8, No. 3. P. 521–527.
21. Savaş E., Gürdal M. Ideal convergent function sequences in random 2-normed spaces. *Filomat*, 2016. Vol. 30, No. 3. P. 557–567. DOI: 10.2298/FIL1603557S
22. Tripathy B. C., Hazarika B. Paranorm $I$-convergent sequence spaces. *Math. Slovaca*, 2009. Vol. 59, No. 4. P. 485–494. DOI: 10.2478/s12175-009-0141-4

URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 23–33
DOI: 10.15826/umj.2022.1.003

# INDUCED $nK_2$ DECOMPOSITION OF INFINITE SQUARE GRIDS AND INFINITE HEXAGONAL GRIDS

**Dinesan Deepthy**

GITAM University, Bangalore, India
ddinesan@gitam.edu

**Joseph Varghese Kureethara**

Christ University, Bangalore, India
frjoseph@christuniversity.in

**Abstract:** The induced $nK_2$ decomposition of infinite square grids and hexagonal grids are described here. We use the multi-level distance edge labeling as an effective technique in the decomposition of square grids. If the edges are adjacent, then their color difference is at least 2 and if they are separated by exactly a single edge, then their colors must be distinct. Only non-negative integers are used for labeling. The proposed partitioning technique per the edge labels to get the induced $nK_2$ decomposition of the ladder graph is the square grid and the hexagonal grid.

**Keywords:** Distance labelling, Channel assignment, $L(h, k)$-colouring, Rectangular grid, Hexagonal grid.

## 1. Introduction

Decomposition of graphs has been an intriguing area of study in Graph Theory. Many of the decomposition problems could be addressed by even a beginner in graph theory. However, it needs a crafty and involved work to achieve certain types of decompositions of graphs. Given a graph with vertices and edges, the task in decomposition is to find subgraphs with a particular property. The disjoint union of these subgraphs is the given graph itself. For a perfect decomposition, there should not be any edges left over apart from the decomposed subgraphs. In an optimal decomposition if there are some edges left over, the collection of such edges is known as leave. In case, with the compromise of some overlapping edges, if we can find subgraphs whose union is the given graph, the set of repeated (overlapping) edges is known as padding.

The problem of decomposition of graphs dates back to some real-life problems. The famous Kirkman's schoolgirl problem and the 9-prisoner's problem are some of them. For many years, decomposition was identified as $G$-design. This has its origin from the design of experiments. If we have $n$ samples to be compared optimally, we can use decomposition as tool. Steiner triple system also needs decomposition techniques.

In this paper, we go for some techniques of a particular type of decomposition known as induced decomposition. We consider only simple connected graphs. We use the definitions and notations mostly from [2] unless otherwise defined here. Two adjacent vertices are referred to as neighbors of each other. The set of neighbors of a vertex $v$ is called the open neighborhood of $v$ and is denoted by $N_G(v)$ or $N(v)$. The set $N[v] = N(v) \cup \{v\}$ is called the closed neighborhood of $v$. Similarly, we can consider the set of neighbors of an edge as well.

An edge-induced subgraph is a subset of the edges of a graph $G$ together with any vertices that are their end vertices. As seen in [1], if $G$ is a connected graph, and $e_1 = (u_1, v_1)$ and $e_2 = (u_2, v_2)$

are two edges of $G$, then the *distance between edges or edge distance* of $e_1$ and $e_2$ is defined as

$$ed(e_1, e_2) = \min\{d(u_1, u_2), d(u_1, v_2), d(v_1, u_2), d(v_1, v_2)\}.$$

If $ed(e_1, e_2) = 0$, then these edges are called *neighbor edges* or adjacent edges. For the induced decomposition, we use the colouring technique known as the $L'(2,1)$–edge coloring of a graph $G$ which is defined as in [7] and used in [4]. For non-negative integers $i$ and $j$, an $L'(i,j)$–edge coloring of a graph $G$ is an assignment of non-negative integers to the edges $e_1$ and $e_2$ of $G$ such that $|c(e_1) - c(e_2)| \geq i$ if $ed(e_1, e_2) = 0$ and $|c(e_1) - c(e_2)| \geq j$ if $ed(e_1, e_2) = 1$. No condition is placed on colors assigned to the edges $e_1$ and $e_2$ if $ed(e_1, e_2) \geq 2$.

In this paper we study the case where $i = 2$ and $j = 1$. For an $L'(i,j)$–edge coloring $c$ of a graph $G$, the $c$-span of $G$ is the maximum value of $|c(e_1) - c(e_2)|$ over all pairs of edges $e_1$ and $e_2$ of $E(G)$. It is denoted by $\lambda'_{i,j}(c)$. That is,

$$\lambda'_{i,j}(c) = \max\left\{|c(e_1) - c(e_2)| : e_1, e_2 \in E(G)\right\}.$$

In particular, for $i = 2$ and $j = 1$, from [7], we have the c-span of $G$ with respect to the $L'(2,1)$–edge coloring as

$$\lambda'_{2,1}(c) = \max\{|c(e_1) - c(e_2)| : e_1, e_2 \in E(G)\}.$$

We use Stiebitz et al. [6] for the terminologies of $\chi'$–critical graph and $\chi'$–critical edge.

## 2. Rectangular grids

We use the rectangular grid graph (RGG) concept from [3, 5]. Given a RGG, we apply $L'(2,1)$–edge coloring technique and thereby obtain the $L'(2,1)$–edge coloring number of infinite rectangular grids. For convenience we denote RGG as $G_{m,n}$, an $m \times n$ rectangular grid graph. A particular case is the *Ladder graph* denoted as $P_2 \times P_m$. The $L'(2,1)$–edge coloring number of ladder graph is obtained as follows.

**Theorem 1** [4, Theorem 4].

$$\lambda'(P_2 \times P_m) = \begin{cases} 4 & \textit{if} \quad m = 2, \\ 6 & \textit{if} \quad m = 3, \\ 7 & \textit{if} \quad m \geq 4. \end{cases}$$

**Theorem 2** [3, Theorem 2]. *The $L'(2,1)$–edge coloring number of the rectangular grid $G_{3,4}$ is 8. That is, $\lambda'(G_{3,4}) = 8$.*

By the following theorem, we obtain the smallest positive integer or the smallest maximum color used among the different $L'(2,1)$–edge coloring of the infinite rectangular grids.

**Theorem 3** [3, Theorem 3]. *The $L'(2,1)$–edge coloring number of $G_{m,n}$ is at most $9$. That is, $\lambda'(G_{m,n}) \leq 9$, for any positive integers $m$ and $n$.*

See Fig. 1 for a sample coloring [3].

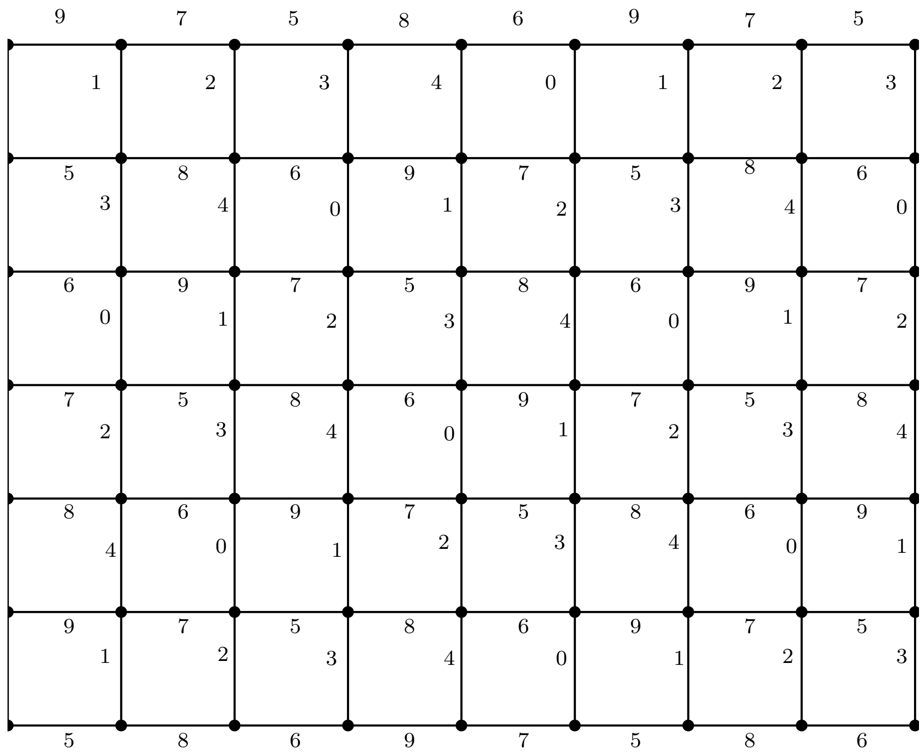



Figure 1. An optimal $L'(2,1)$–labeling of a fragment of rectangular grid.

### 2.1. The $nK_2$ decomposition of the Ladder graph $P_2 \times P_m$

In this section, we find the $nK_2$ decomposition of the ladder graph. By the symbol $\mathcal{P}_n(K_2)(G)$ we denote the graph $G$ that has the property $\mathcal{P}_n(K_2)$, i.e., induced $nK_2$. We use the edge coloring as a tool used in [4], see Fig. 2.

**Theorem 4.** *A ladder graph $P_2 \times P_m$ can have $\mathcal{P}_n(K_2)$ if and only if*

(i) *$n|q$ is such that $2 \le n \le \left\lfloor \frac{q}{6} \right\rfloor$, where $q$ is the size of the ladder graph, and*

(ii) $\operatorname{diam}(P_2 \times P_m) \ge 6$.

P r o o f. We see that the edge partition number of the ladder graph is

$$\pi_\nu{}'(P_2 \times P_m) = d(u) + d(w) = 6,$$

where $u$ is the vertex of maximum degree lying on the cycle $C_4$ and $w$ is such that $d(w)$ is maximum; $w \in N(u)$. Hence, the bound for $n$ in $\mathcal{P}_n(K_2)$ becomes

$$2 \le n \le \left\lfloor \frac{q}{6} \right\rfloor,$$

where $q$ is the size of the ladder graph and $q = 3m - 2$.

By the diameter condition, ladder graph has $\mathcal{P}_n(K_2)$ only if the diameter is at least six. Hence, consider the ladder graph $P_2 \times P_6$ which is of diameter 6 and size, $q = 16$. The bound for $n$ in $\mathcal{P}_n(K_2)$ is calculated as

$$2 \le n \le \left\lfloor \frac{16}{6} \right\rfloor = 2.$$

That is, there is every possibility that the graph $P_2 \times P_6$ can have $\mathcal{P}_2(K_2)$. We now verify the existence of $\mathcal{P}_2(K_2)$ in $P_2 \times P_6$ by the $L'(2,1)$–edge coloring technique given in [4].

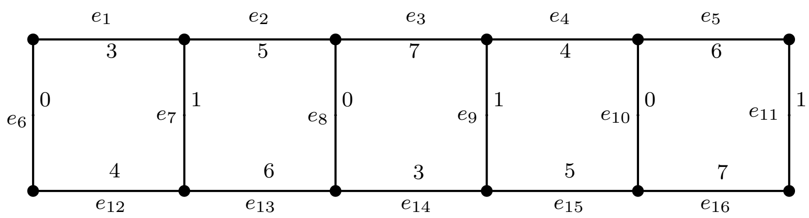



Figure 2. Optimal $L'(2,1)$–coloring of $P_2 \times P_6$.

Partition the edge set $E(P_2 \times P_6)$ into independent sets $E_1,\ E_2,\ \ldots$ such that the set $E_j$ consists of edges which receives color $j$. Note that the edges of $P_2 \times P_6$ are labeled $e_1,\ e_2,\ \ldots$ consecutively and selected under each $E_j$ such that the suffixes of the edge labels are in ascending order.

Consider the following partitioning of $E(P_2 \times P_6)$

$$\begin{array}{lll} E_0 = \{e_6,\ e_8,\ e_{10}\}, & E_1 = \{e_7,\ e_9,\ e_{11}\}, & E_2 = \varnothing, \\ E_3 = \{e_1,\ e_{14}\}, & E_4 = \{e_4,\ e_{12}\}, & E_5 = \{e_2,\ e_{15}\}, \\ E_6 = \{e_5,\ e_{13}\}, & E_7 = \{e_3,\ e_{16}\}. & \end{array}$$

As we aim at $2K_2$ decomposition, select two edges from $E_0$ and $E_1$ respectively such that the remaining two edges having different colors are at edge distance at least two

$$\begin{array}{lllll} {E_0}' = \{e_6,\ e_{10}\}, & {E_1}' = \{e_7,\ e_9\}, & {E_2}' = \varnothing, & {E_3}' = E_3, & {E_4}' = E_4, \\ {E_5}' = E_5, & {E_6}' = E_6, & {E_7}' = E_7, & {E_8}' = \{e_8,\ e_{11}\}. & \end{array}$$

As ${E_2}'$ is empty, we see that the distinct ${E_j}'$ for $j \neq 2$, forms the eight subsets with respect to $\mathcal{P}_2(K_2)(P_2 \times P_6)$ (see Fig. 3).

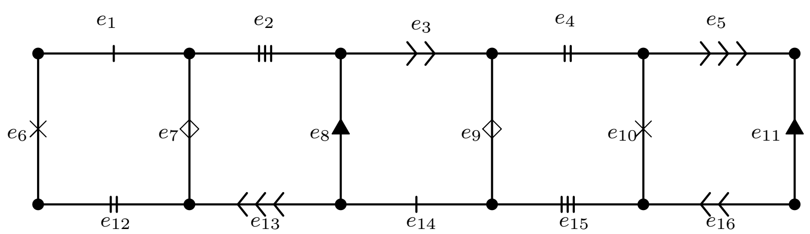



Figure 3. $\mathcal{P}_2(K_2)(P_2 \times P_6)$.

Note that the edges designed in the similar manner come under the same subset of decomposition.

As the size of $P_2 \times P_7$ is 19, it cannot have $\mathcal{P}_2(K_2)$ for any $n$. So consider the ladder graph $P_2 \times P_8$, whose size is 22. As seen earlier, the bound for $n$ in $\mathcal{P}_n(K_2)$ is calculated as,

$$2 \leq n \leq \left\lfloor \frac{22}{6} \right\rfloor,$$

which implies that $n$ takes up values 2 and 3. However, we see that the ladder graph $P_2 \times P_8$ is bound to have only $\mathcal{P}_2(K_2)$. We now verify the existence of $\mathcal{P}_2(K_2)$ in $P_2 \times P_8$ by the $L'(2,1)$–coloring technique and partitioning of the edge set of $P_2 \times P_8$ as follows

$$\begin{array}{ll} E_0 = \{e_8,\ e_{10},\ e_{12},\ e_{14}\}, & E_1 = \{e_9,\ e_{11},\ e_{13},\ e_{15}\}, \\ E_2 = \varnothing, & E_3 = \{e_1,\ e_6,\ e_{18}\}, \\ E_4 = \{e_4,\ e_{16},\ e_{21}\}, & E_5 = \{e_2,\ e_7,\ e_{19}\}, \\ E_6 = \{e_5,\ e_{17},\ e_{22}\}, & E_7 = \{e_3,\ e_{20}\}. \end{array}$$

As we aim at $2K_2$ decomposition, rearrange it to form a new partition as done earlier

$$\begin{array}{lll} E_0{}' = \{e_8,\ e_{10}\}, & E_1{}' = \{e_9,\ e_{11}\}, & E_2{}' = E_2 = \phi, \\ E_3{}' = \{e_1,\ e_6\}, & E_4{}' = \{e_4,\ e_{16}\}, & E_5{}' = \{e_2,\ e_7\}, \\ E_6{}' = \{e_5,\ e_{17}\}, & E_7{}' = \{e_3,\ e_{20}\}, & E_8{}' = \{e_{12}, e_{14}\}, \\ E_9{}' = \{e_{13},\ e_{15}\}, & E_{10}{}' = \{e_{18},\ e_{21}\}, & E_{11}{}' = \{e_{19},\ e_{22}\}. \end{array}$$

They form the eleven subsets with respect to $\mathcal{P}_2(K_2)(P_2 \times P_8)$, see Fig. 4 and Fig. 5. The optimal coloring, labeling and induced $2K_2$ decomposition of the ladder graph $P_2 \times P_8$ are given below.

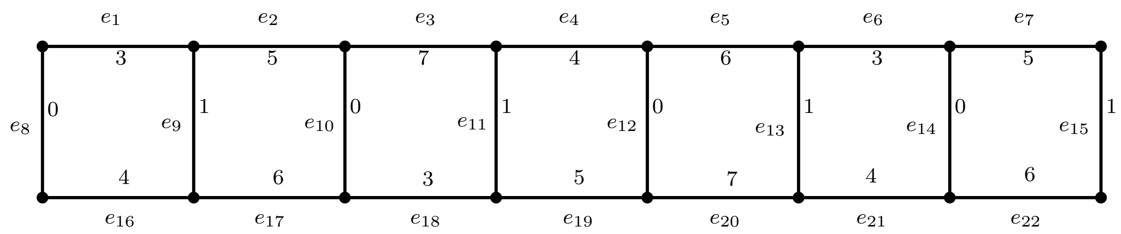



Figure 4. Optimal $L'(2,1)$–coloring in $(P_2 \times P_8)$.

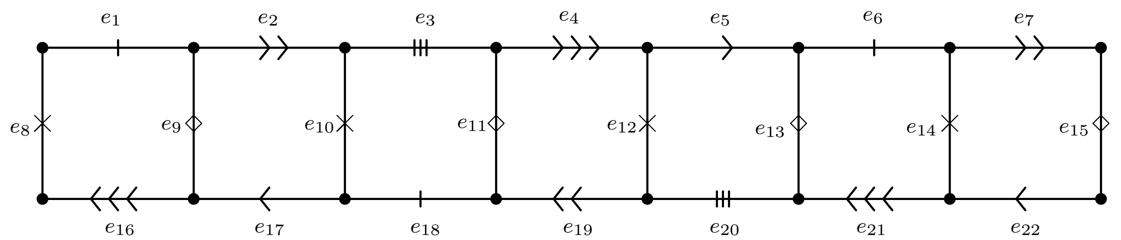



Figure 5. $\mathcal{P}_2(K_2)(P_2 \times P_8)$.

In this manner, we can have $\mathcal{P}_n(K_2)$ for any ladder graph of diameter at least six. Hence, we conclude that a ladder graph $P_2 \times P_m$ can have $\mathcal{P}_n(K_2)$ if and only if

(i) $n|q$ such that $2 \le n \le \left\lfloor \dfrac{q}{6} \right\rfloor$; where $q$ is the size of the ladder graph and

(ii) $\operatorname{diam}(P_2 \times P_m) \ge 6$.

$\square$

### 2.2. Rectangular grid and $\mathcal{P}_n(K_2)$

We now give a process of the induced $nK_2$ decomposition of rectangular grid.

We first find $\mathcal{P}_n(K_2)(P_3 \times P_m)$ or the induced $nK_2$ decomposition in the rectangular grid graph, $(P_3 \times P_m)$ using the optimal edge coloring discussed earlier as a tool. We see that the edge partition number of the rectangular grid graph is $\pi_\nu{}'(P_n \times P_m) = d(u) + d(w) = 8$, where $u$ is the vertex of maximum degree which lying on the cycle $C_4$ and $w$ is such that $d(w)$ is maximum; $w \in N(u)$. Hence, the bound for $n$ in $\mathcal{P}_n(K_2)$ becomes

$$2 \leq n \leq \left\lfloor \frac{q}{8} \right\rfloor,$$

where $q$ is the size of the rectangular grid graph. While considering the grid graph $(P_3 \times P_6)$, whose size is 27, we see that the bound for $n$ in $\mathcal{P}_n(K_2)$ is

$$2 \leq n \leq \left\lfloor \frac{27}{8} \right\rfloor = 3$$

and as $n$ must divide $q$, we have that $n = 3$. That is, $(P_3 \times P_6)$ can have $\mathcal{P}_3(K_2)$ and the verification of its existence is done by the optimal $L'(2,1)$–edge coloring technique as follows.

Partition the edge set $E(P_3 \times P_6)$ into independent sets $E_1,\ E_2,\ \ldots$ such that the set $E_j$ consists of edges with color $j$. Note that the edges of $P_3 \times P_6$ are labeled $e_1,\ e_2,\ \ldots$ consecutively and selected under each $E_j$ such that the suffixes of the edge labels are in ascending order, as done in ladder graph. Consider the following partitioning of $E(P_3 \times P_6)$ (see Fig. 7)

$$\begin{array}{lll} E_0 = \{e_6,\ e_{11},\ e_{20}\}, & E_1 = \{e_7,\ e_{21}\}, & E_2 = \{e_8,\ e_{17},\ e_{22}\}, \\ E_3 = \{e_9,\ e_{18}\}, & E_4 = \{e_{10},\ e_{19}\}, & E_5 = \{e_3,\ e_{12},\ e_{26}\}, \\ E_6 = \{e_5,\ e_{14},\ e_{23}\}, & E_7 = \{e_2,\ e_{16},\ e_{25}\}, & E_8 = \{e_4,\ e_{13},\ e_{27}\}, \\ E_9 = \{e_1,\ e_{15},\ e_{24}\}. & & \end{array}$$

As we aim at induced $3K_2$ decomposition, we will diffuse one of the subsets to have exactly three edges under each set. We also interchange the edge $e_{22}$ in $E_2$ for the same. Hence, the new partition can be considered as follows. Consider the following partitioning of $E(P_3 \times P_6)$ (see Fig. 7)

$$\begin{array}{lll} E_1{}' = \{e_7,\ e_{22},\ e_{20}\} & E_2{}' = \{e_8,\ e_{17},\ e_{21}\} & E_3{}' = \{e_9,\ e_{11},\ e_{18}\} \\ E_4{}' = \{e_6,\ e_{10},\ e_{19}\} & E_5{}' = E_5 = \{e_3,\ e_{12},\ e_{26}\} & E_6{}' = E_6 = \{e_5,\ e_{14},\ e_{23}\} \\ E_7{}' = E_7 = \{e_2,\ e_{16},\ e_{25}\} & E_8{}' = E_8 = \{e_4,\ e_{13},\ e_{27}\} & E_9{}' = E_9 = \{e_1,\ e_{15},\ e_{24}\}. \end{array}$$

This is the required $\mathcal{P}_3(K_2)$ of the grid graph $P_3 \times P_6$ (see Fig. 6 and Fig. 7).

We partition the edge set $E(P_3 \times P_{11})$ according to their edge labels as follows

$$\begin{array}{ll} E_0 = \{e_{11},\ e_{16},\ e_{21},\ e_{35},\ e_{40}\}, & E_1 = \{e_{12},\ e_{17},\ e_{36},\ e_{41}\}, \\ E_2 = \{e_{13},\ e_{18},\ e_{32},\ e_{37},\ e_{42}\}, & E_3 = \{e_{14},\ e_{19},\ e_{33},\ e_{38}\}, \\ E_4 = \{e_{15},\ e_{20},\ e_{34},\ e_{39}\}, & E_5 = \{e_3,\ e_8,\ e_{22},\ e_{27},\ e_{46},\ e_{51}\}, \\ E_6 = \{e_5,\ e_{10},\ e_{24},\ e_{29},\ e_{43},\ e_{48}\}, & E_7 = \{e_2,\ e_7,\ e_{26},\ e_{31},\ e_{45},\ e_{50}\}, \\ E_8 = \{e_4,\ e_9,\ e_{23},\ e_{28},\ e_{47},\ e_{52}\}, & E_9 = \{e_1,\ e_6,\ e_{25},\ e_{30},\ e_{44},\ e_{49}\}. \end{array}$$

In a similar manner, $P_4 \times P_6$ will have only $\mathcal{P}_2(K_2)$ as its size is 38 and due to the divisibility criteria. We can find $\mathcal{P}_n(K_2)$ for any grid $P_m \times P_6$ using the similar conditions as that of a ladder graph mentioned in the earlier section. Now consider the rectangular grid graph, $P_3 \times P_{11}$, whose size is 52 and the bound for $n$ in $\mathcal{P}_2(K_2)$ is obtained to be

$$2 \leq n \leq \left\lfloor \frac{q}{8} \right\rfloor.$$

By applying the divisibility conditions, we see that $n$ takes up the values 2 and 4. That is, the graph $P_3 \times P_{11}$ can have only $\mathcal{P}_2(K_2)$ and $\mathcal{P}_4(K_2)$ (see Fig. 8).

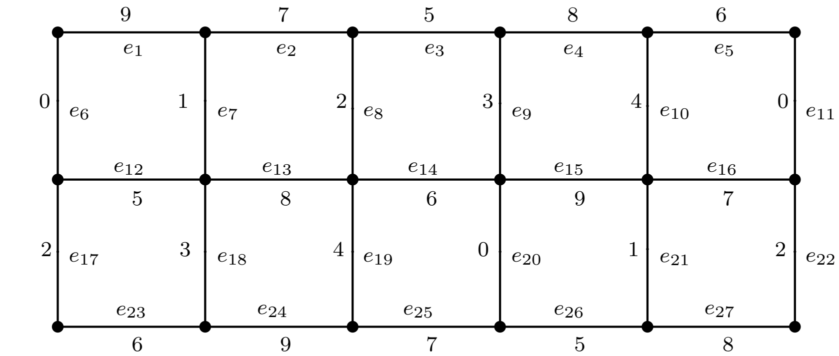



Figure 6. Optimal $L'(2,1)$–edge coloring of $P_3 \times P_6$.

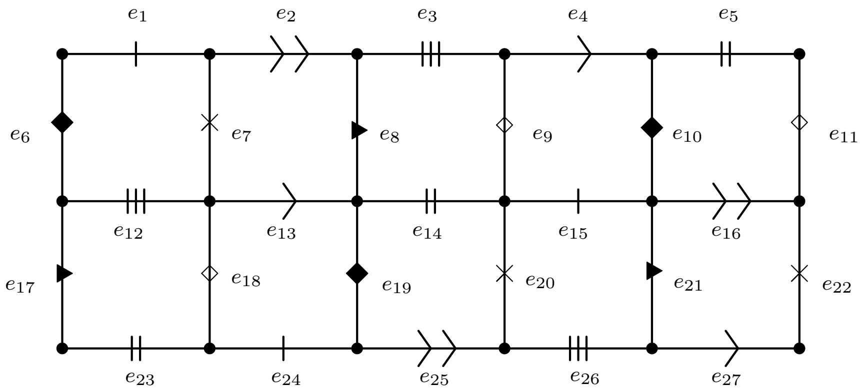



Figure 7. $\mathcal{P}_3(K_2)(P_3 \times P_6)$.

As we aim at $2K_2$ decomposition, further partitioning is required to have induced $2K_2$ in each subset. As the sets $E_0$ and $E_2$ have five edges, we eliminate one from each to form a new subset with two edges at distance at least two. Also eliminate two edges from each subset, with four and six elements, to form new subsets with only two edges in each. The resulting 26 subsets containing two elements each are the required $\mathcal{P}_2(K_2)$ in $E(P_3 \times P_{11})$ (see Fig. 9)

$$
\begin{array}{lll}
E_0{}' = \{e_{11},\ e_{16}\}, & E_1{}' = \{e_{12},\ e_{17}\}, & E_2{}' = \{e_{13},\ e_{18}\}, \\
E_3{}' = \{e_{14},\ e_{19}\}, & E_4{}' = \{e_{15},\ e_{20}\}, & E_5{}' = \{e_3,\ e_8\}, \\
E_6{}' = \{e_5,\ e_{10}\}, & E_7{}' = \{e_2,\ e_7\}, & E_8{}' = \{e_4,\ e_9\}, \\
E_9{}' = \{e_1,\ e_6\}, & E_{10}{}' = \{e_{40},\ e_{42}\}, & E_{11}{}' = \{e_{21},\ e_{35}\}, \\
E_{12}{}' = \{e_{36},\ e_{41}\}, & E_{13}{}' = \{e_{32},\ e_{37}\}, & E_{14}{}' = \{e_{33},\ e_{38}\}, \\
E_{15}{}' = \{e_{34},\ e_{39}\}, & E_{16}{}' = \{e_{22},\ e_{27}\}, & E_{17}{}' = \{e_{46},\ e_{51}\}, \\
E_{18}{}' = \{e_{24},\ e_{29}\}, & E_{19}{}' = \{e_{43},\ e_{48}\}, & E_{20}{}' = \{e_{26},\ e_{31}\}, \\
E_{21}{}' = \{e_{45},\ e_{50}\}, & E_{22}{}' = \{e_{23},\ e_{28}\}, & E_{23}{}' = \{e_{47},\ e_{52}\}, \\
E_{24}{}' = \{e_{25},\ e_{30}\}, & E_{25}{}' = \{e_{44},\ e_{49}\}. &
\end{array}
$$

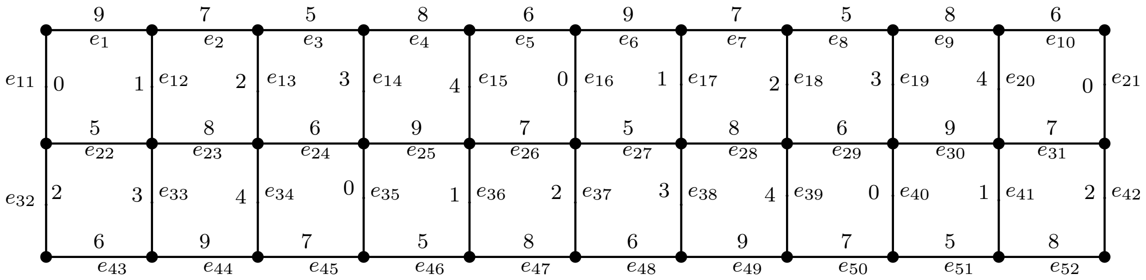


Figure 8. Optimal $L'(2,1)$–labeling of $P_3 \times P_{11}$.

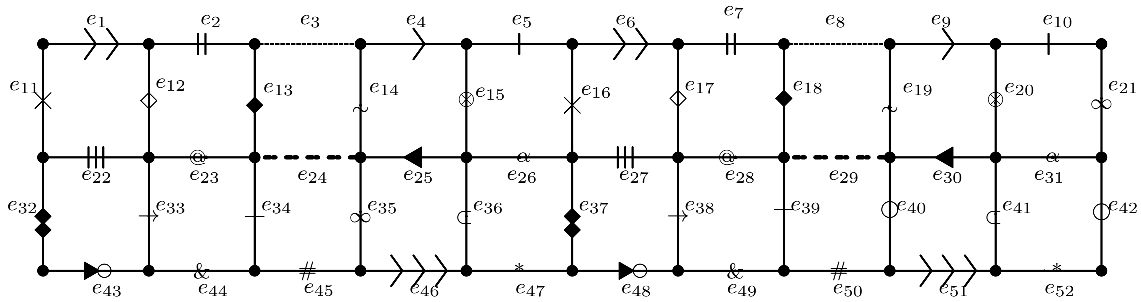


Figure 9. $\mathcal{P}_2(K_2)(P_3 \times P_{11})$.

In a similar manner we can have $\mathcal{P}_4(K_2)(P_3 \times P_{11})$ with the following partition (see Fig. 10)

$$
\begin{array}{ll}
E_0'' = \{e_{11},\ e_{16},\ e_{21},\ e_{35}\}, & E_1'' = \{e_{12},\ e_{17},\ e_{36},\ e_{41}\},\\
E_2'' = \{e_{13},\ e_{18},\ e_{32},\ e_{37}\}, & E_3'' = \{e_{14},\ e_{19},\ e_{33},\ e_{38}\},\\
E_4'' = \{e_{15},\ e_{20},\ e_{34},\ e_{39}\}, & E_5'' = \{e_{22},\ e_{27},\ e_{46},\ e_{51}\},\\
E_6'' = \{e_{5},\ e_{10},\ e_{24},\ e_{29}\}, & E_7'' = \{e_{26},\ e_{31},\ e_{45},\ e_{50}\},\\
E_8'' = \{e_{4},\ e_{9},\ e_{23},\ e_{28}\}, & E_9'' = \{e_{1},\ e_{6},\ e_{25},\ e_{30}\},\\
E_{10}'' = \{e_{40},\ e_{42},\ e_{43},\ e_{48}\}, & E_{11}'' = \{e_{3},\ e_{8},\ e_{44},\ e_{49}\},\\
E_{12}'' = \{e_{2},\ e_{7},\ e_{47},\ e_{52}\}. &
\end{array}
$$

Similarly, we can form $\mathcal{P}_n(K_2)$ for any rectangular grid $P_m \times P_r$, where $r = 5x + 1$ for $x \geq 1$ by following the conditions mentioned under the ladder graph and the above pattern. The same partitioning technique can be applied to study the existence of $\mathcal{P}_n(K_2)$ in hexagonal grid as well.

## 3. Hexagonal Grids and $\mathcal{P}_n(K_2)$

The hexagonal grid or honeycomb topology has wide range of application in Network-on-chip (NoC) which is an effective architecture in chip designing. It is evident that the hexagonal grid or

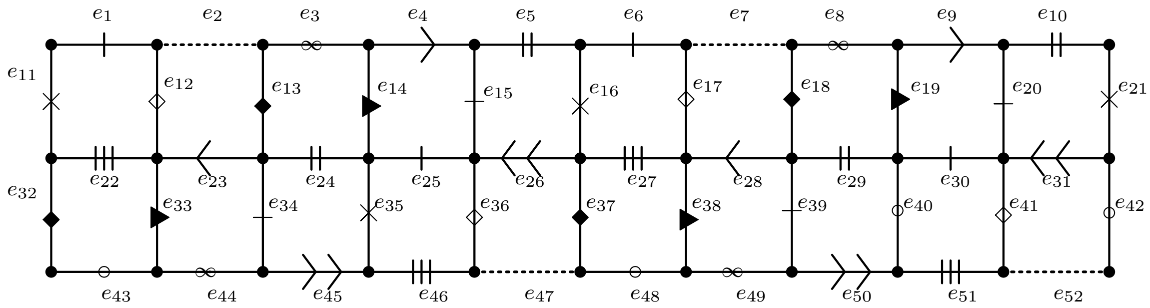
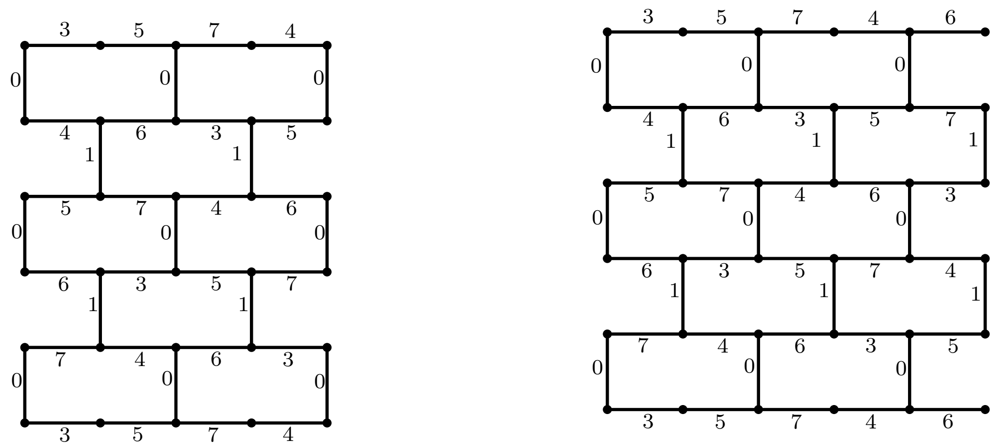



Figure 10. $\mathcal{P}_4(K_2)(P_3 \times P_{11})$.




(a) $H_{6,5}$.

(b) $H_{6,6}$.

Figure 11. Hexagonal grids.

honeycomb structure denoted by $H_{m,n}$ is a spanning subgraph of a rectangular grid with $V(H_{m,n}) = V(G_{m,n})$ and $E(H_{m,n}) \subset E(G_{m,n})$ implying that the value of $\Delta$ decreases and hence $\lambda'(H_{m,n}) < 9$. However, it is proved in [3] that $\lambda'(H_{m,n}) = 7$.

Consider the hexagonal grid $H_{6,5}$ whose size is 37 (Fig. 11 (a)). Then by the condition of $\mathcal{P}_n(K_2)$ we have that there exists no $\mathcal{P}_n(K_2)$ as the size of this grid is prime. Consider the hexagonal grid $H_{6,6}$ whose size is 45. Clearly, nine copies of $\mathcal{P}_5(K_2)$ and five copies of $\mathcal{P}_9(K_2)$ exist in $H_{6,6}$ (Fig. 11 (b)). As $\lambda'(H_{6,6})$ is 7, the edges of $H_{6,6}$, can be partitioned into independent sets $E_0, E_1, E_3, E_4, E_5, E_6, E_7$ such that the set $E_j$ consists of edges which receive color $j$.
Here, $|E_0| = 9,\ |E_1| = 6,\ |E_3| = 6,\ |E_4| = 6,\ |E_5| = 6,\ |E_6| = 6,\ |E_7| = 6$.

For $\mathcal{P}_5(K_2)$, as we aim at induced $5K_2$ in each subset, further partitioning is required as the cardinality of each subset is greater than 5. Let the new partitioning be $|E_0|' = 5, |E_1|' = 5$, $|E_3|' = 5,\ |E_4|' = 5,\ |E_5|' = 5,\ |E_6|' = 5, |E_7|' = 5$. Remaining four edges from $|E_0|$ and one edge each from other subsets of the first partition can be put under two newly formed subsets of cardinality 5. This results in nine copies of $\mathcal{P}_5(K_2)$ as required. For $\mathcal{P}_9(K_2)$, the twelve edges of $|E_6| = 6, |E_7|$ can be distributed equally among the other independent sets of cardinality 6, which

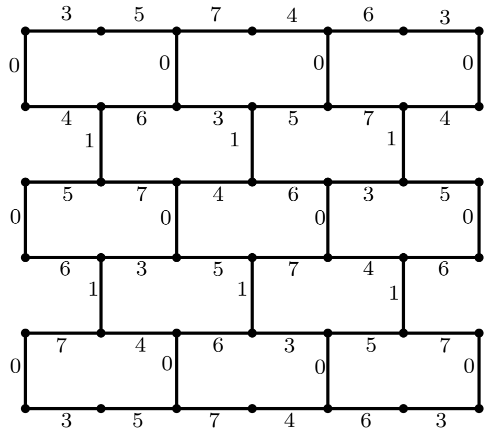



Figure 12. Hexagonal grid $H_{6,7}$.

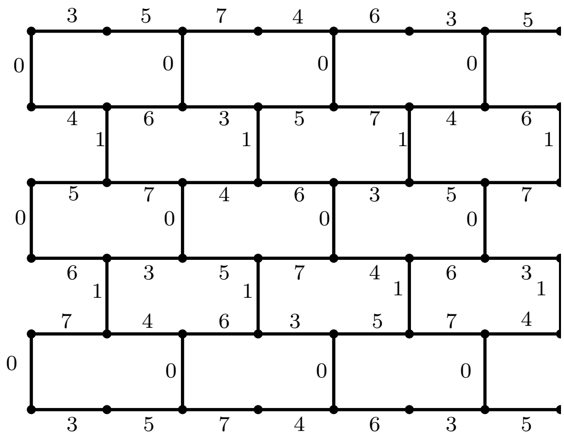



Figure 13. Hexagonal grid $H_{6,8}$.

gives $\mathcal{P}_9(K_2)$ of $H_{6,6}$.

Similarly, $H_{6,7}$ (Fig. 12) is of size 54 and by the condition of $\mathcal{P}_n(K_2)$ we have that there exist twenty seven copies of $\mathcal{P}_2(K_2)$, eighteen copies of $\mathcal{P}_3(K_2)$, nine copies of $\mathcal{P}_6(K_2)$, six copies of $\mathcal{P}_9(K_2)$ and three copies of $\mathcal{P}_{18}(K_2)$. However, as the size of $H_{6,8}$ (Fig. 13) is 62, it has thirty one copies of $\mathcal{P}_2(K_2)$ only.

## Acknowledgements

We thank the Section Editor and the reviewers for their valuable comments and suggestions.

## REFERENCES


1. Balci M. A., Dündar P. Average edge-distance in graphs. *Selçuk J. Appl. Math.,* 2010. Vol. 11, No. 2. P. 63–70.
2. Chartrand G., Lesniak L., Zhang P. *Graphs and Digraphs.* 6th Ed. Boca Raton, US: CRC Press, 2016. 640 p. DOI: 10.1201/b19731
3. Deepthy D., Kureethara J. V. On $L'(2,1)$–edge coloring number of regular grids. *An. Şt. Univ. Ovidius Constanţa,* 2019. Vol. 27, No. 3. P. 65–81. DOI: 10.2478/auom-2019-0034
4. Deepthy D., Kureethara J. V. $L'(2,1)$–edge coloring of trees and cartesian product of path graphs. *Int. J. Pure Appl. Math.,* 2017. Vol. 117, No. 13. P. 135–143.
5. Guo D., Zhan H., Wong M.D.F. On coloring rectangular and diagonal grid graphs for multiple patterning lithography. In: *23rd Asia and South Pacific Design Automation Conference.* IEEE Xplore, 2018. P. 387–392. DOI: 10.1109/ASPDAC.2018.8297354
6. Stiebitz M., Scheide D., Toft B., Favrholdt L. M. *Graph Edge Coloring: Vizing's Theorem and Goldberg's Conjecture.* Hoboken, NJ: John Wiley & Sons Inc., 2012. 344 p.
7. Yeh R. K. *Labelling Graphs with a Condition at Distance Two.* Ph. D. Thesis. Columbia, SC: University of South Carolina, 1990.

URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 34–42

DOI: 10.15826/umj.2022.1.004

# HJB–INEQUALITIES IN ESTIMATING REACHABLE SETS OF A CONTROL SYSTEM UNDER UNCERTAINTY[1]

**Tatiana F. Filippova**

Krasovskii Institute of Mathematics and Mechanics,
Ural Branch of the Russian Academy of Sciences,
16 S. Kovalevskaya Str., Ekaterinburg, 620108, Russian Federation

ftf@imm.uran.ru

**Abstract:** Using the technique of generalized inequalities of the Hamilton–Jacobi–Bellman type, we study here the state estimation problem for a control system which operates under conditions of uncertainty and nonlinearity of a special kind, when the dynamic equations describing the studied system simultaneously contain the different forms of nonlinearity in state velocities. Namely, quadratic functions and uncertain matrices of state velocity coefficients are presented therein. The external ellipsoidal bounds for reachable sets are found, some approaches which may produce internal estimates for such sets are also mentioned. The example is included to illustrate the result.

**Keywords:** Control, Nonlinearity, Uncertainty, Ellipsoidal calculus, State estimation.

## 1. Introduction

In the paper the nonlinear dynamical control systems with unknown but bounded uncertainties with its set-membership description [20–22, 24] are studied and the main goal of the present research is to construct the outer (external) estimates for related reachable sets. Several approaches may be used for these purposes but most of them are suitable only for the case of linear dynamical systems [8, 24, 26]. However researches in nonlinear control systems theory are very important for various applications, e.g., [3–7]. The key issue in nonlinear set-membership estimation theory is to find suitable techniques, which allow to find estimates (of external and internal kind) for unknown system states and do not involve difficult and lengthy computations. Some approaches to achieve this goal may be taken and further developed using techniques of differential inclusions theory [5] but in general these ideas produce very complicated numerical schemes and hard working algorithms.

We use here the advantages of ellipsoidal calculus [8, 24, 26] and further develop the Hamilton-Jacobi-Bellmann (HJB) techniques initiated in researches [12, 17, 19] to construct computationally acceptable set-valued estimates of reachable sets for a new class of nonlinear control systems under uncertainty [9–11, 13–16, 25].

The paper is organized as follows. First, in Section 2 we introduce some notations, give necessary definitions and formulate the main problem. Ellipsoidal external estimates are developed further in Section 3 where equations describing parameters of estimating ellipsoids are presented. The example is given in Section 4 to illustrate the theoretical results.

The study continues previous researches in this field and deals now with a special case, when a nonlinearity of quadratic type together with bilinear terms defined by uncertain matrix are

[1]The work was performed as part of research conducted in the Ural Mathematical Center with the financial support of the Ministry of Science and Higher Education of the Russian Federation (Agreement number 075-02-2022-874).

presented in the dynamical equations having also an uncertainty in initial states. This case is both of theoretical and of applied importance. Note, however, that the structure of the system under consideration differs from that previously studied, including that given in earlier papers [16, 17].

The problems studied in this paper are generated both by the problems of the theory of guaranteed state estimation for nonlinear dynamical systems with uncertain dynamics and by practical problems of control-estimation under unpredictable interferences.

## 2. Main notations and problem statement

We introduce first a short list of main notations. Let $\mathbb{R}^n$ be the $n$–dimensional Euclidean space with the inner product $x'y = \sum_{i=1}^{n} x_i y_i$ for $x, y \in \mathbb{R}^n$, here a prime indicates a transpose, $\|x\| = (x'x)^{1/2}$. Let $\operatorname{comp}\mathbb{R}^n$ be the set of all compact subsets of $\mathbb{R}^n$, $h(A,B)$ be the Hausdorff distance between $A, B \in \operatorname{comp}\mathbb{R}^n$. We denote also $B(a,r) = \{x \in \mathbb{R}^n : \|x-a\| \le r\}$, a symbol $I$ will stand for the identity $n \times n$-matrix.

We use the symbol $\mathbb{R}^{n\times n}$ for a set of all $n \times n$–matrices and $E(a,Q)$ for an ellipsoid in $\mathbb{R}^n$, $E(a,Q) = \{x \in \mathbb{R}^n : \ (Q^{-1}(x-a), (x-a)) \le 1\}$ with a center $a \in \mathbb{R}^n$ and with a symmetric positive definite $n \times n$–matrix $Q$. For any $n \times n$–matrix $M = \{m_{ij}\}$ we denote $\operatorname{Tr}(M) = \sum_{i=1}^{i=n} m_{ii}$.

We will study the control nonlinear system

$$\dot{x} = A(t)x + f(x)d + u(t), \quad t \in [t_0, T], \quad x_0 \in X_0. \tag{2.1}$$

We assume further that $\|x\| \le K$ $(K > 0)$, $x, d \in \mathbb{R}^n$, the nonlinear function $f(x)$ is quadratic in $x$, $f(x) = x'Bx$, and $B$ is a symmetric and positive definite $n \times n$-matrix. Here the coordinates $d_i$ of the vector $d$ are the coefficients with which the nonlinear function $f(x)$ enters the right side of the differential control system (2.1), in particular, they can be interpreted as independent parameters of the studied model or as coefficients of approximate estimates of the state velocities of the simulated system.

The $n \times n$-matrix function $A(t)$ in (2.1) is assumed to be of the form

$$A(t) = A^1(t) + A^0, \tag{2.2}$$

where the matrix $A^0$ (with its dimension $n \times n$) is given and the measurable $n \times n$-matrix function $A^1(t)$ is unknown but bounded, $A^1(t) \in \mathcal{A}^1$ $(t \in [t_0, T])$. Namely, we have

$$A(t) \in \mathcal{A} = A^0 + \mathcal{A}^1, \tag{2.3}$$
$$\mathcal{A}^1 = \big\{A = \{a_{ij}\} \in \mathbb{R}^{n\times n} : |a_{ij}| \le c_{ij}, \ i, j = 1, \dots n\big\},$$

where the numbers $c_{ij} \ge 0$ $(i, j = 1, \dots n)$ are given.

We will assume that $X_0$ in (2.1) is an ellipsoid,

$$X_0 = E(a_0, Q_0),$$

with a symmetric and positive definite matrix $Q_0 \in \mathbb{R}^{n\times n}$ and with a center $a_0$.

It is assumed that $f(x)$ in (2.1) is a scalar function of the form $f(x) = x'Bx$, with a given symmetric positive definite $n \times n$–matrix $B$. The set $\mathcal{U}$ of admissible controls $u(\cdot)$ in (2.1) consists of all functions $u(t)$ which are measurable in Lebesgue sense on $[t_0, T]$ and such that the constraint

$$u(t) \in U \quad \text{for a.e.} \quad t \in [t_0, T] \tag{2.4}$$

is fulfilled, where $U$ is a given set, $U \in \operatorname{comp}\mathbb{R}^n$.

Let the absolutely continuous function $x(t) = x\big(t; u(\cdot), A(\cdot), x_0\big)$ be a solution to dynamical system (2.1)–(2.4) with initial state $x_0 \in X_0$, with admissible control $u(\cdot)$ and with a matrix $A(\cdot)$ satisfying (2.2)–(2.3).

**Definition 1.** *The reachable set $X(t)$ at time $t$ ($t_0 < t \le T$) of system (2.1)–(2.4) is defined as follows*

$$X(t) = \{x \in \mathbb{R}^n : \exists u(\cdot) \in \mathcal{U},\ \exists x_0 \in X_0,\ \exists A(\cdot) \in \mathcal{A},\quad x = x(t) = x(t; u(\cdot), A(\cdot), x_0)\}.$$

It is well known that the exact construction of reachable sets $X(t)$ of a control system is very difficult even for systems with a linear dynamics. The theory based on ideas of construction external and internal ellipsoidal estimates of reachable sets for systems with a linear dynamics have been deeply developed in [8, 24]. Recent new results devoted to construction of external (and in some special cases internal) set-valued estimates of reachable sets $X(t)$ for separate kinds of nonlinear systems may be found in [10–12, 14, 15, 25].

The approach presented here for estimating reachable sets of the system (2.1)–(2.3) is based on the techniques of developed ellipsoidal calculus and uses also approach connected with Hamilton–Jacobi–Bellman equations which is applied to nonlinear control systems of the class described above. Therefore this research establishes a connection between these two approaches to the estimation of unknown states of uncertain dynamical systems of the considered type.

We need to define also an additional trajectory tube $X(t; u(\cdot))$ ($t_0 < t \le T,\ u(\cdot) \in \mathcal{U}$) which depends on a control $u(\cdot)$.

**Definition 2.** *Let $u(\cdot)$ be an admissible control. The set $X(t; u(\cdot))$ at time $t$ ($t_0 < t \le T$) of system (2.1) is defined as the set*

$$X(t; u(\cdot)) = \{x \in \mathbb{R}^n : \exists x_0 \in X_0,\ \exists A(\cdot) \in \mathcal{A},\ x = x(t; u(\cdot), A(\cdot), x_0)\}.$$

Note that for each fix $t$ ($t_0 < t \le T$) and for a fixed control $u(\cdot)$ ($u(\cdot) \in \mathcal{U}$) the set $X(t; u(\cdot))$ represents the reachable set of system (2.1) taken with respect to $x_0 \in X_0$ only. Accordingly, the estimating ellipsoidal tubes and their cross-sections, generally speaking, also depend on admissible controls $u(\cdot)$, so in the formulations of main problems we would like to emphasize this circumstance by using a slightly modified notation $E(\hat{a}, \hat{Q}; T, u(\cdot))$ for estimating ellipsoids, adding a time moment and control here as additional arguments.

Thus, the main two problems considered here are as follows.

**Problem 1.** For each feasible control $u(\cdot) \in \mathcal{U}$, find the optimal (closest with respect to inclusion of sets) external ellipsoidal estimate $E(\hat{a}, \hat{Q}; T, u(\cdot))$ of the reachable set $X(T; u(\cdot))$ of the dynamical system (2.1),

$$X(T; u(\cdot)) \subset E(\hat{a}, \hat{Q}; T, u(\cdot)).$$

**Problem 2.** Given a vector $x^* \in \mathbb{R}^n$, find the feasible control $u^*(\cdot) \in \mathcal{U}$ and a number $\epsilon^* > 0$ such that

$$d(x^*, E(\hat{a^*}, \hat{Q}^*; T, u^*(\cdot))) = \inf_{u(\cdot) \in \mathcal{U}} d(x^*, E(\hat{a^*}, \hat{Q}^*; T, u(\cdot))) = \epsilon^*.$$

## 3. Main results

Here we describe the general scheme which allows to find the solutions of Problems 1–2. This scheme uses the dynamic programming ideas which are slightly modified to apply to the class of systems under study.

Let us mention first some important results [19, 21] from the optimal control theory which serve as the basis for further constructions.

Consider the control system described by the ordinary differential equation

$$\dot{x} = f(t, x, u(t)), \quad t \in [t_0, T] \tag{3.1}$$

with function $f : [t_0, T] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ measurable in $t$ and continuous in other variables. Here $x$ stands for the state vector, $t$ stands for time and control $u(\cdot)$ is a measurable function satisfying the constraints

$$u(\cdot) \in \mathcal{U} = \big\{u(\cdot) : u(t) \in U, \quad t \in [t_0, T]\big\} \tag{3.2}$$

where $U \in \operatorname{comp} \mathbb{R}^m$.

Assume that the initial condition $x(t_0)$ to the system (3.1) is unknown but bounded

$$x(t_0) = x_0, \quad x_0 \in X_0 \in \operatorname{comp} \mathbb{R}^n. \tag{3.3}$$

Assume that an absolutely continuous function $x(t) = x(t, u(\cdot), t_0, x_0)$ is a solution to (3.1) and we have $x(t_0) = x_0$ which satisfies (3.3) and a related control function $u(t)$ satisfies (3.2).

We study the control system (3.1)–(3.3) and we will assume further that $f(t, x, u)$ in (3.1) is continuous in $\{t, x, u\}$ and continuously differentiable in $x$. We also suppose that solutions to (3.1)–(3.3) may be extended to the whole interval $[t_0, T]$.

We use the notation $X(t) = X(t;\ t_0, X_0)$ for the reachable set of the system (3.1)–(3.3) at time $t$. It is well known that the set $X(t)$ may be interpreted as a level set of a value function $V(t, x)$ for an special auxiliary control problem [19, 24]. This value function for the new auxiliary problem satisfies the HJB equation of the following type

$$V_t(t, x) + \max_{u \in U} \,(V_x, f(t, x, u)) = 0.$$

Generally the value function may not be differentiable. So a solution to the HJB equation may be treated as a minmax or viscosity solution [9]. The precise solutions to such HJB-equations are difficult to find and the corresponding variational inequalities and related comparison theorems may be used to obtain approximate estimates of reachable sets [19].

### 3.1. Auxiliary constructions

The following auxiliary result will be needed further.

**Lemma 1** [21]. *Assume that there exists a function $\mu(t)$ which is integrable on $[t_0, T]$ and such that the inequality*

$$\max_{u \in U} \,(V_x, f(t, x, u)) + V_t(t, x) \le \mu(t)$$

*is fulfilled. Then the following external estimate of the reachable set $X(t)$ of the system (3.1)–(3.3) is true*

$$X(t) \subseteq \Big\{x : V(t, x) \le \int_{t_0}^{t} \mu(s) ds + \max_{x \in X_0} V(t_0, x)\Big\}, \quad t_0 \le t \le T.$$

*Remark 1.* It is known that we may take here $\mu(s) = 0$ [19].

The following more general inequality may be used also in the estimation context, namely

$$V_t(t, x) + \max_{u \in U} \,(V_x, f(t, x, u)) \le g(t, V(t, x)) \tag{3.4}$$

where $g(t, V)$ is integrable in $t \in [t_0, T]$ and is continuously differentiable in $V$.

Due to above property the ordinary differential equation

$$\dot{U}(t) = g(t, U), \quad U(t_0) = U_0, \tag{3.5}$$

is called a comparison equation for (3.1)–(3.3). We will need further the following result.

**Theorem 1** [19]. *Assume that the relations* (3.4) *and* (3.5) *are satisfied and the inequality*

$$\max_{x\in X_0} V(t_0,x) \le U_0$$

*is true. Then the upper estimate*

$$X(t) \subseteq \{x : V(t,x) \le U(t)\}, \quad t_0 \le t \le T.$$

*is valid.*

### 3.2. External estimates of reachable sets under uncertainty via HJB techniques

A number of approaches had been proposed recently to derive differential equations which describe the dynamics of external ellipsoidal estimates of reachable sets of uncertain control systems. In particular, the differential equations of ellipsoidal estimates for reachable sets of a nonlinear dynamical control system were derived in [12]. There the case was studied when state velocities of the contain special quadratic forms but in that case the presence of uncertainty in matrix coefficients was not investigated.

The following result continues this research and describes the dynamics of external ellipsoidal estimates of the reachable set $X(t) = X(t;t_0,X_0)$ $(t_0 \le t \le T)$ for the special case when $U = E(\hat a, \hat Q)$ with a center $\hat a$ and a positive definite matrix $\hat Q$ given.

First we find the smallest number $k > 0$ for which the inclusion

$$X_0 = E(a_0, Q_0) \subseteq E(a_0, k^2 B^{-1}) \tag{3.6}$$

is true, this initial step will help to get better resulting estimate for the whole trajectory tube $X(t) = X(t;t_0,X_0)$ $(t_0 \le t \le T)$. The smallest number $k > 0$ satisfying (3.6) may be determined using the procedure described, for example, in [15].

The following main result is true.

**Theorem 2.** *For any $t \in [t_0, T]$ the following inclusion is true*

$$X(t;t_0,X_0) \subseteq E(a^+(t), r^+(t)B^{-1}), \tag{3.7}$$

*here functions $a^+(t)$, $r^+(t)$ are the solutions of the following system*

$$\begin{gathered}
\dot a^+(t) = A^0 a^+(t) + ((a^+(t))'Ba^+(t) + r^+(t))d + \hat a, \quad t_0 \le t \le T,\\
\dot r^+(t) = \max_{\|l\|=1} \Big\{ l'\big(2r^+(t)B^{1/2}(A_0 + 2d(a^+(t))'B)B^{-1/2}\\
+ (q(r^+(t)))^{-1} r^+(t) B^{1/2}\hat Q^* B^{1/2}\big)l\Big\} + q(r^+(t))r^+(t),\\
q(r) = ((nr)^{-1}\mathrm{Tr}(B\hat Q^*))^{1/2},
\end{gathered} \tag{3.8}$$

*the matrix $\hat Q^*$ is positive definite and satisfies the inclusion*

$$\mathcal{A}^1 a_0 + E(0,\hat Q) + k_0 D^{1/2} B^{1/2} B(0,1) \subseteq E(0, \hat Q^*) \tag{3.9}$$

*and the initial state is the following*

$$a^+(t_0) = a_0, \quad r^+(t_0) = k^2. \tag{3.10}$$

P r o o f. Using the result of Theorem 1 and basing on the scheme of reasoning discussed in [15] and [19] with necessary modifications because of the special structure of the system (2.1)–(2.3), we derive the estimates (3.7)–(3.10). Note that the above result is essentially and ideologically close to the estimates given in [13] (see Theorem 2), but it differs significantly in details and in final relations, thereby supplementing the already existing range of methods and providing a deeper theoretical basis for solving problems of estimating the states of uncertain systems of the class under study. □

Despite the seeming cumbersomeness of the formulas describing the ellipsoid that is external in terms of inclusion for the reachable set (at the current moment of time), the calculations of the external estimates given in the Theorem 2 are fast enough (performed "in real time") and easy to implement.

The outer ellipsoids obtained by the scheme of Theorem 2 are optimal in the sense that they touch the real reachable sets at some points and cannot be reduced without violating the basic requirement to contain the estimated reachable set.

Theorem 2 solves the Problem 1 and gives the way to find an approximate solution for the Problem 2.

## 4. Example

We illustrate here the proposed state estimation scheme for a nonlinear uncertain system of the studied kind. The external estimates calculated on the base of Theorem 2 remain ellipsoidal-valued (and therefore convex) and contain reachable sets of the considered system.

*Example 1.* Consider the control system

$$\begin{cases} \dot{x}_1 = (2+\nu)x_1 + u_1, \\ \dot{x}_2 = (2+\nu)x_2 + u_2, \\ \dot{x}_3 = (2+\nu)x_3 + x_1^2 + x_2^2 + x_3^2 + u_3. \end{cases} \tag{4.1}$$

Here we take $x_0 \in X_0 = B(0,1)$, $0 \le t \le T = 0.4$ and $U = B(0, 0.1)$, a parameter $\nu$ is unknown but bounded, namely $\nu \in [0, 0.1]$. We emphasize that the constraint on the unknown parameter $\nu$ in the control system (4.1) has a different form than in the example (13) in [17], in accordance with a different formulation of the main problem studied here and because of a different technique used for its solution.

The reachable set $X(T)$ and its external ellipsoidal estimate found on the base of Theorem 2 $E(a^+(t), Q^+(t))$ for $t = T$ are shown in Fig. 1.

The following Fig. 2 for which different results [12, 17] were used is included to illustrate the possibilities of the approach in common, in particular of obtaining two-sided (external and internal) ellipsoidal estimates for the reachable sets of control systems with uncertainty. Fig. 2 presents two sets also, the same reachable set $X(T)$ and its internal estimating ellipsoid $E^-(T) = E(a^-(T), Q^-(T))$,

Here, in both Fig. 1 and Fig. 2, one can see a possible gap between the external and internal estimates of the reachable sets under study, this gap cannot be eliminated within the framework of the approach described here. We also note that the ellipsoidal estimates constructed above (each in its own class, of internal or external kind) are exact in the sense that these estimates are unimprovable (they cannot be reduced or increased, respectively) without violating the basic requirements for their construction. It can also be underlined that the algorithms for constructing these ellipsoidal estimates are very simple to implement (for example, through the Matlab system) and do not require much computation time.

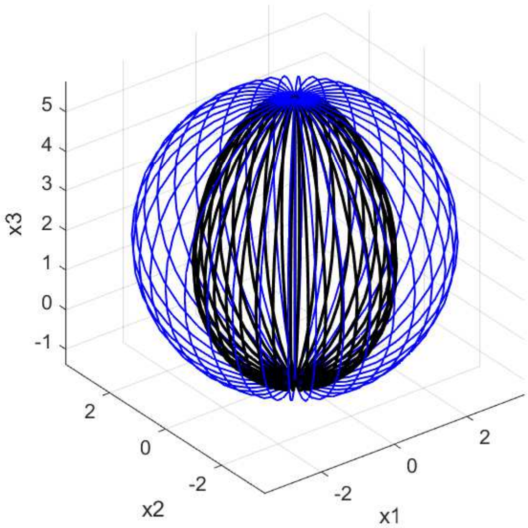



Figure 1. External estimating ellipsoid $E^+(T) = E(a^+(T), Q^+(T))$ (blue color) and the reachable set (black color) $X(T)$ in the space of $\{x_1, x_2, x_3\}$-coordinates.

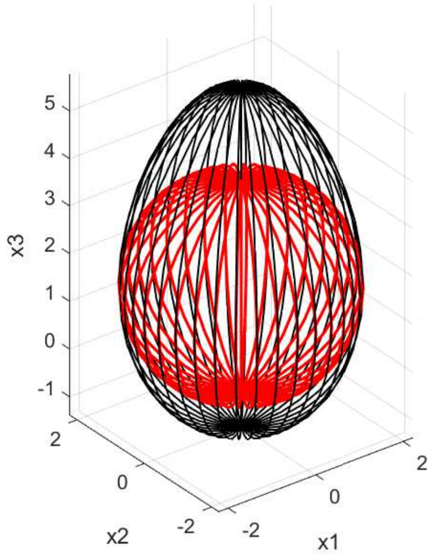



Figure 2. Internal estimating ellipsoid $E^-(T) = E(a^-(T), Q^-(T))$ (red color) and the reachable set (black color) $X(T)$ in the space of $\{x_1, x_2, x_3\}$-coordinates.

*Remark 2.* The above results are also applicable to more complicated classes of problems of control under nonlinearity and uncertainty including [1, 12, 15, 17–19, 25] and to the case of presence of additional state constraints, in this case it is also possible to use basic ideas of the research [23].

*Remark 3.* A detailed description of somewhat different, but similar in essence, approaches to solving problems of control and state estimation and using special information sets for studying control problems under uncertainty can be found in [2].

## 5. Conclusion

A new method of external estimation of the states of a nonlinear control system with uncertainty is proposed, based on the ideas and results of the theory of Hamilton–Jacobi–Bellmann equations. The relationship between the new approaches proposed here and the ideas and results of earlier studies in the theory of estimating the states of dynamical systems under conditions of uncertainty and nonlinearity is established. The possibilities of computer simulation for problems of this class are discussed. The numerical simulation results for constructing upper and inner estimates of reachable sets related to the proposed techniques and illustrating the basic ideas and algorithms are included.

## Acknowledgements

The author is sincerely grateful to two anonymous reviewers for their valuable remarks, suggestions and comments that contributed to the improvement of the text.

## REFERENCES

1. Althoff M., Stursberg O., Buss M. Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization. In: *Proc. 2008 47th IEEE Conference on Decision and Control*. IEEE Xplore, 2008. P. 4042–4048. DOI: 10.1109/CDC.2008.4738704
2. Ananyev B. I., Gusev M. I., Filippova T. F. *Upravlenie i ocenivanie sostoyanij dinamicheskikh sistem s neopredelennost'yu* [Control and Estimation of Dynamical Systems States with Uncertainty]. Novosibirsk: Izdatel'stvo SO RAN, 2018. 193 p. (in Russian)
3. Apreutesei N. C. An optimal control problem for a prey-predator system with a general functional response. *Appl. Math. Lett.*, 2009. Vol. 22, No. 7. P. 1062–1065. DOI: 10.1016/j.aml.2009.01.016
4. August E., Lu J., Koeppl H. Trajectory enclosures for nonlinear systems with uncertain initial conditions and parameters. In: *Proc. 2012 American Control Conference*. IEEE Xplore, 2012. P. 1488–1493. DOI: 10.1109/ACC.2012.6314741
5. Baier R., Gerdts M., Xausa I. Approximation of reachable sets using optimal control algorithms. *Numer. Algebra Control Optim.*, 2013. Vol. 3, No. 3. P. 519–548. DOI: 10.3934/naco.2013.3.519
6. Boscain U., Chambrion T., Sigalotti M. On some open questions in bilinear quantum control. In: *Proc. 2013 European Control Conference (ECC)*. IEEE Xplore, 2013. P. 2080–2085. DOI: 10.23919/ECC.2013.6669238
7. Ceccarelli N., Di Marco M., Garulli A., Giannitrapani A. A set theoretic approach to path planning for mobile robots. In: *Proc. 2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*. IEEE Xplore, 2004. P. 147–152. DOI: 10.1109/CDC.2004.1428621
8. Chernousko F. L. *State Estimation for Dynamic Systems*. Boca Raton: CRC Press, 1994. 304 p.
9. Crandall M. G., Lions P.-L. Viscosity solutions of Hamilton–Jacobi equations. *Trans. Amer. Math. Soc.*, 1983. Vol. 277, No. 1. P. 1–42. DOI: 10.2307/1999343
10. Filippova T. F. Ellipsoidal estimates of reachable sets for control systems with nonlinear terms. *IFAC-PapersOnLine*, 2017. Vol. 50, No. 1. P. 15355–15360. DOI: 10.1016/j.ifacol.2017.08.2460
11. Filippova T. F. Reachable sets of nonlinear control systems with state constraints and uncertainty and their estimates. *Cybern. Phys.*, 2017 . Vol. 6, No. 4. P. 185–194.
12. Filippova T. F. Estimates of reachable sets of a nonlinear dynamical system with impulsive vector control and uncertainty. In: *Proc. 2018 14th Int. Conf. "Stability and Oscillations of Nonlinear Control Systems" (Pyatnitskiy's Conference) (STAB)*. IEEE Xplore, 2018. P. 1–4. DOI: 10.1109/STAB.2018.8408353
13. Filippova T. F. The HJB approach and state estimation for control systems with uncertainty. *IFAC-PapersOnLine*, 2018. Vol. 51, No. 13. P. 7–12. DOI: 10.1016/J.IFACOL.2018.07.246
14. Filippova T. F. Differential equations for ellipsoidal estimates of reachable sets for a class of control systems with nonlinearity and uncertainty. *IFAC-PapersOnLine*, 2018. Vol. 51, No. 32. P. 770–775. DOI: 10.1016/j.ifacol.2018.11.452

15. Filippova T. F. Estimation of star-shaped reachable sets of nonlinear control systems. In: *Lecture Notes in Comput. Sci., vol. 10665: Proc. Large-Scale Scientific Computing. LSSC 2017. Lirkov I., Margenov S., Wasniewski J. (eds.)*. Cham: Springer, 2018. P. 210–218. DOI: 10.1007/978-3-319-73441-5_22
16. Filippova T. F. Control and estimation for a class of impulsive dynamical systems. *Ural Math. J.*, 2019. Vol. 5, No. 2. P. 21–30. DOI: 10.15826/umj.2019.2.003
17. Filippova T. F. Control of a bilinear system in conditions of uncertainty. In: *Proc. 2020 15th Int. Conf. Stability and Oscillations of Nonlinear Control Systems (Pyatnitskiy's Conference) (STAB)*. IEEE Xplore, 2020. P. 1–4. DOI: 10.1109/STAB49150.2020.9140458
18. Goubault E., Putot S. Robust under-approximations and application to reachability of non-linear control systems with disturbances. *IEEE Control Syst. Lett.*, 2020. Vol. 4, No. 4. P. 928–933. DOI: 10.1109/LCSYS.2020.2997261
19. Gusev M. I., Kurzhanski A. B. On the Hamiltonian techniques for designing nonlinear observers under set-membership uncertainty. *IFAC Proceedings Volumes*, 2007. Vol. 40, No. 12. P. 633–638. DOI: 10.3182/20070822-3-ZA-2920.00104
20. Kurzhanski A. B. Identification – a theory of guaranteed estimates. In: *From Data to Model. Willems J.C. (eds.)*. Berlin, Heidelberg: Springer, 1989. P. 135–214. DOI: 10.1007/978-3-642-75007-6_4
21. Kurzhanski A. B. Comparison principle for equations of the Hamilton-Jacobi type in control theory. *Proc. Steklov Inst. Math.*, 2006. Vol. 253. P. S185–S195. DOI: 10.1134/S0081543806050130
22. Kurzhanski A. B. Hamiltonian techniques for the problem of set-membership state estimation. *Internat. J. Adapt. Control Signal Process.*, 2010. Vol. 25, No. 3. P. 249–263. DOI: 10.1002/acs.1207
23. Kurzhanskii A. B., Filippova T. F. On a description of the set of viable trajectories of a differential inclusion. *Dokl. Akad. Nauk SSSR*, 1986. Vol. 289, No. 1. P. 38–41. (in Russian)
24. Kurzhanski A. B., Varaiya P. *Dynamics and Control of Trajectory Tubes. Theory and Computation.* Systems Control Found. Appl., vol. 85. Basel: Birkhäuser, 2014. 445 p. DOI: 10.1007/978-3-319-10277-1
25. Ornik M. Guaranteed reachability for systems with unknown dynamics. *Proc. 2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE Xplore, 2020. P. 2756–2761. DOI: 10.1109/CDC42340.2020.9304326
26. Polyak B. T., Nazin S. A., Durieu C., Walter E. Ellipsoidal parameter or state estimation under model uncertainty. *Automatica*, 2004. Vol. 40, No. 7. P. 1171–1179. DOI: 10.1016/j.automatica.2004.02.014

URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 43–54

DOI: 10.15826/umj.2022.1.005

# APPROXIMATION OF POSITIONAL IMPULSE CONTROLS FOR DIFFERENTIAL INCLUSIONS

**Ivan A. Finogenko**

Matrosov Institute for System Dynamics and Control Theory,
Siberian Branch of Russian Academy of Sciences,
134 Lermontova st., Irkutsk, 664033, Russian Federation

fin@icc.ru

**Alexander N. Sesekin**

Ural Federal University,
19 Mira str., Ekaterinburg, 620002, Russian Federation

Krasovskii Institute of Mathematics and Mechanics,
Ural Branch of the Russian Academy of Sciences,
16 S. Kovalevskaya Str., Ekaterinburg, 620108, Russian Federation

a.n.sesekin@urfu.ru

**Abstract:** Nonlinear control systems presented as differential inclusions with positional impulse controls are investigated. By such a control we mean some abstract operator with the Dirac function concentrated at each time. Such a control ("running impulse"), as a generalized function, has no meaning and is formalized as a sequence of correcting impulse actions on the system corresponding to a directed set of partitions of the control interval. The system responds to such control by discontinuous trajectories, which form a network of so-called "Euler's broken lines." If, as a result of each such correction, the phase point of the object under study is on some given manifold (hypersurface), then a slip-type effect is introduced into the motion of the system, and then the network of "Euler's broken lines" is called an impulse-sliding mode. The paper deals with the problem of approximating impulse-sliding modes using sequences of continuous delta-like functions. The research is based on Yosida's approximation of set-valued mappings and some well-known facts for ordinary differential equations with impulses.

**Keywords:** Positional impulse control, Differential inclusion, Impulse-sliding mode.

## 1. Introduction

We consider a dynamic system of the form

$$\dot{x}(t) \in F(t, x(t)) + B(t, x(t))u, \quad x(t_0) = x_0, \tag{1.1}$$

where $F : \mathbb{R}^1 \times \mathbb{R}^n \to \mathbb{R}^n$ is a set-valued function whose values are convex compact sets in the space $\mathbb{R}^n$ with the Euclidean norm $\| \cdot \|$, the matrix function $B(t, x)$ of dimension $n \times m$ is continuous in a set of variables, the column vector $u = (u_1, \ldots, u_m)$ is some function that describes the control action on the system.

We make the following assumptions about $F(t, x)$:

(B1) The set-valued mapping $F(t, x)$ is upper semicontinuous at each point $(t, x)$. This means that, for an arbitrary $\epsilon > 0$, there exists $\delta = \delta(t, x, \epsilon) > 0$ such that $F(t', x') \subset F^\epsilon(t, x)$ for all $(t', x') \in W_\delta(t, x)$, where $F^\epsilon(t, x)$ is an $\epsilon$-neighborhood of the set $F(t, x)$ and $W^\delta(t, x)$ is a $\delta$-neighborhood of the point $(t, x)$.

(B2) The set-valued mapping $F(t,x)$ satisfies the condition of sublinear growth: the inequality $\|w\| \le L(t)(1+\|x\|)$ with some continuous function $L(t)$ holds for any $(t,x) \in \mathbb{R}^{n+1}$ and $w \in F(t,x)$.

Conditions (B1) and (B2) ensure the existence of a solution to the differential inclusion

$$\dot{x} \in F(t,x), \tag{1.2}$$

on any segment $I = [t_0, \vartheta]$ (see, for example, [3]).

It is assumed that the matrix $B(t,x)$ satisfies the Frobenius condition

$$\sum_{\nu=1}^{n} \frac{\partial b_{ij}(t,x)}{\partial x_\nu} b_{\nu l}(t,x) = \sum_{\nu=1}^{n} \frac{\partial b_{il}(t,x)}{\partial x_\nu} b_{\nu j}(t,x).$$

This condition will ensure the uniqueness of the reaction of system (1.1) for impulse control $u$ [13, 14].

By a positional impulse control, we mean some abstract operator $(t,x) \longrightarrow U(t,x)$ that maps the space of variables $(t,x)$ into the space $m$ of vector distributions [14] according to the rule: $U(t,x) = r(t,x)\,\delta_t$, where $r(t,x)$ is a vector function with values in $\mathbb{R}^m$ and $\delta_t$ is the Dirac impulse function concentrated at the point $t$. The expression $r(t,x)\,\delta_t$ (“running impulse”) has no meaning as a generalized function and means only the fact that the system has impulse control, which implies a discrete implementation of the “running impulse” in the form of a sequence of correcting impulses concentrated at points of some partition $h\colon t_0 < t_1 < \ldots < t_N = \vartheta$ of the segment $I$. The result of such a sequential correction is a discontinuous curve $x^h(\cdot)$, here called “Euler's broken line.”

According to [14], we define a network of “Euler's broken lines” $x^h(\cdot)$ corresponding to the set of partitions $h:\ t_0 < t_1 < ... < t_p = \vartheta$ of the segment $I$. To do this, we first define a jump function by the equations

$$S(t,x,r(t,x)) = z(1) - z(0), \quad \dot{z}(\xi) = B(t,z(\xi))r(t,x), \quad z(0) = x. \tag{1.3}$$

Here we take into account that there are dependencies $S = S(t,x,r)$ and $z = z(\xi,t,x,r)$. Note also that the jump function is a vector function $S = (S^1, \ldots, S^n)$.

The jumps of the “Euler broken lines” at the points of the partitions $h$ of the segment $I$ are determined by the equations

$$S(t_i, x^h(t_i), r(t_i, x^h(t_i))) = z(1) - z(0), \quad \dot{z}(\xi) = B(t_{t_i}, z(\xi))r(t_i, x^h(t_i)),$$

with the initial conditions $z(0) = x^h(t_i)$.

On each interval $(t_i, t_{i+1}]$, “Euler's broken line” $x^h(t)$ is constructed as a function that coincides with the solution of the differential inclusion (1.2) with the initial conditions

$$x(t_i) = x^h(t_i) + S(t_i, x^h(t_i), r(t_i, x^h(t_i))), \quad x^h(t_0) = x_0, \quad i = 0, \ldots, p-1.$$

In this case, the following relations are valid:

$$x^h(t_i + 0) = x^h(t_i) + S(t_i, x^h(t_i), r(t_i, x^h(t_i))), \quad S = 0 \Leftrightarrow r = 0.$$

We assume that the following condition holds for all $(t,x)$:

$$r(t, x + S(t,x,r(t,x))) = 0. \tag{1.4}$$

This means that, after an impulsive action on the system at time $t$, the phase point $x(t)$ turns out to be on the manifold

$$\Phi = \{(t,x) : r(t,x) = 0\}.$$

In this case, the "Euler broken line" is called the impulse-sliding mode. We also assume that the functions $S(t,x,r)$ and $r(t,x)$ are continuously differentiable.

Under some assumptions, one can consider a subsequence of the sequence of Euler broken lines convergent as $d(h) = \max(t_{k+1} - t_k) \to 0$, whose limit is on the surface $\Phi$. This is called the ideal impulse-sliding mode. The purpose of the impulse control is to keep the phase point on the manifold $\Phi$.

In [10], a differential inclusion of an ideal impulse-sliding mode was obtained in the form

$$\begin{gathered}\dot{x} \in \frac{\partial S(t,x,r(t,x))}{\partial t} + \frac{\partial S(t,x,r(t,x))}{\partial r}\frac{\partial r(t,x)}{\partial t} + \\ + \left(E + \frac{\partial S(t,x,r(t,x))}{\partial x} + \frac{\partial S(t,x,r(t,x))}{\partial r}\frac{\partial r(t,x)}{\partial x}\right) F(t,x), \\ x(t_0+0) = x(t_0) + S(t_0, x(t_0), r(t_0, x(t_0))).\end{gathered} \tag{1.5}$$

In [11], a differential inclusion with discontinuous positional controls under constraints on control resources was constructed, for which the ideal impulse-sliding mode of inclusion (1.5) is an ordinary sliding mode in the sense of the theory of discontinuous systems. This makes it possible to use combinations of positional impulse and conventional discontinuous controls for controlled systems in situations without enough control resources for the latter. Note that the sliding mode of controlled systems with discontinuous feedback is the primary mode of operation and allows solving such problems as stabilization, complete controllability, and tracking (movement along a predetermined trajectory). Many studies were devoted to these issues.

Here we continue our research from the papers mentioned above and consider the approximation of Euler broken lines for system (1.1) with positional impulse control.

There are various ways to describe discontinuous trajectories (generalized solutions) of dynamical systems. One of them is to establish rules by which the trajectory jumps (see, for example, [4, 9]). If the jump function is somehow defined, then to describe the solution of the differential equation

$$\dot{x} = f(t,x) + g(t,x)\delta(t) \tag{1.6}$$

with the $\delta$-function $\delta(t)$ concentrated at a point (for convenience, at zero), one can justify the passage to the limit on the solutions of this equation after the replacement of the ideal momentum $\delta(t)$ in it by a sequence of its smooth or continuous approximations. The precise definitions are given below. For convenience, we will call them $\delta$-shaped functions.

Note that different approaches also give different concepts of a generalized solution. Even within the framework of the approximation approach, which goes back to the study of Kurzweil [12], the concept of a solution is not uniquely defined and depends on the nature of the passage to the limit (see, for example, [5, pp. 34–37]).

In this paper, we construct the approximation of "Euler broken lines" for the differential inclusion (1.1) using $\delta$-shaped functions and the jump function (1.3). The research method is based on Yosida's approximations of set-valued mappings from [7, 8] and theorems from [5] on differential equations (1.6) with $\delta$-functions in the coefficients.

## 2. Yosida's approximations

We consider Yosida's approximations for the set-valued mapping $F(t,x)$ under the following assumption.

**Condition $\mathcal{A}$.** *For any points $(t,x,y)$, the inequality*

$$(x-y)^T A(t,x)(u-v) \le l\|x-y\|^2 \tag{2.1}$$

*holds for any $u \in F(t,x)$ and $v \in F(t,y)$, where $l > 0$ is a constant and $A(t,x) = [a_{ij}(t,x)]_{i,j=1}^n$ is a symmetrical, positive definite, and continuously differentiable matrix whose eigenvalues are from a segment $[c,d]$, $0 < c \le d < +\infty$.* (In (2.1), we understand vectors as columns, and $T$ means transpose.)

Denote by $z = J_\lambda(t,x)$ the solution of the inclusion $z \in x + \lambda F(t,z)$. Let $F_\lambda(t,x) = (J_\lambda(t,x) - x)/\lambda$. Note that $J_\lambda(t,x)$ and $F_\lambda(t,x)$ are the resolvent and the Yosida approximation, respectively, for the set-valued mapping $x \to -F(t,x)$ (see [1]) for every fixed $t$. Therefore, here we also call the function $F_\lambda(t,x)$ Yosida's approximation for the set-valued mapping $(t,x) \to F(t,x)$.

*Remark 1.* Provided that $A(t,x) \equiv E$, inequality (2.1) is called the condition of right Lipschitz property and is used to study the property of right uniqueness of solutions to differential equations [5, p. 8]. In particular, it follows from the usual Lipschitz condition, which no longer gives the right uniqueness of solutions for set-valued mappings. Provided that the right-hand side of inequality (2.1) is equal to zero and the mapping $F$ does not depend on the variable $t$, Condition $\mathcal{A}$ turns into a condition of monotonicity type for set-valued mappings, which ensures the existence and some properties of the Yosida approximants [1]. Inequality (2.1) is more general than the right Lipschitz condition and the monotonicity condition. The use of the matrix $A(t,x)$ in it is convenient for studying differential equations with a matrix at the derivatives, for example, in Lagrange equations of the second kind when describing mechanical systems with discontinuous nonlinearities (dry friction and discontinuous feedbacks).

In what follows, we suppose that assumptions (B1) and (B2) and condition $\mathcal{A}$ are satisfied. The following Assertions 1–3 follow from Lemma 1 and Theorems 1 and 3 from [8].

**Assertion 1.** *For any segment $I = [a,b]$ and any bounded region $\Omega \subset \mathbb{R}^n$, there is a number $\lambda' > 0$ such that, for all $\lambda \in [0,\lambda']$ and $(t,x) \in \Omega$, the Yosida approximation $F_\lambda(t,x)$ with the following properties is uniquely defined*:

(1) *the mapping $(\lambda,t,x) \to F_\lambda(t,x)$ is continuous in $(\lambda,t,x)$ and Lipschitz in $x$. The latter means that, for each fixed $\lambda \in (0,\lambda']$, there exists a constant $L_\lambda$ such that*

$$\|F_\lambda(t,x) - F_\lambda(t,y)\| \le L_\lambda\|x-y\|$$

*holds for any $(t,x)$, $(t,y) \in I \times \Omega$;*

(2) *there are constants $l_1 > 0$ and $L > 0$ such that the following inequality holds for any $(t,x)$, $(t,y) \in I \times \Omega$ and $\lambda \in (0,\lambda']$:*

$$(x-y)^T A(t,x)\big(F_\lambda(t,x) - w\big) \le l_1\|x-y\|^2 + \lambda L; \tag{2.2}$$

(3) *for every fixed point $(t,x)$, $F_\lambda(t,x) \to m(F(t,x))$ as $\lambda \to +0$, where $m(F(t,x)) \in F(t,x)$ is the minimum point of the quadratic form $z^T A(t,x) z$ on the set $F(t,x)$.*

We define a mapping $F_\lambda(t,x)$ for $\lambda = 0$, setting $F_0(t,x) = m(F(t,x))$ for any $(t,x) \in \Omega$ and consider a one-parameter family of equations

$$\dot{x} = F_\lambda(t,x). \tag{2.3}$$

**Assertion 2.** *The following statements are valid*:

(1) *for any initial state* $(t_0, x_0)$, *equation* (2.3) *has a unique solution* $x_\lambda(t)$ *for all sufficiently small values* $\lambda > 0$;

(2) *for* $\lambda = 0$, *a solution* $x_0(t)$ *of equation* (2.3) *exists and is the right-unique solution to the differential inclusion* (1.2), *i.e., any two solutions can merge but cannot fork as* $t$ *increases.* (*Such solutions are called slow in* [1]);

(3) *for solutions* $x_\lambda(t)$ *of equations* (2.3) *with the same initial conditions,* $x_\lambda(t) \to x_0(t)$ *uniformly on any segment* $[t_0, t_1]$ *on which these solutions exist; more precisely,*

$$\|x_\lambda(t) - x_0(t)\|^2 = O(\lambda) \quad \textit{for all} \quad t \in [t_0, t_1].$$

**Assertion 3.** *Let the mapping* $F$ *on the right-hand side of the inclusion* (1.2) *and the matrix* $A$ *in inequality* (2.1) *do not depend on the variable* $t$. *Then*

(1) *for any solution* $x(t)$ *of the differential inclusion* (1.2) *defined in an interval, the function* $t \to m(F(x(t))$ *is right-continuous;*

(2) *any solution* $x(t)$ *to the inclusion* (1.2) *for* $\lambda = 0$ *is right-handed. This means that, for all* $t$ *from the domain of this solution,* $D^+x(t) = m(F(x(t))$, *where* $D^+x(t)$ *is the right-hand derivative of the function* $x(t)$.

Assertions 1–3 used below define some qualitative properties of Euler broken lines and can be useful in developing algorithms for numerical calculations.

Consider the differential inclusion

$$\dot{x} \in F(t, x) + \delta_*(t) g(t, x(t - \tau)) \tag{2.4}$$

and the equation

$$\dot{x} = F_\lambda(t, x) + \delta_*(t) g\big(t, x(t - \tau)\big), \tag{2.5}$$

where $F_\lambda(t, x)$ is Yosida's approximation of the mapping $F(t, x)$ and $\tau > 0$ is a positive parameter.

**Lemma 1.** *Let* $g(t, x)$ *be a continuous vector function satisfying the Lipschitz condition in* $x$ *with constant* $L_p$, *and let* $\delta_*(t)$ *be a continuous scalar function.*

*Then there are positive constants* $K_1$, $K_2$, $K_3$, *and* $\lambda'$ *such that, for any solutions* $x_\lambda(t)$ *and* $x(t)$ *to equations* (2.4) *and the inclusion* (2.5), *respectively, defined on the segment* $[t_0 - \tau, t_0 + T]$ *with the initial functions* $x_\lambda(t) = x_\lambda(t_0)$ *and* $x(t) = x(t_0)$ *on the segment* $[t_0 - \tau, t_0]$, *the following inequality holds for all* $t \in [t_0, t_0 + T]$ *and* $\lambda \in (0, \lambda']$:

$$\|x_\lambda(t) - x(t)\|^2 \le (K_1\lambda + K_2\|x_\lambda(t_0) - x(t_0)\|) e^{\int_{t_0}^{t_0+T} K_3|\delta_*(s)|ds}. \tag{2.6}$$

P r o o f. According to Assertion 1, there are numbers $\lambda' > 0$, $L > 0$, and $l_1 > 0$ such that, for all $\lambda \in (0, \lambda']$, a mapping $F_\lambda(t, x)$ is defined, which is continuous and Lipschitz in $x$. Define

$$\Gamma_\lambda(t, x, x') = F_\lambda(t, x) + \delta_*(t) g(t, x')$$

and take an arbitrary $w(t, y, y') \in F(t, y) + \delta_*(t) g(t, y')$. Then there is $u(t, y) \in F(t, y)$ such that $w(t, y, y') = u(t, y) + \delta_*(t) g(y')$. From inequality (2.2), we get

$$\begin{aligned} &(x - y)^T A(t, x)\big(\Gamma_\lambda(t, x, x') - w(t, y, y')\big) = \\ &= (x - y)^T A(t, x)\Big(F_\lambda(t, x) - u(t, y) + \delta_*(t)\big(p(t, x') - p(t, y')\big)\Big) = \\ &= (x - y)^T A(t, x)\big(F_\lambda(t, x) - u(t, y)\big) + \delta_*(t)(x - y)^T A(t, x)\big(p(t, x') - p(t, y')\big) \le \\ &\le l_1\|x - y\|^2 + L\lambda + |\delta_*(t)| L_p \|A(t, x)\| \|x - y\| \|x' - y'\|. \end{aligned} \tag{2.7}$$

Let $y(t) = x(t) - x_\lambda(t)$ and

$$\xi(t) = \frac{1}{2}\big(y(t)\big)^T A\big(t, x(t)\big) y(t).$$

Then

$$\dot{\xi}(t) = \big(y(t)\big)^T A\big(t, x_(t)\big)\dot{y}(t) + \frac{1}{2}\big(y(t)\big)^T \dot{A}\big(t, x(t)\big) y(t)$$

for almost all $t \in [t_0, t_0 + T]$.

In our proof, we will use some quite obvious estimates and Lipschitz conditions for functions and matrices that assumed or follow from the continuous differentiability of the matrix $A(t, x)$, as well as some well-known inequalities related to the properties and norms of matrices. In particular, we will use the following property of quadratic forms with symmetric positive definite matrices (see, for example, [2, p. 13]):

$$c\|x - y\|^2 \le (x - y)^T A(t, x)(x - y) \le d\|x - y\|^2, \tag{2.8}$$

where the segment $[c, d]$ contains all eigenvalues of the matrix $A(t, x)$ for all $(t, x)$.

Setting

$$x = x(t), \quad y = x_\lambda(t), \quad x' = x(t - \tau), \quad y' = x_\lambda(t - \tau)$$

in inequalities (2.7) and (2.8), we obtain

$$\dot{\xi}(t) \le l_2 \xi(t) + L\lambda + l_3\|\delta_*(t)\|\sqrt{\xi(t)\xi(t - \tau)} \tag{2.9}$$

with some positive constants $l_2$ and $l_3$.

Let

$$\eta(t) = \max\{\xi(s) : t_0 \le s \le t\}.$$

Then $\xi(s) \le \eta(t)$ for all $s \in [t_0 - \tau, t]$, $\xi(t') = \eta(t)$ for some $t' \in [t_0, t]$, and (2.9) implies

$$\dot{\xi}(t) \le (l_4 + l_5|\delta_*(t)|)\eta(t) + L\lambda \tag{2.10}$$

with some positive constants $l_4$ and $l_5$. Integrating (2.10), we get

$$\eta(t) = \xi(t') = \eta(t_0) + \int\limits_{t_0}^{t'} \big((l_4 + l_5|\delta_*(s)|\big)\eta(s) + L\lambda)\, ds \le \eta(t_0) + \int\limits_{t_0}^{t} \big((l_4 + l_5|\delta_*(s)|)\eta(s) + L\lambda\big)\, ds.$$

Now Granwall's lemma (see, for example, [3, p. 122]) implies

$$\eta(t) \le (\eta(t_0) + TL\lambda)e^{l_4 T} e^{\int_{t_0}^{t_0+T} l_5|\delta_*(t)|\, dt}.$$

Since $\xi(t) \le \eta(t)$ and $\eta(t_0) = \xi(t_0)$, using this inequality and inequalities (2.8) for quadratic forms, it is easy to find constants $K_1$, $K_2$, and $K_3$ such that inequality (2.6) holds.

The lemma is proved. □

## 3. Differential inclusions with delay and delta functions involved in coefficients

Consider a problem written in the form

$$\begin{cases} \dot{x}(t) \in F\big(t, x(t)\big) + \delta(t) g\big(t, x(t - 0)\big), \\ x(t_0) = x_0, \end{cases} \tag{3.1}$$

where $\delta(t)$ is the Dirac $\delta$-function concentrated at the point $t=0$, and the sequence of problems

$$\begin{cases} \dot{x}(t) \in F\big(t, x(t)\big) + \delta_i(t) g\big(t, x(t-\tau_i)\big), & i = 1, 2, \ldots, \\ x(t_0) = x_{i0} \end{cases} \tag{3.2}$$

where $x_{i0} \to x_0$ and $\delta_i(t)$ form a sequence of continuous ($\delta$-shaped) functions satisfying the conditions

(D1) $\delta_i(t) = 0$ ($t \le \alpha_i$, $t \ge \beta_i$), $\delta_i(t) \ge 0$ ($\alpha_i < t < \beta_i$), where $\alpha_i \to 0$, $\beta_i \to 0$, and $\beta_i - \alpha_i \le \tau_i \to 0$ as $i \to +\infty$;

(D2) $\int_{\alpha_i}^{\beta_i} \delta_i(t) dt \to 1$ as $i \to +\infty$.

Let us introduce auxiliary problems:

$$\dot{u} \in F(t, u), \quad u(t_0) = x_0, \quad t \in [t_0, 0]; \tag{3.3}$$

$$\dot{z} \in F(t, z), \quad z(0) = u(0) + g\big(t, u(0)\big), \quad t \in [0, t_0 + T]. \tag{3.4}$$

**Theorem 1.** *Let $g(t,x)$ be a function continuous and Lipschitz in $x$, and let the functions $\delta_i(t)$ satisfy conditions* (D1)–(D2). *Then, for any sequence of solutions $x_i(t)$ of problems* (3.2)*, the following holds as $i \to +\infty$:*

$$\begin{aligned} &x_i(t) \to u(t), \quad t_0 \le t < 0; \\ &x_i(t) \to z(t), \quad 0 < t \le t_0 + T, \end{aligned}$$

*where $u(t)$ and $z(t)$ are the solutions of the inclusions* (3.3) *and* (3.4)*, respectively.*

P r o o f. First, we consider the sequence of ordinary differential equations

$$\begin{cases} \dot{x}_i = F_\lambda(t, x_i(t)) + \delta_i(t) g\big(t, x_i(t-\tau_i)\big), \\ x_i(t_0) = x_{i0} \end{cases} \tag{3.5}$$

and related auxiliary problems

$$\begin{cases} \dot{u}^\lambda = F_\lambda(t, u^\lambda), \\ u^\lambda(t_0) = x_0, \quad t_0 \le t \le 0; \end{cases} \tag{3.6}$$

$$\begin{cases} \dot{z}^\lambda = F_\lambda(t, z^\lambda), \\ z^\lambda(0) = u^\lambda(0) + g\big(t, u^\lambda(0)\big), \quad 0 \le t \le t_0 + T. \end{cases} \tag{3.7}$$

Here $F_\lambda(t,x)$ is a continuous and Lipschitz in $x$ approximation of the Yosida set-valued mapping $F(t,x)$. Therefore, we can use the well-known results for ordinary differential equations with $\delta$-functions in the coefficients. Note that, by Assertion 1 and 2, problems (3.2)–(3.4) have right-unique solutions, and problems (3.5)–(3.7) have unique solutions for their initial data.

Let $x_i^\lambda(t)$, $i = 1, 2, \ldots$, be solutions of equations (3.5). From Theorem 4 [5, pp. 36–37] and remarks to it there, we obtain the following for any fixed $0 < \lambda < \lambda'$ as $i \to +\infty$:

$$\begin{aligned} &x_i^\lambda(t) \to u^\lambda(t), \quad t_0 \le t < 0, \\ &x_i^\lambda(t) \to z^\lambda(t), \quad 0 < t \le t_0 + T, \end{aligned} \tag{3.8}$$

where $u^\lambda(t)$ and $z^\lambda(t)$ are solutions to equations (3.6) and (3.7), respectively.

By Lemma 1, for arbitrary $\epsilon > 0$, there exist a number $\eta$ and an index $N_1$ such that, for all $t \in [t_0, t_0 + T]$, $0 < \lambda < \eta$, and $i \ge N_1$, we have

$$\|x_i(t) - x_i^\lambda(t)\| \le K\sqrt{\lambda}, \quad \forall i = 1, 2, \ldots \tag{3.9}$$

where $x_i(t)$ are the solutions to problems (3.2), and

$$\|u^\lambda(t) - u(t)\| \le K\sqrt{\lambda} \tag{3.10}$$

for all $t \in [t_0, 0]$, where $u(t)$ is the solution of the differential inclusion (3.6). The first line of (3.8) implies that, for any $t \in [t_0, 0)$ and the same value $\lambda$, there exists a number $N_2 \ge N_1$ such that

$$\|x_i^\lambda(t) - u^\lambda(t)\| < \frac{\varepsilon}{3} \tag{3.11}$$

for all $i \ge N_2$.

Let $\sqrt{\lambda} < \varepsilon/3$. Then, from (3.9)–(3.11), we get $\|x_i(t) - u(t)\| < \varepsilon$ for any fixed $t \in [t_0, 0)$ for all $i \ge N_2$.

Therefore, it is established that $x_i(t) \to u(t)$ as $i \to +\infty$ for any fixed $t \in [t_0, 0)$.

It follows from inequality (2.6) that, for the solutions $z^\lambda(t)$ of equations (3.7) and solution $w(t)$ of the inclusion (3.4), there exists $0 < \eta < \lambda'$ such that

$$\|z^\lambda(t) \to z(t)\| \le \frac{\varepsilon}{3}$$

for all $0 < \lambda < \eta$ and $t \in [0, t_0 + T]$. The second line of (3.8) implies that, for any fixed $0 < \lambda < \lambda'$ and $t \in (0, t_0 + T]$, there exists a number $N_3$ such that

$$\|x_i^\lambda(t) - w^\lambda(t)\| < \frac{\varepsilon}{3}$$

for all $i \ge N_3$. Now, in view of (3.9), similarly to the above, for any $\varepsilon > 0$, there exists a positive integer $N_4 \ge N_3$ such that

$$\|x_i(t) - w(t)\| \le \|w(t) - w^\lambda(t)\| + \|w^\lambda(t) - x_i^\lambda(t)\| + \|x_i^\lambda(t) - x_i(t)\| < \varepsilon$$

for any fixed $t \in (0, t_0 + T]$ for all $i \ge N_4$.

Hence, $x_i(t) \to z(t)$ as $i \to +\infty$ for any fixed $t \in (0, t_0 + T]$. $\square$

**Definition 1.** *By a generalized solution of inclusion* (3.1) *we mean a function* $x(t)$ *satisfying the differential inclusion* (3.3) *on the segment* $[t_0, 0]$ *and the differential inclusion* (3.4) *on* $(0, t_0+T]$ *with the initial condition* $x(+0) = x(0) + p\big(t, x(0)\big)$.

By this definition, Theorem 1 ensures the existence and structure of generalized solutions of the inclusion (3.1). A convenient convention for us is to extend the generalized solution $x(t)$ at the discontinuity point $t = 0$ by a limit on the left (which obviously exists).

Consider a differential equation of the form

$$\dot{x}(t) = F_\lambda\big(t, x(t)\big) + \delta(t)p\big(t, x(t-0)\big), \tag{3.12}$$

where $F_\lambda(t, x)$ is the Yoshida approximation of the set-valued mapping $F(t, x)$.

**Corollary 1.** *Let the assumptions of Theorem 1 be satisfied. Then there exist positive constants* $\lambda'$ *and* $K$ *such that, for any generalized solutions* $x(t)$ *and* $x^\lambda(t)$ *of problems* (3.1) *and* (3.6), *respectively, the following holds*:

$$\|x(t) - x_\lambda(t)\| \le K(\sqrt{\lambda} + \|x(t_0) - x_\lambda(t_0)\|)$$

*for all* $t \in I$ *and* $\lambda \in (0, \lambda']$.

**Corollary 2.** *Let the assumptions of Theorem 1 be satisfied, let $x_i^\lambda(t)$ be solutions of equations (3.5) such that $x_i^\lambda(t_0) \to x_0$ for $\lambda \to +0$, $i \to +\infty$, and let $x(t)$ be a generalized solution of inclusion (3.2) with the initial condition $x(t_0) = x_0$. Then, for any $\varepsilon > 0$, there are numbers $\lambda' > 0$ and $N$ such that*

$$\|x_i^\lambda(t) - x(t)\| < \varepsilon$$

*for all $0 < \lambda < \lambda'$ and $i \geq N$.*

*Remark 2.* Theorem 1 and its corollaries are formulated for differential inclusions with impulsive action at time $t = 0$. However, this does not limit the generality of the results since the change of variable $s = t - t'$ allows us to consider inclusions with impulsive action at the time $t = t'$.

*Remark 3.* In Definition 1 of the generalized solution of the differential inclusion (3.1), the set-valued mapping $F(t,x)$ can be replaced by the mapping $m(F(t,x))$ from Assertion 1. In the case when the matrix $A(t,x)$ in Condition $\mathcal{A}$ is identity, $m(F(t,x))$ is the point of the set $F(t,x)$ closest to the origin in the Euclidean norm.

## 4. Euler broken line approximation

Euler broken line approximations for system (1.1) are based on Theorem 1 and its corollaries. Suppose that assumptions (B1) and (B2) and Condition $\mathcal{A}$ are satisfied and consider a jump function $g(t,x) = S(t,x,r(t,x))$ defined by the equality (1.3).

We define a partition $h$: $t_0 < t_1 < \ldots < t_N = t_0 + T$ of the segment $I = [t_0, t_0 + T]$, and Euler broken lines $x^h(t)$, which on each interval $(t_k, t_{k+1}]$ coincide with the solutions of the Cauchy problems for the differential inclusion

$$\dot{x} \in F(t,x), \quad x(t_i) = x^h(t_k) + g\big(t_k, x^h(t_k)\big), \quad k = 0, \ldots, N-1.$$

In this case, the following conditions are satisfied for $k = 0, \ldots, N-1$:

$$x^h(t_0) = x_0, \quad x^h(t_k + 0) = x^h(t_k) + g\big(t_k, x^h(t_k)\big).$$

For the partition $h$ of the segment $I$, we introduce the sequence of problems

$$\begin{cases} \dot{x}(t) \in F\big(t, x(t)\big) + g\big(t, x(t - \tau_i^k)\big) \sum\limits_{k=0}^{N-1} \delta_i^k(t - t_k), \quad i = 1, 2, \ldots, \\ x(t_0) = x_0 + g(t_0, x_0) \end{cases} \tag{4.1}$$

as $i \to +\infty$. For each fixed $k = 1, \ldots, N-1$, we impose the following conditions on the functions $\delta_i^k(t)$:

(D1k) $\delta_i^k(t) = 0$ ($t \leq \alpha_i^k, t \geq \beta_i^k$) and $\delta_i^k(t) \geq 0$ ($\alpha_i^k < t < \beta_i^k$), where $\alpha_i^k \to 0$, $\beta_i^k \to 0$, and $\beta_i^k - \alpha_i^k \leq \tau_i^k \to 0$ as $i \to +\infty$;

(D2k) $\int\limits_{\alpha_i^k}^{\beta_i^k} \delta_i^k(t)dt \to 1$ for all $i = 1, 2, \ldots$.

Since $\alpha_i^k \to 0$, $\beta_i^k \to 0$, and $\tau_i^k \to 0$ as $i \to +\infty$, we initially consider these quantities to be so small that the intervals $(t_k + \alpha_i^k, t_k + \beta_i^k)$, $k = \overline{1, N-1}$, are pairwise disjoint.

**Theorem 2.** *Let $F(t,x)$ and $g(t,x)$ satisfy the assumptions of Lemma 1, and let the functions $\delta_i^k(t)$ satisfy conditions* (D1k)–(D2k). *Then, for every fixed partition $h$ of the segment $I$, the sequence of solutions $x_i^h(t)$ of problems* (4.1) *converges as $i \to +\infty$ to the Euler broken line $x^h(t)$ at every point $t \in I$ such that $t \neq t_k$, $k = \overline{0, N-1}$.*

P r o o f. Taking into consideration Remark 2, we apply Theorem 1 to the inclusion (4.1) on the segment $I_1^\varepsilon = [t_0, t_2 - \varepsilon]$ for an arbitrary $\varepsilon > 0$ so small that $t_1 \in I_1^\varepsilon$. Here we take into account that, starting from some number $i$, $\delta_i^2(t) = 0$ will hold for all $t \in I_1^\varepsilon$. As a result, we get

$$x_i^h(t) \to x^h(t) \tag{4.2}$$

for any $t \in I_1^\varepsilon$, $t \neq t_1$ and $t \neq t_0$. Then, in view of the right uniqueness of the solutions of the inclusion $\dot{x} \in F(t,x)$ and the arbitrariness of $\varepsilon > 0$, we conclude that (4.2) holds at all points of the segment $[t_0, t_2]$ except for the points $t_k$, $k = 0, 1, 2$.

Now, as initial data, we take some point $s \in (t_1, t_2)$ (for example, the midpoint of this interval) and the value $x_i^h(s)$ of the Euler broken line at this point. Applying similar reasoning to the segment $I_2^\varepsilon = [s, t_3 - \varepsilon]$ and taking into account the right uniqueness of the solutions, we conclude that (4.2) holds at all points of the segment $[t_0, t_3]$ except for the points $t_k$, $k = 0, 1, 2, 3$. Here we took into account that, starting from some number $i$, $\delta_i^k(t) = 0$ holds for all $t \in I_2^\varepsilon$, $k = 1, 2$. This process continues up to the point $t_{N-1}$, and at this last step, we consider the segment $[s, t_0 + T]$, where $s$ is the midpoint of the segment $[t_{N-2}, t_{N-1}]$. □

Consider the problems

$$\begin{cases} \dot{x}(t) \in F\big(t, x(t)\big) + p\big(x(t-0)\big) \sum\limits_{k=1}^{N-1} \delta(t - t_k), \\ x(t_0) = x_0 + g(x_0). \end{cases} \tag{4.3}$$

$$\begin{cases} \dot{x}(t) = F_\lambda\big(t, x(t)\big) + p\big(x(t-0)\big) \sum\limits_{k=1}^{N-1} \delta(t - t_k), \\ x(t_0) = x_0 + g(x_0), \end{cases} \tag{4.4}$$

where $F_\lambda(t,x)$ is the Yoshida approximation for $F(t,x)$.

For (4.3) and (4.4), the concepts of generalized solutions $x(t)$ and $x_\lambda(t)$ are introduced by analogy with Definition 1.

**Corollary 3.** *Let all assumptions of Theorem 2 be satisfied. Then, for any fixed partition $h$ of the segment $I$, there exists a constant $K$ depending on the number $N$ of points of the partition $h$ such that, for any generalized solutions $x(t)$ and $x_\lambda(t)$ of the inclusions* (4.3) *and equations* (4.4), *respectively, the inequality*

$$\|x(t) - x_\lambda(t)\| \le K\sqrt{\lambda}$$

*holds for any $0 < \lambda < \lambda'$ and $t \in I$.*

P r o o f. The proof follows from the successive application of Corollary 3 to the segments $[t_{k-1}, t_k]$ and the initial conditions $x(t_{k-1} + 0)$ and $x_\lambda(t_{k-1} + 0)$ for $k = \overline{1, N-1}$. □

**Corollary 4.** *Let all assumptions of Theorem 2 be satisfied. Then, for any fixed partition $h$ of the segment $I$, there exists a constant $K$ depending on the number $N$ of points of the partition $h$*

*such that, for any generalized solutions $x_\lambda(t)$ of equation (4.4) and "Euler's broken lines" $x^h(t)$ of inclusions $\dot{x} \in F(t,x)$, the inequality*

$$\|x^h(t) - x_\lambda(t)\| \le K\sqrt{\lambda}$$

*holds for any $0 < \lambda < \lambda'$ and $t \in (t_0, t_0 + T]$.*

P r o o f. The definition of the Euler broken line $x^h(t)$ implies that, on the interval $(t_0, t_0+T]$, it coincides with the generalized solution to the inclusion (4.3), and then the corollary follows from Corollary 3. □

*Remark 4.* Let condition (1.4) be satisfied. Then, using Theorem 2 and its corollaries, we can formulate statements about the approximation of ideal impulse-sliding modes, which satisfy the inclusion (1.5).

## 5. Conclusion

Let us make a number of concluding remarks.

1. The Yosida approximation has a rather complex structure, and for its applications, it is necessary to calculate the resolution $J_\lambda$, which reduces to finding fixed points of set-valued mappings. It is not always possible to solve such a problem in an analytical form in the general case. At the same time, the results of Sections 2–4 remain valid for any other continuous approximations of set-valued mappings for which inequality (2.2) holds. For a set-valued function $u(x) = \operatorname{sgn} x$ equal to $-1$ for $x > 0$, 1 for $x < 0$, and segment $[-1, 1]$ for $x = 0$, the Yosida approximation is

$$u_\lambda(x) = \begin{cases} x/\lambda, & |x| \le \lambda; \\ \operatorname{sgn} x, & |x| > \lambda. \end{cases}$$

But instead of it, for example, the function $u_\lambda(x) = 2/\pi \cdot \arctan(\lambda x)$ can be used.

2. Consider the system

$$P(t,x)\dot{x} \in R(t,x) - H(t,x)\operatorname{sgn} x, \tag{5.1}$$

where $P(t,x)$ is a symmetric, positive definite $n \times n$ matrix, $H(t,x)$ is a diagonal $n \times n$ matrix with nonzero elements, and $\operatorname{sgn} x = (\operatorname{sgn} x_1 \times \cdots \times \operatorname{sgn} x_n)$ is a set-valued function. If the function $R(t,x)$ and the elements of the matrices $P$ and $H$ are continuous and locally Lipschitz functions in $x$, then the mapping $F(t,x) = P^{-1}(t,x)(R(t,x) - H(t,x)\operatorname{sgn} x)$ satisfies assumptions (B1) and (B2) and Condition $\mathcal{A}$ with matrix $A(t,x) = P(t,x)$, and all the results of Sections 2–4 are valid. In this case, to approximate $\operatorname{sgn} x_i$, one can use functions of the form $u_\lambda(x_i)$.

If the matrix $P(t,x)$ and the function $H(t,x)$ are constant, then, using the scheme of the proof of Lemma 1, we can obtain the estimate

$$v(t) \le \frac{H\gamma\lambda}{2L}(e^{2L/\gamma(t-t_0)} - 1),$$

where $L$ is the Lipschitz constant of the function $f(t,x)$ in the variable $x$, $\gamma$ is the smallest eigenvalue of the matrix $P$, $H = \sum_{i=1}^n H_i$, and

$$v(t) = (x(t) - x_\lambda(t))^T P(x(t) - x_\lambda(t)), \quad x(t_0) = x_\lambda(t_0).$$

3. A large class of systems that lead to differential inclusions are differential equations with a discontinuous right-hand side. As stated in [6], near the discontinuity points of the function $F(t,x)$, the approximate solution obtained by some numerical method usually differs from the exact one by $O(h)$, where $h$ is an integration step, regardless of the order of accuracy of the approximate method. The guaranteed in Lemma 1 accuracy of the approximation of solutions to the differential inclusion (1.2) by solutions to the approximating inclusions (2.3) is of order $O(\sqrt{\lambda})$. Numerical experiments using computers and graphical visualization of the integration results have shown that if an integration step is much smaller than the parameter $\lambda$, then the qualitative behavior of the approximate solutions to the approximating equations is closer to the exact sliding modes of the original discontinuous equations; in the presence of discontinuities of the signature type, there are no sawtooth curves characteristic of approximate solutions to discontinuous systems.

4. Note that discontinuous characteristics, as a rule, are included in systems of equations in the form of terms or factors with continuous functions. Therefore, for some classes of systems, it is expedient to approximate discontinuous characteristics directly.

## REFERENCES


1. Aubin J.-P., Ekeland I. *Applied Nonlinear Analysis.* NY: Willey & Sons Inc., 1984. 532 p.
2. Barbashin E. A. *Funkcii Liapunova* [Lyapunov functions]. Moscow: Nauka, 1970. 240 p. (in Russian)
3. Borisovich Yu. G., Gel'man B. D., Myshkis A. D., Obukhovskii V. V. *Vvedenie v teoriyu mnogoznachnykh otobrazhenii i differencial'nykh vklyuchenii* [Introduction to the Theory of Set-Valued Mappings and Differential Inclusions]. M.: URSS, 2005. 216 p. (in Russian)
4. Dykhta V. A., Samsonyuk O. N. *Optimal'noe impul'snoe upravlenie s prilozheniyami* [Optimal impulse control with applications]. M.: FIZMATLIT, 2003. 256 p. (in Russian)
5. Filippov A. F. *Differential Equations with Discontinuous Righthand Sides.* Math. Appl. Ser., vol. 18. Netherlands: Springer Science+Business Media, 1988. 304 p. DOI: 10.1007/978-94-015-7793-9
6. Filippov A. F. On approximate computation of solutions of ordinary differential equations with discontinuous right hand sides. *Vestnik Moskov. Univ. Ser. 15. Vychisl. Mat. Kibernet.*, 2001. No. 2. P. 18–20.
7. Finogenko I. A. On the right Lipschitz condition for differential equations with piecewise continuous right-hand sides. *Differ. Equ.,* 2003. Vol. 39, No. 8. P. 1124–1131. DOI: 10.1023/B:DIEQ.0000011286.59758.0d
8. Finogenko I. A. On continuous approximations and right-sided solutions of differential equations with piecewise continuous right-hand sides. *Differ. Equ.*, 2005. Vol. 41, No. 5. P. 677–686. DOI: 10.1007/s10625-005-0202-6
9. Finogenko I. A., Ponomarev D. V. On differential inclusions with positional discontinuous and pulse controls. *Tr. Inst. Mat. Mekh. UrO RAN*, 2013. Vol. 19, no. 1. P. 284–299. (in Russian)
10. Finogenko I. A., Sesekin A. N. Impulse position control for differential inclusions. *AIP Conf. Proc.,* 2018. Vol. 2048, No. 1. Art. no. 020008. DOI: 10.1063/1.5082026
11. Finogenko I. A., Sesekin A. N. Positional impulse and discontinuous controls for differential inclusions. *Ural Math. J.*, 2020. Vol. 6, No. 2. P. 68–75. DOI: 10.15826/umj.2020.2.007
12. Kurzweil J. Generalized ordinary differential equations. *Czechoslovak Math. J.*, 1958. Vol. 8, No. 3. P. 360–388.
13. Miller B. M., Rubinovich E. Ya. Discontinuous solutions in the optimal control problems and their representation by singular space-time transformations. *Autom. Remote Control*, 2013. Vol. 74. P. 1969–2006. DOI: 10.1134/S0005117913120047
14. Zavalishchin S. T., Sesekin A. N. *Dynamic Impulse Systems: Theory and Applications.* Dordrecht: Kluwer Academic Publishers, 1997. 256 p. DOI: 10.1007/978-94-015-8893-5

**URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 55–63**

DOI: 10.15826/umj.2022.1.006

# ON SOLVING AN ENHANCED EVASION PROBLEM FOR LINEAR DISCRETE–TIME SYSTEMS

**Elena K. Kostousova**

Krasovskii Institute of Mathematics and Mechanics,
Ural Branch of the Russian Academy of Sciences,
16 S. Kovalevskaya Str., Ekaterinburg, 620108, Russian Federation

kek@imm.uran.ru

**Abstract:** We consider the problem of an enhanced evasion for linear discrete-time systems, where there are two conflicting bounded controls and the aim of one of them is to be guaranteed to avoid the trajectory hitting a given target set at a given final time and also at intermediate instants. First we outline a common solution scheme based on the construction of so called solvability tubes or repulsive tubes. Then a much more quick and simple for realization method based on the construction of the tubes with parallelepiped-valued cross-sections is presented under assumptions that the target set is a parallelepiped and parallelotope-valued constraints on controls are imposed. An example illustrating this method is considered.

**Keywords:** Linear control systems, Discrete-time systems, Uncertainty, Evasion problem, Parallelepipeds.

## 1. Introduction

We consider linear discrete-time systems under conflicting controls that may have different aims. Namely, the aim of the one control may be to guide the trajectory to a given target set, and the aim of the other control may be the opposite. This gives rise to two subproblems under conditions of uncertainty, namely, the approach problem and the evasion one.

There are well known approaches for solving the problems of such sort that are based on the construction of special tubes of trajectories known as stable bridges or solvability tubes [2, 16–19]. Since an exact construction of trajectory tubes is usually a very complicated problem, different numerical methods are being devised. Of the many works, we only indicate as examples [1–15, 17–22, 24], including those based on estimation of sets by more simple sets such as ellipsoids [1, 2, 4–6, 17–20] and parallelepipeds/parallelotopes [9–15, 21, 22]. In particular, for discrete-time systems, ellipsoidal and polyhedral solution schemes have been developed for the terminal target approach problem and the terminal evasion problem at a given final time (see, for example, [2, 11, 12, 20] and [14] for both problems respectively). But two methods presented in [14] guarantee the evasion from the given set only at the given final time. Computer simulations corroborate this.

The present paper is devoted to solving the problem of enhanced evasion for linear discrete-time systems with two bounded controls, where the aim of one of them is to avoid, regardless of the actions of the other, the trajectory hitting the given target set not only at the final time, but also at intermediate instants. First the common solution scheme is outlined. Then a much more quick and simple for realization method based on the construction of repulsive parallelepiped-valued tubes is presented. In fact, here, in contrast to [14], a pair of polyhedral tubes is constructed and explicit formulas for feedback control strategies are given on the base of both tubes. A corresponding illustrative example is included.

Note that the solutions to the evasion problems may be useful for construction of dangerous disturbances, for example, in problems of aircraft control [3, 22].

We will use the following notation. Let the symbol $\mathbb{R}^n$ denotes the $n$-dimensional vector space; $\|x\|_\infty = \max_{1\le i\le n} |x_i|$ be the vector norm for $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ (we use $\top$ as the transposition symbol); $\mathbb{R}^{n\times m}$ be the space of real $n\times m$-matrices $A = \{a_i^j\} = \{a^j\}$ with elements $a_i^j$ and columns $a^j$ (the superscript numbers the columns of the matrix and the subscript numbers the components of vectors). Let $\operatorname{diag}\pi$ be the diagonal matrix $A$ with $a_i^i = \pi_i$ (where $\pi_i$ are the components of the vector $\pi$); $\operatorname{Abs} A = \{|a_i^j|\}$ for $A = \{a_i^j\} \in \mathbb{R}^{n\times m}$. Let the symbol $I$ stands for the identity matrix and $0$ stands for zero matrices and vectors; $\det A$ be the determinant of $A \in \mathbb{R}^{n\times n}$. The value of $\operatorname{sign} z$ is equal to $-1$, $0$, $1$ for $z < 0$, $z = 0$, $z > 0$ respectively. We use the notation $k = 1, \dots, N$ instead of $k = 1, 2, \dots, N$ for the sake of brevity.

## 2. Problem statements

Consider the following discrete-time system

$$x[k] = A[k]\,x[k-1] + B[k]u[k] + C[k]v[k], \quad k = 1, \dots, N, \tag{2.1}$$

where

$$u[k] \in \mathcal{R}[k], \quad v[k] \in \mathcal{Q}[k], \quad k = 1, \dots, N, \tag{2.2}$$

with the target set $\mathcal{M} \subset \mathbb{R}^n$. Here $x[k] \in \mathbb{R}^n$ are states, the matrices $A[k] \in \mathbb{R}^{n\times n}$, $B[k] \in \mathbb{R}^{n\times n_u}$, $C[k] \in \mathbb{R}^{n\times n_v}$ are given and matrices $A[k]$ are nonsingular; $u[k] \in \mathbb{R}^{n_u}$ and $v[k] \in \mathbb{R}^{n_v}$ are controls, which may have different aims; $\mathcal{M}$, $\mathcal{R}[k]$, and $\mathcal{Q}[k]$ are given compact sets. The functions $u[\cdot]$ and $v[\cdot]$ satisfying (2.2) are called *admissible*.

We consider the evasion problem, where the aim of $v$ is to guarantee $x[N] \notin \mathcal{M}$ and moreover $x[k] \notin \mathcal{M}$ for all $k = 1, \dots, N-1$ regardless of the admissible realizations of $u$, and formulate it as follows.

**Problem 1** (Evasion problem, enhanced evasion problem). *For system (2.1)–(2.2), find sets $\hat{\mathcal{W}}[k]$, $k = 0, 1, \dots, N$, satisfying $\hat{\mathcal{W}}[k] \supseteq \mathcal{M}$, $k = 1, \dots, N$, and find a corresponding feedback control strategy $v = v[k, x]$ satisfying $v[k, x] \in \mathcal{Q}[k]$, $k = 1, \dots, N$, such that each solution $x[\cdot]$ to the equation*

$$x[k] = A[k]\,x[k-1] + B[k]u[k] + C[k]v[k, x[k-1]], \quad k = 1, \dots, N,$$

*that starts from any initial point $x[0] = x^0$ with $x^0 \notin \hat{\mathcal{W}}[0]$ would satisfy $x[k] \notin \hat{\mathcal{W}}[k]$, $k = 1, \dots, N$, for all admissible functions $u[\cdot]$.*

Similarly to [3, 22] the multivalued function $\hat{\mathcal{W}}[\cdot]$ with the crossections $\hat{\mathcal{W}}[k]$, $k = 0, 1, \dots, N$, can be called a repulsive tube.

It is possible to formulate this problem in a form more close to statements and constructions from [16, Ch. 2] if we consider the solvability tube $\check{\mathcal{W}}[\cdot]$ with cross-sections $\check{\mathcal{W}}[k] = \mathbb{R}^n \setminus \hat{\mathcal{W}}[k]$, where the symbol $\mathbb{R}^n \setminus \mathcal{X}$ means the complement of $\mathcal{X}$: $\mathbb{R}^n \setminus \mathcal{X} = \{x \in \mathbb{R}^n \,|\, x \notin \mathcal{X}\}$. The term "enhanced evasion problem" is used to emphasize the difference from the following problem.

**Problem 1′** (Terminal evasion problem). *It is similar to Problem 1, but we require $\hat{\mathcal{W}}[k] \supseteq \mathcal{M}$ only for $k = N$.*

As it will be presented below the solution to Problem 1 can be obtained through relations which involve rather labor-consuming operations with sets, namely, Minkowski's sum ($\mathcal{X}^1 + \mathcal{X}^2 = \{y|\ y = x^1 + x^2, x^k \in \mathcal{X}^k\}$), Minkowski's difference ($\mathcal{X}^1 \dot{-} \mathcal{X}^2 = \{y \,|\, y + \mathcal{X}^1 \subseteq \mathcal{X}^2\}$), union, and intersection.

Therefore we consider a similar polyhedral evasion problem under the assumptions which are similar to [14].

**Assumption 1.** *Let matrices* $A[k]$ *in* (2.1) *be nonsingular*: $\det A[k] \neq 0$, $k = 1, \dots, N$, *the sets* $\mathcal{R}[k]$ *and* $\mathcal{Q}[k]$ *in* (2.2) *be parallelotopes, and the target set* $\mathcal{M}$ *be a parallelepiped*:

$$\begin{gathered} \mathcal{R}[k] = \mathcal{P}[r[k], \bar{R}[k]], \quad \bar{R}[k] \in \mathbb{R}^{n_u \times n_u}; \qquad \mathcal{Q}[k] = \mathcal{P}[q[k], \bar{Q}[k]], \quad \bar{Q}[k] \in \mathbb{R}^{n_v \times n_v}; \\ \mathcal{M} = \mathcal{P}(p_{\mathrm{f}}, P_{\mathrm{f}}, \pi_{\mathrm{f}}). \end{gathered} \tag{2.3}$$

By a *parallelepiped* $\mathcal{P}(p, P, \pi) \subset \mathbb{R}^n$ we call a set defined as

$$\mathcal{P} = \mathcal{P}(p, P, \pi) = \{x \in \mathbb{R}^n \,|\, x = p + \sum_{i=1}^{n} p^i \pi_i \xi_i, \ \|\xi\|_\infty \leq 1\}.$$

Here $p \in \mathbb{R}^n$; $P = \{p^i\} \in \mathbb{R}^{n \times n}$ is a nonsingular matrix ($\det P \neq 0$) with columns $p^i$ such that $\|p^i\|_2 = 1$; $\pi \in \mathbb{R}^n$, $\pi \geq 0$. We call $p$ the center of the parallelepiped and $P$ its orientation matrix. Note that the above conditions $\|p^i\|_2 = 1$ for the Euclidean norm may be omitted to simplify formulas. We say that a parallelepiped is *nondegenerate* if all $\pi_i > 0$.

By a *parallelotope* $\mathcal{P}[p, \bar{P}] \subset \mathbb{R}^n$ we call a set defined as

$$\mathcal{P} = \mathcal{P}[p, \bar{P}] = \{x \in \mathbb{R}^n \,|\, x = p + \bar{P}\xi, \ \|\xi\|_\infty \leq 1\}.$$

Here $p \in \mathbb{R}^n$ and $\bar{P} = \{\bar{p}^i\} \in \mathbb{R}^{n \times m}$, $m \leq n$, i.e., the matrix $\bar{P}$, which determines the shape, may be singular and not square. We say that a parallelotope $\mathcal{P}$ is *nondegenerate* if $m = n$ and $\det \bar{P} \neq 0$. Note that each parallelepiped $\mathcal{P}(p, P, \pi)$ is a parallelotope $\mathcal{P}[p, \bar{P}]$, where $\bar{P} = P \cdot \operatorname{diag} \pi$, and each nondegenerate parallelotope $\mathcal{P}[p, \bar{P}]$ is a parallelepiped $\mathcal{P}(p, P, \pi)$ with $P = \bar{P}$, $\pi = (1, 1, \dots, 1)^\top$.

**Problem 2** (Polyhedral evasion problem). *Under Assumption 1, find a solution of Problem 1 in a class of polyhedral tubes* $\mathcal{P}[\cdot] = \mathcal{P}(p[\cdot], P[\cdot], \pi[\cdot])$ *with parallelepiped-valued cross-sections. Moreover introduce a family of such tubes* $\mathcal{P}[\cdot]$ (*i.e. instead of* $\hat{\mathcal{W}}[\cdot]$ *there are the tubes* $\mathcal{P}[\cdot]$).

Recall that in [11, 12] the solutions to terminal target polyhedral approach problems are given even for more general classes of systems, namely, for systems (2.1) with uncertainties / controls in the matrices $A[k]$ and with state constraints. There the families of the tubes $\mathcal{P}^-[\cdot] = \mathcal{P}[p^-[\cdot], \bar{P}^-[\cdot]]$ with parallelotope-valued cross-sections and corresponding control strategies $u[k, x]$ have been constructed to guarantee $x[N] \in \mathcal{M}$ at the given final time $N$.

In [14], the following problem was considered.

**Problem 2′** (Polyhedral terminal evasion problem). *Under Assumption 1, find a solution of Problem 1′ in a class of polyhedral* (*parallelotope-valued*) *tubes* $\mathcal{P}^+[\cdot] = \mathcal{P}[p^+[\cdot], \bar{P}^+[\cdot]]$.

In [14], two techniques to solve Problem 2′ to ensure $x[N] \notin \mathcal{M}$ were presented using two families of the tubes $\mathcal{P}^+[\cdot] = \mathcal{P}[p^+[\cdot], \bar{P}^+[\cdot]]$ and $\mathcal{P}^{\mathrm{e}}[\cdot] = \mathcal{P}[p^{\mathrm{e}}[\cdot], \bar{P}^{\mathrm{e}}[\cdot]]$.

In the present paper, the first of these techniques is extended to solve Problem 2. Note that now we will use the tubes $\mathcal{P}[\cdot]$ with parallelepiped-valued cross-sections because this is more convenient in order to take into account the set $\mathcal{M}$ at times $k < N$.

## 3. Main results

To solve Problem 1 let us consider the following system of recurrence relations for calculating the tubes $\hat{\mathcal{W}}[\cdot]$ and $\hat{\mathcal{W}}^1[\cdot]$:

$$\begin{gathered} \hat{\mathcal{W}}^0[k-1] = \hat{\mathcal{W}}[k] + (-B[k]\mathcal{R}[k]), \quad k = N, \dots, 1; \\ \hat{\mathcal{W}}^1[k-1] = A[k]^{-1}(\hat{\mathcal{W}}^0[k-1] \dot{-} C[k]\mathcal{Q}[k]), \quad k = N, \dots, 1; \\ \hat{\mathcal{W}}[k-1] = \hat{\mathcal{W}}^1[k-1] \bigcup \mathcal{M}, \quad k = N, \dots, 2; \\ \hat{\mathcal{W}}[N] = \mathcal{M}; \quad \hat{\mathcal{W}}[0] = \hat{\mathcal{W}}^1[0]. \end{gathered} \tag{3.1}$$

**Theorem 1.** *Let the tubes $\hat{\mathcal{W}}[\cdot]$ and $\hat{\mathcal{W}}^1[\cdot]$ satisfy (3.1) and all of their cross-sections appear to be nonempty. Then the tube $\hat{\mathcal{W}}[\cdot]$ together with the control strategy $v[k,x]$ of the following form*

$$
\begin{gathered}
v[k,x] \in \begin{cases} \mathcal{V}[k,x] \text{ for } x \notin \hat{\mathcal{W}}^1[k{-}1]; \\ \mathcal{Q}[k], \text{ otherwise}, \end{cases} \\
\mathcal{V}[k,x] = \mathcal{Q}[x] \bigcap \{v \mid C[k]v \in (\mathbb{R}^n \setminus \hat{\mathcal{W}}^0[k{-}1]) - A[k]x\}
\end{gathered}
\tag{3.2}
$$

*give a solution to Problem 1.*

P r o o f (Sketch of the proof). The lines of reasoning from [2, 10] with necessary modifications can be used. First the relations for $\check{\mathcal{W}}[k] = \mathbb{R}^n \setminus \hat{\mathcal{W}}[k]$ and $\check{\mathcal{W}}^1[k] = \mathbb{R}^n \setminus \hat{\mathcal{W}}^1[k]$ can be written using duality interconnections basing, for example, on [23, p. 137]. Let $\check{\mathcal{W}}[\cdot]$ and $\check{\mathcal{W}}^1[\cdot]$ be found. Inclusions $\check{\mathcal{W}}[k] \subseteq \mathbb{R}^n \setminus \mathcal{M}$ and $\hat{\mathcal{W}}[k] \supseteq \mathcal{M}$ follow from (3.1). Then we can verify, for any $k$, that if $x = x[k{-}1] \in \check{\mathcal{W}}^1[k{-}1]$, then we obtain $\mathcal{V}[k,x] \neq \emptyset$ and

$$
x[k] = A[k]x + B[k]u[k] + C[k]v[k,x] \in \check{\mathcal{W}}[k]
$$

for any $v[k,x] \in \mathcal{V}[k,x]$ and arbitrary $u[k] \in \mathcal{R}[k]$. □

So, to guarantee $x[k] \notin \mathcal{M}$ for all $k = 1,\dots,N$ we first need to find the tubes $\hat{\mathcal{W}}[\cdot]$ and $\hat{\mathcal{W}}^1[\cdot]$ by solving recurrence relations (3.1) backward starting from $\hat{\mathcal{W}}[N] = \mathcal{M}$. Then starting from any $x^0 \notin \hat{\mathcal{W}}[0]$ we can apply an arbitrary control strategy $v$ that satisfies (3.2). According to the proof, if $x^0 \notin \hat{\mathcal{W}}[0]$, then only the first line in (3.2) can be implemented.

Also note that, in general, the sets $\hat{\mathcal{W}}[k]$ satisfying (3.1) are not guaranteed to be convex even if $\mathcal{R}[k]$, $\mathcal{Q}[k]$, and $\mathcal{M}$ are convex, and the sets $\mathbb{R}^n \setminus \hat{\mathcal{W}}^0[k{-}1]$ are the sets with holes.

To solve Problem 2, we use elementary external polyhedral estimates for results of operations with sets. Recall that the result of a linear transformation of a parallelepiped is a parallelepiped or a parallelotope. The so called touching external estimate $\boldsymbol{P}_V^+(\mathcal{Q})$ for the set $\mathcal{Q}$ with the orientation matrix $V$ can be found on the base of the values of the support function for $\mathcal{Q}$ [15]. It is easy to find touching estimates $\boldsymbol{P}_V^+(\mathcal{P}^1 + \mathcal{P}^2)$ and $\boldsymbol{P}_V^+(\mathcal{P}^1 \bigcup \mathcal{P}^2)$ for a sum and for a union of parallelepipeds/parallelotopes using explicit formulas [15]. Minkowski's difference $\mathcal{P}^1 \dot{-} \mathcal{P}^2$ of a parallelepiped and a parallelotope is either a parallelepiped or an empty set (concrete formulas can be found in [9]).

Notice that to check whether a point $x$ belongs to the parallelepiped $\mathcal{P}(p,P,\pi)$ it is useful to use relative coordinates $\xi = P^{-1}(x-p)$.

**Lemma 1.** *Given $x \in \mathbb{R}^n$ and $\mathcal{P} = \mathcal{P}(p,P,\pi)$, let $\xi = P^{-1}(x-p)$. Then $x \notin \mathcal{P}$ iff $|\xi_{i_*}| > \pi_{i_*}$ for some $i_* \in \{1,\dots,n\}$.*

To solve Problem 2 we introduce the system of the following recurrence relations for calculating parallelepipeds $\mathcal{P}[k] = \mathcal{P}(p[k],P[k],\pi[k])$ and $\mathcal{P}^1[k] = \mathcal{P}(p^1[k],P^1[k],\pi^1[k])$, which determine the couple of the tubes $\mathcal{P}[\cdot]$ and $\mathcal{P}^1[\cdot]$:

$$
\begin{gathered}
\mathcal{P}^0[k{-}1] = \boldsymbol{P}^+_{P[k]}(\mathcal{P}[k] + (-B[k]\mathcal{R}[k])), \quad k = N,\dots,1; \\
\mathcal{P}^1[k{-}1] = A[k]^{-1}(\mathcal{P}^0[k{-}1] \dot{-} C[k]\mathcal{Q}[k]), \quad k = N,\dots,1; \\
\mathcal{P}[k{-}1] = \boldsymbol{P}^+_{P[k-1]}(\mathcal{P}^1[k{-}1] \bigcup \mathcal{M}), \quad k = N,\dots,2; \\
\mathcal{P}[N] = \boldsymbol{P}^+_{P[N]}(\mathcal{M}); \quad \mathcal{P}[0] = \mathcal{P}^1[0].
\end{gathered}
\tag{3.3}
$$

Given the tubes $\mathcal{P}[\cdot]$ and $\mathcal{P}^1[\cdot]$, let us introduce the notation:

$$\xi[k,x] = P^1[k-1]^{-1}(x - p^1[k-1]);$$
$$\Theta[k] = P[k]^{-1}C[k]\bar{Q}[k];$$
$$\Phi[k,x] = (\operatorname{diag}\pi^1[k-1])^{-1}\operatorname{Abs}\xi[k,x].$$

Here $\xi[k,x]$ stands for the relative coordinates of $x$ with respect to the cross-section

$$\mathcal{P}^1[k-1] = \mathcal{P}(p^1[k-1], P^1[k-1], \pi^1[k-1])$$

of the tube $\mathcal{P}^1[\cdot]$; the matrix $\Theta[k]$ is determined by the parameters of system (2.1), (2.2), and (2.3), and by the cross-section $\mathcal{P}[k] = \mathcal{P}(p[k], P[k], \pi[k])$ of the tube $\mathcal{P}[\cdot]$; the vector $\Phi[k,x]$ is determined by $\xi[k,x]$ and the cross-section $\mathcal{P}^1[k-1]$ of the tube $\mathcal{P}^1[\cdot]$.

Then we can apply several formulas for construction of the control strategy $v[k,x]$ basing on the main formula of the form

$$\begin{gathered} v^0[k,x] = q[k] + \bar{Q}[k]\chi[k,x], \\ \chi_j[k,x] = \operatorname{sign}\Theta^j_{i_*}[k]\cdot\operatorname{sign}\xi_{i_*}[k,x], \quad j = 1,\dots,n_v. \end{gathered} \tag{3.4}$$

Let us consider three following variants of the formulas:

$$v^{(0)}[k,x] = \begin{cases} v^0[k,x] \text{ for } x \notin \mathcal{P}^1[k-1]; \\ \text{arbitrary } v \in \mathcal{Q}[k], \text{ otherwise,} \end{cases} \tag{3.5}$$

where $i_*$ in (3.4) is any index $i_* = i_*[k] \in \{1,\dots,n\}$ such that $|\xi_{i_*}[k,x]| > \pi^1_{i_*}[k-1]$;

$$v^{(1)}[k,x] = v^0[k,x], \quad \forall x \in \mathbb{R}^n, \tag{3.6}$$

where $i_* = i_*[k] \in \operatorname{Argmax}_{1\le i\le n}\Phi_i[k,x]$;

$$v^{(2)}[k,x] = \begin{cases} v^{(1)}[k,x] \text{ for } x \notin \mathcal{P}^1[k-1]; \\ q[k], \text{ otherwise.} \end{cases} \tag{3.7}$$

**Theorem 2.** *Under Assumption 1, let $P[k]$, $k = N,\dots,1$, be arbitrary nonsingular orientation matrices (i.e., arbitrariness is allowed when choosing $P[N]$ and matrices $P[k-1]$ in the 3rd line in (3.3)) and system (3.3) has a solution such that all parallelepipeds $\mathcal{P}^1[k]$, $k = N,\dots,1$, turn out to be nondegenerate. Then the tube $\mathcal{P}[\cdot]$ together with each of the control strategies $v^{(l)}[k,x]$, $l \in \{0,1,2\}$, from (3.5)–(3.7), which are determined by the couple of the tubes $\mathcal{P}[\cdot]$ and $\mathcal{P}^1[\cdot]$, give a particular solution to Problem 2.*

P r o o f (Sketch of the proof). It can be verified, using Lemma 1, that if $x[0] \notin \mathcal{P}[0]$, then the control strategy $v^{(0)}[\cdot,\cdot]$ ensures that $x[k] \notin \mathcal{P}[k]$ for all $k > 0$. The control strategies $v^{(1)}$ and $v^{(2)}$ are in fact special cases when $v^{(0)}$ is concretized. □

*Remark 1.* Theorem 2 depicts the parametric family of the tubes $\mathcal{P}[\cdot]$. Here the matrix function $P[\cdot]$ appears as a parameter. We note two following heuristic techniques to choose $P[k]$ for $k < N$ in the 3rd line in (3.3) (then only $P[N]$ is the parameter).

(a) Given $P[k]$, put $P[k-1] = P^1[k-1] = A[k]^{-1}P[k]$.

(b) Put $P[k-1]$ using arguments of local volume minimization of the type:

$$V \in \operatorname{Argmin}_{V\in\{P^1,P^2\}}\operatorname{vol}\boldsymbol{P}^+_V(\mathcal{P}^1\bigcup\mathcal{P}^2),$$

where $\mathcal{P}^k = \mathcal{P}(p^k, P^k, \pi^k)$, $k = 1,2$.

**Corollary 1.** *Theorem 2 is true, with an evident modification, if in the evasion problem the aim of $v$ is to ensure $x[N] \notin \mathcal{M}$ and $x[k] \notin \mathcal{M}$ only for $k \in \mathcal{K}$, where $\mathcal{K}$ is some subset of $\{1,\dots,N-1\}$ (in particular, we have $\mathcal{K}=\emptyset$ if we require only $x[N]\notin\mathcal{M}$). Namely, it is sufficient to replace the formulas for $\mathcal{P}[k-1]$ in the 3rd line of (3.3)) by $\mathcal{P}[k-1]=\mathcal{P}^1[k-1]$ for all $k-1$ such that $k-1\notin\mathcal{K}$. Then, for $\mathcal{K}=\emptyset$, the parallelepiped-valued tubes $\mathcal{P}[\cdot]$ turn out to coincide with the parallelotope-valued tubes $\mathcal{P}^+[\cdot]$ from [14, Theorem 1].*

Note that the solutions to Problem 2 described by Theorem 2 can be easily calculated by the explicit formulas.

So, to guarantee $x[k]\notin\mathcal{M}$ for all $k=1,\dots,N$ we can find several pairs of the tubes $\mathcal{P}[\cdot]$ and $\mathcal{P}^1[\cdot]$, which are determined by recurrence relations (3.3). Then for a given $x^0$ we can choose the most suitable tube $\mathcal{P}[\cdot]$, for example, similarly to [14, Sec. IV] (we need to fulfill the condition $x^0\notin\mathcal{P}[0]$ to meet the above claims for the trajectory) and apply any of the control strategies $v^{(l)}[k,x]$, $l\in\{0,1,2\}$, from (3.5)–(3.7), which are determined by the selected tube $\mathcal{P}[\cdot]$ and the corresponding tube $\mathcal{P}^1[\cdot]$. If we get $x^0\in\mathcal{P}[0]$ for all calculated tubes, then we, generally speaking, cannot guarantee $x[k]\notin\mathcal{M}$ for all $k=1,\dots,N$, but this can happen for some of realizations of $u[\cdot]$.

## 4. Example

Let us illustrate the presented constructions on the example of the same system as in [14, Sec. IV]. The system is obtained by Euler's approximations of a differential one considered on an interval $t\in[0,\theta]$:

$$A[k]\equiv I+h_N\cdot\begin{bmatrix}0 & 1\\ -8 & 0\end{bmatrix},\quad B[k]\equiv h_N\cdot(0,1)^\top,\quad \mathcal{R}[k]\equiv\mathcal{P}(0,I,1)\subset\mathbb{R}^1,$$
$$C[k]\equiv h_N\cdot(1,0)^\top,\quad \mathcal{Q}[k]\equiv\mathcal{P}(0,I,0.2)\subset\mathbb{R}^1,\quad \mathcal{M}=\mathcal{P}((-0.5,0)^\top,I,(0.5,0.5)^\top),$$
$$h_N=\theta/N,\quad \theta=2,\quad N=200.$$

Given $x^0$, let us denote by $\boldsymbol{A}_v^1$; $\boldsymbol{A}_v^2$; $\boldsymbol{A}_u^1$ the following three aims: to ensure $x[N]\notin\mathcal{M}$; $x[k]\notin\mathcal{M}$, $k=1,\dots,N$; $x[N]\in\mathcal{M}$ via control strategies $v$; $v$; $u$ respectively. To construct these controls $v$; $v$; $u$ we will use the solutions to Problem $2'$; to Problem 2; to the terminal target approach problem from [11, 12] through construction of several tubes $\mathcal{P}^{+,\alpha}[\cdot]$; $\mathcal{P}^{\beta}[\cdot]$; $\mathcal{P}^{-,\gamma}[\cdot]$ from parametric families of the tubes described in [14, Theorem 1] and also in Corollary 1; in Theorem 2; in [11, 12] respectively (see [14, the end of Sec. III] about using the families of the tubes for more details).

We consider 5 initial points

$$x^{0,1}=(-0.6,2)^\top,\quad x^{0,2}=(0,1.5)^\top,\quad x^{0,3}=(0.87,-1.5)^\top,$$
$$x^{0,4}=(0.88,-1.5)^\top,\quad x^{0,5}=(1,-1.5)^\top$$

and construct corresponding trajectories $x^{j,(i)}[\cdot]$, $j=1,\dots,5$, $i=1,2$, under controls $v^{(i)}$, $i=1,2$, from (3.6) and (3.7). We consider two Tests. In Test 1, we apply controls $v$, $u$ with aims $\boldsymbol{A}_v^1$, $\boldsymbol{A}_u^1$ respectively; in Test 2, we apply controls $v$, $u$ with aims $\boldsymbol{A}_v^2$, $\boldsymbol{A}_u^1$. Note that in Test 2 the aim $\boldsymbol{A}_u^1$ is not opposite to $\boldsymbol{A}_v^2$ (constructing $u$ with the aim $\boldsymbol{A}_u^2$ opposite to $\boldsymbol{A}_v^2$ is out of the scope of this paper). It is possible to use several formulas to construct $u$ basing on tubes $\mathcal{P}^{-,\gamma}[\cdot]$ (see, for example, [12]). Here we have applied the formulas which are similar to [13, Formula (13)]. We do not supply the trajectories $x^{j,(i)}[\cdot]$ by numbers of the Tests to simplify the notation.

The results of computer simulations are visualized in Fig. 1, where cross-sections $\mathcal{P}^{+,\alpha}[0]$, $\alpha=1,\dots,4$, and $\mathcal{P}^{\beta}[0]$, $\beta=1,\dots,4$, of several tubes calculated for solving Problem $2'$ and

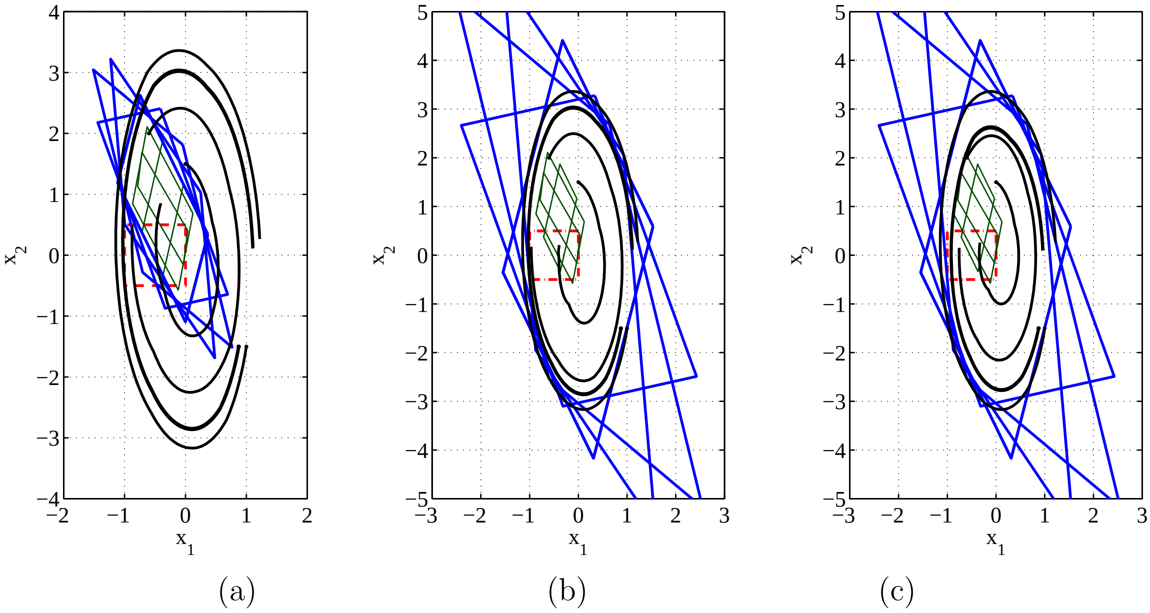



Figure 1. Used constructions and results of evasion from $\mathcal{M}$ (dashed red lines) in Example: several cross-sections $\mathcal{P}^{+,\alpha}[0]$ (left figure, blue thick lines), $\mathcal{P}^{\beta}[0]$ (two right figures, blue lines), $\mathcal{P}^{-,\gamma}[0]$ (green thin lines), and the controlled trajectories under suitable control strategies $v^i$, $i \in \{1,2\}$, and $u$. (a) Test 1: using $v^{(1)}$ based on $\mathcal{P}^{+,\alpha}[\cdot]$. (b) Test 2: using $v^{(1)}$ based on $\mathcal{P}^{\beta}[\cdot]$. (c) Test 2: using $v^{(2)}$ based on $\mathcal{P}^{\beta}[\cdot]$

Problem 2 respectively are shown by thick lines, cross-sections $\mathcal{P}^{-,\gamma}[0]$, $\gamma = 1,\ldots,3$, by thin lines; the target set $\mathcal{M}$ is presented by dashed lines. The tubes $\mathcal{P}^{+,\alpha}[\cdot]$ and $\mathcal{P}^{-,\gamma}[\cdot]$ are the same as in [14, Sec. IV]; $\mathcal{P}^{\beta}[\cdot]$ are constructed as described in Theorem 2 and Remark 1(b) under the same orientation matrices $P[N]$ at the final instant as for $\mathcal{P}^{+,\alpha}[\cdot]$.

The point $x^{0,1} \in \mathcal{P}^{-,\gamma_*}[0]$ for some $\gamma_*$ and we obtained $x[N] \in \mathcal{M}$ (aim $\boldsymbol{A}_u^1$ is achieved) for all trajectories started at $x^{0,1}$ in both Test 1 and Test 2 as it is theoretically guaranteed similarly to [12, Theorem 3.1].

Each of the points $x^{0,j}$, $j = 2,\ldots,5$, is outside at least one of $\mathcal{P}^{+,\alpha}[0]$ (see Fig. 1(a)). In Test 1, we obtained $x[N] \notin \mathcal{M}$ (aim $\boldsymbol{A}_v^1$ is achieved) for all trajectories started at these $x^{0,j}$ as it was theoretically guaranteed by [14, Theorem 1] and also by Corollary 1. Note that in Test 1 $x^{j,(i)}[\cdot]$, $j = 2,3$, $i = 1,2$, hit $\mathcal{M}$ at some instants $k < N$. The reason is that in Test 1 we used controls $v$ designed for solving Problem $2'$ but not Problem 2.

It is also curious that in Test 2 we obtained $x^{2,(i)}[N] \in \mathcal{M}$, $i = 1,2$, in opposite to Test 1. The reason here is that we have used controls $v^{(i)}$ basing on the tubes $\mathcal{P}^{\beta}[\cdot]$ without the guarantee to achieve the aim $\boldsymbol{A}_v^2$ because we have $x^{0,2} \in \bigcap_{\beta=1}^{4} \mathcal{P}^{\beta}[0]$.

The point $x^{0,5} \notin \mathcal{P}^{\beta_*}[0]$ for some $\beta_*$ (see Fig. 1(b), Fig. 1(c)). In Test 2, we obtained $x[k] \notin \mathcal{M}$, $k = 1,\ldots,N$, (aim $\boldsymbol{A}_v^2$ is achieved) for the trajectories $x^{5,(i)}[\cdot]$, $i = 1,2$, started at $x^{0,5}$ as it is theoretically guaranteed by Theorem 2.

For the trajectories started at $x^{0,j}$, $j = 2,3,4$, we have no any guarantees about hitting $\mathcal{M}$ in Test 2 because we have $x^{0,j} \in \bigcap_{\beta=1}^{4} \mathcal{P}^{\beta}[0]$ for these $j$. And we obtained, in particular, the following results in Test 2. For very close initial points $x^{0,3}$ and $x^{0,4}$ we obtained that $x^{3,(1)}[k] \in \mathcal{M}$ and $x^{3,(2)}[k] \in \mathcal{M}$ for 1 and 13 instants $k$ respectively; $x^{4,(1)}[k] \notin \mathcal{M}$ for all $k$, and $x^{4,(2)}[k] \in \mathcal{M}$ for 14 instants $k$. Thus, in this example, $v^{(1)}$ turned out to be more successful than $v^{(2)}$ for the initial points without guarantees.

## 5. Conclusion

We deal with linear discrete-time systems under two conflicting controls and given target sets. Two subproblems arrise, namely the approach problem and the evasion one. Formerly we elaborated the polyhedral control synthesis for the terminal approach problem and for the terminal evasion problem using polyhedral (parallelotope-valued) tubes. In this paper, the enhanced evasion problem is considered to avoid the trajectory hitting the given target set not only at the given final time, but also at intermediate instants. The common solution scheme is outlined. Then the solution technique is elaborated based on polyhedral (parallelepiped-valued) tubes. The recurrence relations with explicit formulas are presented for the couple of such tubes, the finding of which is much less time-consuming than the construction of the exact solutions. Control strategies, which can be calculated also by explicit formulas on the base of these tubes, are constructed. The illustrative example demonstrating the theoretical results is presented.

### REFERENCES

1. Ananyev B. I. An application of motion correction methods to the alignment problem in navigation. *Ural Math. J.,* 2016. Vol. 2, No. 2. P. 16–26. DOI: 10.15826/umj.2016.2.002
2. Bertsekas D. P., Rhodes I. B. On the minimax reachability of target sets and target tubes. *Automatica*, 1971. Vol. 7, No. 2. P. 233–247. DOI: 10.1016/0005-1098(71)90066-5
3. Botkin N., Martynov K., Turova V., Diepolder J. Generation of dangerous disturbances for flight systems. *Dyn. Games and Appl.*, 2019. Vol. 9. P. 628–651. DOI: 10.1007/s13235-018-0259-5
4. Chernousko F. L. *Estimation for Dynamic Systems.* Boca Raton: CRC Press, 1994. 304 p.
5. Filippova T. F. Control and estimation for a class of impulsive dynamical systems. *Ural Math. J.,* 2019. Vol. 5, No. 2. P. 21–30. DOI: 10.15826/umj.2019.2.003
6. Filippova T. F., Matviychuk O. G. Optimality principles for solving nonlinear control problems under uncertainty. In: *CHAOS 2021: Proc. 14th Chaotic Modeling and Simulation Int. Conf., Skiadas C.H., Dimotikalis Y. (eds).* Springer Pros. Complexity, 2022. P. 143–154. DOI: 10.1007/978-3-030-96964-6_12
7. Gusev M. I., Zykov I. V. External estimates and comparison principle for trajectory tubes of nonlinear control systems. *AIP Conf. Proc.,* 2019. Vol. 2164. Art. no. 060008. 10 p. DOI: 10.1063/1.5130810
8. Kamneva L. V., Patsko V. S. Construction of the solvability set in differential games with simple motion and nonconvex terminal set. *Proc. Steklov Inst. Math.,* 2018. Vol. 301, Suppl. 1. P. S57–S71. DOI: 10.1134/S008154381805005X
9. Kostousova E. K. State estimation for dynamic systems via parallelotopes: Optimization and parallel computations. *Optim. Methods Softw.,* 1998. Vol. 9, No. 4. P. 269–306. DOI: 10.1080/10556789808805696
10. Kostousova E. K. On polyhedral estimates in problems of the synthesis of control strategies in linear multistep systems. In: *Algoritmy i programmnye sredstva parallel'nykh vychislenii* [Algorithms and Software for Parallel Computing]. Ekaterinburg: IMM UrO RAN, 2006. Vol. 9. P. 84–105. (in Russian)
11. Kostousova E. K. On the polyhedral method of solving problems of control strategy synthesis. *Proc. Steklov Inst. Math.,* 2016. Vol. 292, Suppl. 1. P. S140–S155. DOI: 10.1134/S0081543816020127
12. Kostousova E. K. On polyhedral control synthesis for dynamical discrete-time systems under uncertainties and state constraints. *Discrete Contin. Dyn. Syst.*, 2018. Vol. 38, No. 12. P. 6149–6162. DOI: 10.3934/dcds.2018153
13. Kostousova E. K. On a polyhedral method of solving an evasion problem for linear dynamical systems. *AIP Conf. Proc.,* 2019. Vol. 2164. Art. no. 060009. 10 p. DOI: 10.1063/1.5130811
14. Kostousova E. K. On a polyhedral method for solving an evasion problem for linear discrete-time systems. In: *Proc. of 2020 15th Int. Conf. on Stability and Oscillations of Nonlinear Control Systems (Pyatnitskiy's Conference) (STAB).* IEEE Xplore, 2020. P. 1–4. DOI: 10.1109/STAB49150.2020.9140471
15. Kostousova E. K. External polyhedral estimates of reachable sets of discrete-time systems with integral bounds on additive terms. *Math. Control Relat. Fields*, 2021. Vol. 11, No. 3. P. 625-641. DOI: 10.3934/mcrf.2021015
16. Krasovskii N. N., Subbotin A. I. *Game-Theoretical Control Problems.* New York: Springer, 1988. 517 p.

17. Kurzhanski A. B., Daryin A. N. *Dynamic Programming for Impulse Feedback and Fast Controls: The Linear Systems Case.* Lect. Notes Control Inf. Sci., vol. 468. London: Springer, 2020. 275 p. DOI: 10.1007/978-1-4471-7437-0
18. Kurzhanski A. B., Vályi I. *Ellipsoidal Calculus for Estimation and Control.* Boston: Birkhäuser, 1997. 321 p.
19. Kurzhanski A. B., Varaiya P. *Dynamics and Control of Trajectory Tubes. Theory and Computation.* Systems Control Found. Appl., vol. 85. Basel: Birkhäuser, 2014. 445 p. DOI: 10.1007/978-3-319-10277-1
20. Kurzhanskiy A. A., Varaiya P. Theory and computational techniques for analysis of discrete-time control systems with disturbances. *Optim. Methods Softw.,* 2011. Vol. 26, No. 4–5. P. 719–746. DOI: 10.1080/10556788.2010.548867
21. Martynov K., Botkin N., Turova V., Diepolder J. Real-time control of aircraft take-off in windshear. Part I: Aircraft model and control schemes. In: *Proc. of 2017 25th Mediterranean Conf. on Control and Automation (MED 2017).* IEEE Xplore, 2017. P. 277–284. DOI: 10.1109/MED.2017.7984131
22. Martynov K., Botkin N., Turova V., Diepolder J. Quick construction of dangerous disturbances in conflict control problems. In: *Advances in Dynamic Games. Ramsey D.M., Renault J. (eds).* Ann. Internat. Soc. Dynam. Games, vol. 17. Cham: Birkhäuser, 2020. P. 3–24. DOI: 10.1007/978-3-030-56534-3_1
23. Schneider R. G. *Convex Bodies: The Brunn-Minkowski Theory.* Cambridge: Cambridge Univ. Press, 1993. 490 p.
24. Ushakov V. N., Uspenskii A. A., Matviychuk A. R., Malev A. G. Stability defect of sets in game problems of approaching. *IFAC-PapersOnline,* 2012. Vol. 45, No. 25. P. 89–93. DOI: 10.3182/20120913-4-IT-4027.00063

URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 64–75

DOI: 10.15826/umj.2022.1.007

# ON DOUBLE SIGNAL NUMBER OF A GRAPH

**X. Lenin Xaviour**

Department of Mathematics, Nesamony Memorial Christian College,
Marthandam – 629165, Tamil Nadu, India
leninxaviour93@gmail.com

**S. Ancy Mary**

Department of Mathematics,
St. John's College of Arts and Science, Ammandivilai,
Affliated to Manonmaniam Sundaranar University,
Abishekapatti, Tirunelveli – 627012, Tamil Nadu, India
ancymary369@gmail.com

**Abstract:** A set $S$ of vertices in a connected graph $G=(V,E)$ is called a *signal set* if every vertex not in $S$ lies on a signal path between two vertices from $S$. A set $S$ is called a *double signal set* of $G$ if $S$ if for each pair of vertices $x,y \in G$ there exist $u,v \in S$ such that $x,y \in L[u,v]$. The double signal number $\operatorname{dsn}(G)$ of $G$ is the minimum cardinality of a double signal set. Any double signal set of cardinality $\operatorname{dsn}(G)$ is called dsn-set of $G$. In this paper we introduce and initiate some properties on double signal number of a graph. We have also given relation between geodetic number, signal number and double signal number for some classes of graphs.

**Keywords:** Signal set, Geodetic set, Double signal set, Double signal number.

## 1. Introduction

By a graph $G=(V,E)$ we mean a finite, connected, undirected graph with neither loops nor multiple edges. The *order* $|V|$ and *size* $|E|$ of $G$ are denoted by $p$ and $q$ respectively. For graph theoretic terminology we refer to [1]. The *open neighborhood* of any vertex $v$ in $G$ is $N(v)=\{x : xv \in E(G)\}$ and *closed neighborhood* of a vertex $v$ in $G$ is $N[v]=N(v)\cup\{v\}$. The *degree* of a vertex in the graph $G$ is denoted by $\deg(v)$ and the maximum degree (minimum degree) in the graph $G$ is denoted by $\triangle(G)(\delta(G))$. For a set $S\subseteq V(G)$ the open (closed) neighborhood $N(S)(N[S])$ in $G$ is defined as

$$N(S)=\bigcup_{v\in S} N(v)(N[S]=\bigcup_{v\in S} N[v]).$$

A graph $G$ is said to be *connected* if any two vertices in $G$ are joined by a path. A maximal connected subgraph of $G$ is called a *component* of $G$. A graph is said to be *disconnected* if it has at least two components. A *cut-vertex* of a connected graph is a vertex whose removal results a disconnected graph. A graph $G$ is said to be *regular* if every vertex of $G$ has equal degree.

If $G$ is a connected graph the *distance* $d(x,y)$ is the length of a shortest $x$-$y$ path in $G$. The *diameter* is defined by $\operatorname{diam}(G)=\max_{x,y\in V(G)} d(x,y)$. Two vertices $u$ and $v$ are said to be *antipodal vertices* if $d(u,v)=\operatorname{diam}(G)$. If $e=\{u,v\}$ is an edge of a graph $G$ with $\deg(u)=1$ and $\deg(v)>1$, then we call $e$ a pendant edge, $u$ a pendant vertex and $v$ a support vertex. A vertex $v$ of $G$ is said to be an *extreme vertex* if the subgraph induced by its neighborhood is complete. The set of all extreme vertices is denoted by $\operatorname{Ext}(G)$. An acyclic connected graph is called a *tree*. An $x$-$y$ path of length $d(x,y)$ is called *geodesic*.

A set $S \subseteq V(G)$ is called a *geodetic set* of $G$, if every vertex in $G$ lies on a geodesic joining a pair of vertices of $S$. The *geodetic number* of $G$, denoted by $g(G)$, is the minimum cardinality of a geodetic set of $G$. The geodetic number of a disconnected graph is the sum of the geodetic number of its components. Any geodetic set of cardinality $g(G)$ is called $g$-set of $G$.

A set $S$ of vertices in $G$ is called a *double geodetic set* of $G$ if for each pair of vertices of $G$ lie on any geodesic joining pair of vertices from $S$. The *double geodetic number* $\operatorname{dg}(G)$ is the minimum cardinality of a double geodetic set. Any geodetic set of cardinality $\operatorname{dg}(G)$ is called dg-set of $G$. The double geodetic number of a graph was introduced and studied in [7]. Various concepts inspired by geodetic sets are introduced in [1, 3, 4].

On a various study on the distance in graphs, we refer to [1]. In the meantime, Chartrand et al. introduced a new type of distance parameter called the detour distance in graphs. Once a new type of distance between two vertices was introduced by Chartrand et al., various new distance parameters such as Supreme distance, D-distance and many more, were introduced by different researchers. In continuation, Kathiresan et al. introduced a distance parameter, called the signal distance in graphs [5]. The signal distance $d_{SD}(u, v)$ between a pair of vertices $u$ and $v$ is defined by

$$d_{SD}(u,v) = \min\Big\{d(u,v) + \sum_{w\in V(G)} (\deg w - 2) + (\deg u - 1) + (\deg v - 1)\Big\},$$

where $S$ is a path connecting $u$ and $v$, $d(u, v)$ is the length of the path $S$ and the sum $\sum_{w\in V(G)}$, where sum runs over all the internal vertices between $u$ and $v$ in the path $S$. The $u$-$v$ signal path of length $d_{SD}(u, v)$ is also called *geosig*. A vertex $v$ is said to lie on a geosig $P$ if $v$ is an internal vertex of $P$. The signal interval $L[x, y]$ consists of $x, y$ and all vertices lying on some $x$-$y$ geosig of $G$ and for a non empty set $S \subseteq V(G)$, $L[S] = \bigcup_{x,y\in S} L[x, y]$.

A set $S \subseteq V(G)$ in a connected graph is a *signal set* of $G$ if $L[S] = V(G)$. The *signal number* $\operatorname{sn}(G)$ is the minimum cardinality of a signal set of $G$. A signal set of cardinality $\operatorname{sn}(G)$ is called a sn-set of $G$. The signal number of a graph was introduced in [8] and further studied in [2, 5]. The concept of signal number can be applied in the fields of electrical engineering and irrigation systems. It was shown that the determining the signal number of a graph is an $NP$-hard problem. Let $2^V$ denote the set of all the subsets of $V$. The mapping $L\colon V \times V \to 2^V$ defined by

$$L[x, y] = \{z \in V : z \text{ lies on a } x\text{-}y \text{ geosig in } G\}$$

is the signal function of $G$. One of the basic properties of $L$ is that $x, y \in L[x, y]$ for any pair $x, y \in V$. Hence the signal function captures every pair of vertices and so the problem of double signal sets is trivially well-defined while it is clear that this fails in many graphs already for triplets (for example, complete graphs). This is the motivation for introducing and studying double signal sets.

The concepts of distance in graphs is a major component in graph theory with its convexity concepts having numerous applications in real life problems. There are several interesting applications of these concepts to facility location in real life situations, routing of transport problems and communication network designs. As the path involved in this discussion of this paper are geosig, no intervention by hackers or enemies is possible to the respective facilities provided. Further, as signal paths are secured and longer than geodesic paths, it is advantageous to more customers in getting the service with protection.

The following theorems will be used in the subsequent sections.

**Theorem 1** [2]. *For any connected graph $G$, the set of all end vertices is a subset of every signal set of $G$.*

**Theorem 2** [3]. *Each extreme vertex of a connected graph $G$ belongs to every geodetic set of $G$.*

**Theorem 3** [7]. *Each extreme vertex of a connected graph $G$ belongs to every double geodetic set of $G$.*

The signal number of some standard classes of graph can be easily found and are given below:

- Path $P_p$ of $p \geq 2$ vertices, $\operatorname{sn}(P_p) = 2$.
- Cycle $C_p$ of $p \geq 3$ vertices, $\operatorname{sn}(C_p) = \begin{cases} 2, & \text{if} \quad p \text{ is even}, \\ 3, & \text{if} \quad p \text{ is odd}. \end{cases}$
- Complete graph $K_p$ of $p \geq 2$ vertices, $\operatorname{sn}(K_p) = p$.
- Peterson graph $G$, $\operatorname{sn}(G) = 4$.
- Star graph $K_{1,p-1}$ of $p \geq 2$ vertices, $\operatorname{sn}(K_{1,p-1}) = p - 1$.
- Complete bipartite graph $K_{m,n}(2 \leq m \leq n)$, $\operatorname{sn}(C_p) = \begin{cases} m, & \text{if} \quad m \leq 3, \\ 4, & \text{otherwise}. \end{cases}$

## 2. Double signal number of a graph

**Definition 1.** *Let $G$ be a connected graph with at least two vertices. A set $S$ of vertices of $G$ is called a double signal set of $G$ if for each pair of vertices $x, y \in G$ there exist $u, v \in S$ such that $x, y \in L[u, v]$. The double signal number* $\operatorname{dsn}(G)$ *of $G$ is the minimum cardinality of a double signal set. Any double signal set of cardinality* $\operatorname{dsn}(G)$ *is called* dsn*-set of $G$.*

*Example 1.* For the graph $G$ in Fig. 1, it is clear that no 2-element subset of $G$ is a signal set of $G$. Now $S = \{v_1, v_4, v_5\}$ is a signal set of $G$ and so $\operatorname{sn}(G) = 3$. Clearly the pair of vertices $v_3$, $v_6$ lies only the $v_3$-$v_6$ geosig. Similarly, the vertices $v_6, v_8$ lies only the $v_6$-$v_8$ geosig. Also the vertices $v_2$, $v_6$ and $v_6$, $v_7$ lies only the $v_2$-$v_6$ and $v_6$, $v_7$ geosig, respectively. Therefore that $S$ is not a double signal set of $G$. Since $v_2, v_3, v_7, v_8$ be an internal vertices of $v_1$-$v_4$ geosig path, we need at least 6

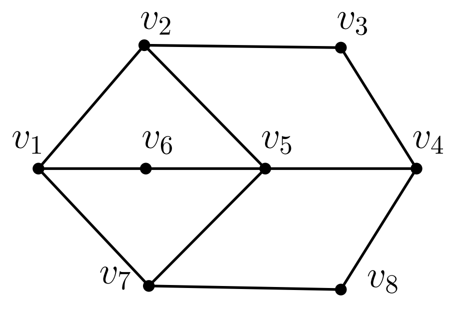



Figure 1. Graph G.

vertices to form a double signal set of $G$ and so $\operatorname{dsn}(G) \geq 6$. Now, since $S_1 = \{v_1, v_3, v_4, v_5, v_6, v_8\}$ is a double signal set, it follows that $\operatorname{dsn}(G) = 6$.

*Remark 1.* For the graph $G$ in Fig. 1, $S = \{v_1, v_4\}$ is the unique $g$-set and dg-set of $G$ and so $g(G) = \operatorname{dg}(G) = 2$. Thus the double signal number is different from geodetic number and double geodetic number.

The following theorem directly follows by the definition of signal number and double signal number.

**Theorem 4.** *For any connected graph $G$ of order $p$, $2 \leq \operatorname{sn}(G) \leq \operatorname{dsn}(G) \leq p$.*

*Remark 2.* The bounds in Theorem 4 are sharp. For the complete graph $K_p (p \geq 2)$, $\operatorname{dsn}(K_p) = p$. The set of the two end vertices of path graph $P_p (p \geq 2)$ forms a unique double signal set and so $\operatorname{dsn}(P_p) = 2$. Thus the nontrivial complete graph $K_p$ has the largest possible double signal number and the nontrivial path graph $P_p$ has the smallest double signal number. Also Example 2 shows that the bounds in Theorem 4 is sharp.

**Theorem 5.** *Each extreme vertex of a connected graph $G$ belongs to every signal set of $G$.*

P r o o f. Let $u$ be an extreme vertex of $G$ and let $S$ be a signal set of $G$. If $u$ is an end-vertex, then by Theorem 1 $u \in S$. Suppose $u \notin S$ be non end-vertex. Then $u$ is an internal vertex of an $x$-$y$ geosig path, say $P$, for some $x, y \in S$. Since $\deg(u) \geq 2$, $u$ has at least two neighbours in $P$ which are not adjacent and so that $u$ is not an extreme vertex, which is a contradiction. Hence $u \in S$. □

The following result is an easy consequences of Theorem 5.

**Result 1.** For the complete graph $K_p (p \geq 2)$, $\operatorname{dsn}(K_p) = p$.

To aid in our discussion throughout this paper, we define a definition as follows.

**Definition 2.** *Let $G$ be a connected graph of order $p \geq 2$. A vertex $v \in G$ is said to be a weak extreme vertex, if there exists a vertex $u$ in $G$ such that $v$ is either an initial vertex or a terminal vertex of any signal interval containing both $u$ and $v$.*

**Theorem 6.** *Every double signal set of a connected graph $G$ contains all the weak extreme vertices of $G$. In particular, if the set $S$ of all weak extreme vertices is a double signal set, then $S$ is the unique* dsn*-set of $G$.*

P r o o f. Let $S$ be a double signal set of $G$ and let $x$ be a weak extreme vertex of $G$. Suppose $x \notin S$. Let $y$ be any vertex in $G$ such that $x \neq y$. Since $S$ is a double signal set of $G$, we have for some $u, v \in S$, that $x, y$ lie on an $u$-$v$ geosig path. Also, that $x$ is a weak extreme vertex of $G$ shows either $x = u$ or $x = v$. It follows that $x \in S$, which is a contradiction. □

**Corollary 1.** *Each extreme vertex of a connected graph $G$ belongs to every double signal set of $G$.*

P r o o f. Since every extreme vertex of $G$ is weak extreme, the result follows from Theorem 6. □

*Example 2.* For the graph $G$ in Fig. 2, the set $S = \{v_1, v_5, v_7\}$ of extreme vertices form unique minimum signal set of $G$ and so $\operatorname{sn}(G) = 3$. Since the pair of vertices $v_3, v_6$ does not lie on any geosig of any pair of vertices from $S$, that $S$ is not a double signal set of $G$. Also the vertex $v_6$ is the only non-extreme vertex which became weak extreme. It is clear that the set $S_1 = S \cup \{v_6\}$ of all weak extreme vertices form a double signal set of $G$ and so by Theorem 6 $\operatorname{dsn}(G) = 4$.

**Result 2.** For any cycle $C_p (p \geq 3)$,

$$\operatorname{dsn}(C_p) = \begin{cases} 2, & \text{if } p \text{ is even,} \\ p, & \text{if } p \text{ is odd.} \end{cases}$$

**Result 3.** For any wheel $W_p = K_1 + C_{p-1} (p \geq 3)$, $\mathrm{dsn}\,(W_p) = p$.
**Result 4.** For the complete bipartite graph $K_{m,n} (m, n \geq 2)$, $\mathrm{dsn}\,(K_{m,n}) = \min\{m, n\}$.
**Result 5.** For any fan $F_p = K_1 + P_{p-1} (p \geq 3)$,

$$\mathrm{dsn}\,(F_p) = \begin{cases} p-2, & \text{if} \quad p \text{ is even,} \\ p, & \text{if} \quad p \text{ is odd.} \end{cases}$$

**Result 6.** For the star graph $K_{1,p-1}$, $\mathrm{dsn}\,(K_{1,p-1}) = p - 1$.

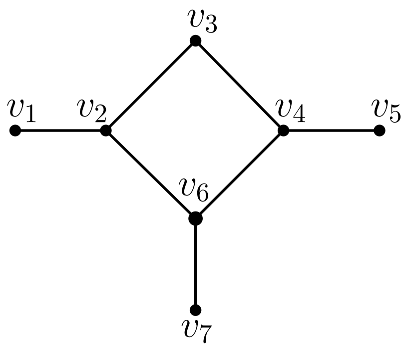



Figure 2. Graph G.

**Theorem 7.** *Let $G$ be a connected graph with cut vertices and let $S$ be a double signal set of $G$. If $v$ is a cut vertex of $G$, then every component of $G - v$ contains at least one element of $S$.*

P r o o f. Let $v$ be cut vertex of $G$ and $S$ be a double signal set of $G$. Suppose to the contrary, there exists a component, say $H$ of $G - v$ such that $H$ contains no vertex of $S$. By Theorem 6, $S$ contains all the weak extreme vertices of $G$ and hence, by assumption $H$ does not contain any weak extreme vertex of $G$. Let $u \in V(H)$. Since, $S$ is a double signal set of $G$, there exist vertices $x, y \in S$ such that $u, v \in L[x, y] \subseteq L[S]$. Let the $x$-$y$ geosig path in $G$ be $P : x = u_0, u_1, .., u, ..., u_l = y$ such that $u \neq x, y$. Since, $v$ is a cut vertex, the $x$-$u$ subpath of $P$ and the $u$-$y$ subpath of $P$ both contain $v$, it implies that $P$ is not a geosig path, which is a contradiction. Hence, every component of $G - v$ contains an element of $S$. □

**Theorem 8.** *No cut-vertex of a connected graph $G$ belongs to any* dsn*-set of $G$.*

P r o o f. Suppose $S$ be a dsn-set of a connected graph $G$ that contains a cut-vertex $v$. Let $G_1, G_2, ..., G_n (n \geq 2)$ be the components of $G - v$. Let $S_1 = S - \{v\}$. We show that $S_1$ is a double signal set of $G$. Let $x, y \in V(G)$. Since $S$ is a double signal set, then $x, y$ lies on a geosig $P$ joining a pair of vertices $a, b \in S$. If $v \notin \{a, b\}$, then $\{a, b\} \subseteq S_1$ and so that $S_1$ is a double signal set of $G$, which contradicts the minimality of $S$. Therefore, assume that $v \in \{a, b\}$ such that $v = b$ and $a \in G_1$. Since $S_1 \subseteq S$, that $a \in S_1$. By Theorem 7 we can fix a vertex $u \in G_k$ for $k \neq 1$ such that $u \in S$. Since $u \neq v$, that $u \in S$. Now, since $v$ is a cut vertex of $G$, the signal interval of the path between $a$ and $v$ contained in the signal interval of the path between $a$ and $u$. This shows that $x, y$ lies on the geosig between $a, u \in S_1$. Therefore, that $S_1$ is a double signal set of $G$, which again contradicts the minimality of $S$. Hence no cut-vertex of $G$ belongs to any dsn-set of $G$. □

**Definition 3.** *Let $u$ be a vertex in $G$. A vertex $v$ in $G$ is said to be an $u -$ signal vertex if for any vertex $w \neq u, v$ with $d_{SD}(u, v) < d_{SD}(u, w)$, $w$ lies on an $u$-$v$ signal path.*

**Theorem 9.** *For any connected graph $G$, $\operatorname{sn}(G) = 2$ if and only if there exist vertices $u, v$ such that $v$ is an $u$-signal vertex of $G$.*

P r o o f. Let $\operatorname{sn}(G) = 2$ and $S = \{u, v\}$ be a sn-set of $G$. Then every vertex $w$ in $G$ lies on this $u$-$v$ signal path and so that $d_{SD}(u, v)$ is minimum. Thus, $d_{SD}(u, v) < d_{SD}(u, w)$ for every $w \neq u, v$. Hence $v$ is an $u$-signal vertex of $G$. The converse part is obvious. □

**Theorem 10.** *For a nontrivial connected graph $G$, $\operatorname{dsn}(G) = 2$ if and only if $\operatorname{sn}(G) = 2$.*

P r o o f. Let $S = \{u, v\}$ be a sn-set of $G$ such that $\operatorname{sn}(G) = 2$. Then every pair of vertices of $G$ lies on a $u$-$v$ geosig and so that $S$ itself forms a double signal set. Hence, $\operatorname{dsn}(G) = 2$. Converse part follows from Theorem 4. □

The following result follows from Theorem 6 and Theorem 8.

**Result 7.** If $T$ is a tree with $l$ end vertices, then $\operatorname{dsn}(T) = l$. In fact, the set of all end vertices of $T$ is the unique dsn-set of $T$.

**Lemma 1.** *Let $G$ be a connected graph of order $p \geq 2$. If there exists a vertex $v \in G$ such that*

$$\deg(v) > \sum_{w \in G} \deg(w) + l(P),$$

*where $l(P)$ is the length of a geosig path $P$ between any two antipodal vertices and the sum $\sum_{w \in G} \deg(w)$ runs over all the internal vertices between the antipodal vertices in $P$, then $\operatorname{dsn}(G) = p$.*

For every connected graph $G$, it is clear that $\operatorname{rad}(G) \leq \operatorname{diam}(G) \leq 2 \operatorname{rad}(G)$ [6]. Ostrand showed that any two positive integers $a$ and $b$ with $a \leq b \leq 2a$ are realizable as the radius and diameter, respectively. This theorem can be extended so that the double geodetic number can be prescribed as well.

**Theorem 11.** *For positive integers $r, d$ and $a \geq 2$ with $r \leq d \leq 2r$, there exists a connected graph $G$ with $\operatorname{rad}(G) = r$, $\operatorname{diam}(G) = d$ and $\operatorname{dsn}(G) = a$.*

P r o o f. If $r = 1$, then consider $G = K_a$ or $G = K_{1,a}$ according to whether $d = 1$ or $d = 2$, respectively. If $r = d \geq 2$ and $a = 2$, then we take $G = C_{2r}$.

Now assume that $r = d \geq 2$ and $a \geq 3$. Let $C_{2r} : u_1, u_2, ..., u_r, u_{r+1}, ..., u_{2r}, u_1$ be a cycle of order $2r$. Add $p - 1$ pendant edges $v_1u_1, v_2u_1, ..., v_{a-1}u_1$ to obtained the graph $G$. Clearly $\operatorname{rad}(G) = \operatorname{diam}(G) = r$. The graph $G$ has $a - 1$ extreme vertices, that is, $S = \{v_1, v_2, ..., v_{a-1}\}$. By Corollary 1, each double signal set of $G$ must contain $S$ and that $L[S] \neq V(G)$. Hence, $\operatorname{dsn}(G) \geq a - 1$. On the other hand, we have $L[S \cup \{u_{r+1}\}] = V(G)$ and every pair of vertices of $G$ lies on a geosig of some pair of vertices from $S \cup \{u_{r+1}\}$, implying that $\operatorname{dsn}(G) = a$.

Finally assume $2 \leq r < d$. First assume $a \geq 3$. Let $G$ be the graph obtained from the disjoint union of a cycle $C_{2r} : u_1, u_2, ..., u_r, u_{r+1}, ..., u_{2r}, u_1$ of order $2r$ and a path $P_{d-r+1} : v_0, v_1, ..., v_{d-r}$ of order $d - r + 1$ by identifying $u_1$ and $v_0$. Add new pendant edges $u_rw_1, u_rw_2, u_rw_3, ..., u_rw_{a-3}$. Then $G$ has radius $r$ and diameter $d$. This graph $G$ is shown in Fig. 3.

Now, we prove the set $\{w_1, w_2, ....w_{a-3}, u_{r+1}, u_{2r}, v_{d-r}\}$ forms a double signal set of $G$. By Corollary 1, $w_1, w_2, ..., w_{a-3}, v_{d-r} \in S$, where $S$ is a double signal set of $G$. Further, as vertices $u_{r+1}, u_{r+2}, ..., u_{2r}$ in $V(G) - \{w_1, w_2, ..., w_{a-3}, v_{d-r}\}$ cannot covered by using the vertices $w_1, w_2, ..., w_{a-2}, v_{d-r}$, $|S| \geq a - 3 + 1 = a - 2$. Now it is clear that $u_{r+1}$ is either an internal

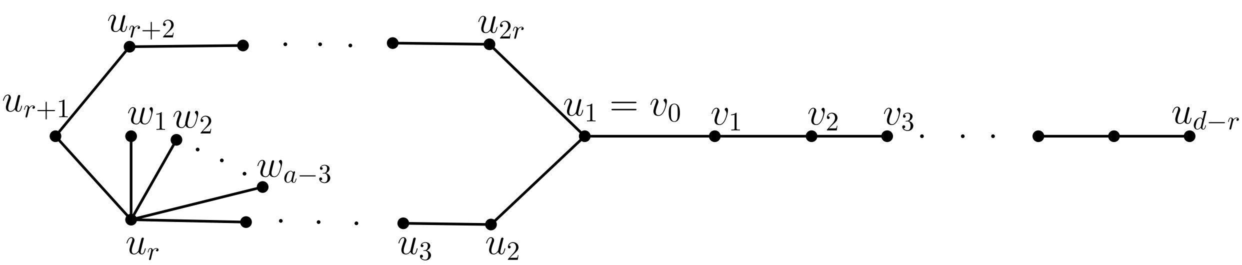



Figure 3. Graph G.

vertex or a terminal vertex of any signal path containing the pair of vertices $u_{r+1}, v_i$. Similarly, $u_{2r}$ is either an internal vertex or a terminal vertex of any signal path containing the pair of vertices $u_{2r}, w_i$. Thus $u_{r+1}, u_2r$ are weak extreme vertices. Therefore by Theorem 6, $u_{r+1}, u_{2r} \in S$ and so $\operatorname{dsn}(G) \geq a$. Since $\{w_1, w_2, ....w_{a-3}, u_{r+1}, u_{2r}, v_{d-r}\}$ forms a double signal set of $G$, that $\operatorname{dsn}(G) = a$. For the case $a = 2$, we remove the pendant edges $u_r w_1, u_r w_2, u_r w_3, ..., u_r w_{a-3}$ of $G$ in Fig. 3. Clearly $G$ has radius $r$ and diameter $d$. Also $\{u_{r+1}, v_{d-r}\}$ is the unique double signal set of $G$ and so by Theorem 4 we conclude that $\operatorname{dsn}(G) = 2$. This complete the proof. □

**Theorem 12.** *For every pair $a, p$ of integers with $2 \leq a \leq p$, there exists a connected graph $G$ of order $p$ such that* $\operatorname{dsn}(G) = a$.

P r o o f. If $2 \leq a = p$, we take $G = K_p$. For $2 \leq a < p$, we consider a tree graph $G$ of order $p$ with $a$ end-vertices. □

## 3. The double signal number and double geodetic number of a graph

In this section, we consider the realization result connecting the double signal number and double geodetic number of connected graphs. For this, first we focus the signal number and geodetic number. Because, for the graph $G$ in Fig. 2, $g(G) = 3$ and $\operatorname{sn}(G) = 3$ so that $\operatorname{sn}(G) = g(G)$. Similarly, for the graph $G$ in Fig. 1, $g(G) = 2$ and $\operatorname{sn}(G) = 3$ so that $\operatorname{sn}(G) > g(G)$ and for the graph $G$ in Fig. 4, $g(G) = 3$ and $\operatorname{sn}(G) = 2$ so that $\operatorname{sn}(G) < g(G)$.

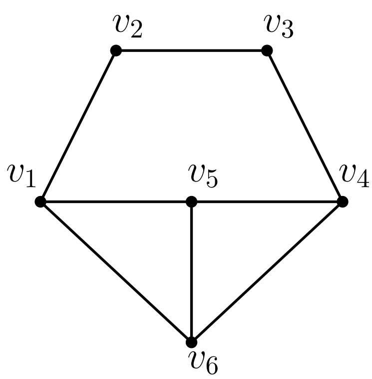



Figure 4. Graph G.

It is easily seen that a signal set is not in general a geodetic set in a graph $G$. Also the converse ir true. We verify that if $S$ is a signal set and $D$ is a geodetic set of $G$, either $S \subseteq D$ or $D \subseteq S$. Hence the signal set and the geodetic set depend one to another. Therefore we can't find out which one is the bigger set.

**Result 8.** If $G$ is a tree, then $\operatorname{sn}(G) = g(G)$.

**Theorem 13.** *If $G$ is a regular graph, then* $\operatorname{sn}(G) = g(G)$.

P r o o f. Since $G$ is regular, the degree of every vertices of $G$ is unique. So the signal distance between any pair of vertices depends only the geodesic distance between this pair of vertices. Hence, $\operatorname{sn}(G) = g(G)$. □

In view of Theorem 4, we have the following realization theorems.

**Theorem 14.** *For any integers $a, b$ and $c$ with $3 \le a \le b \le c$, there exists a connected graph $G$ with $g(G) = a$, $\operatorname{sn}(G) = b$ and $\operatorname{dsn}(G) = c$.*

P r o o f. This theorem is proved by considering four cases.

*Case 1.* $a = b = c$. Then for the complete graph $K_a$, $g(G) = \operatorname{sn}(G) = \operatorname{dsn}(G) = a$.

*Case 2.* $a = b < c$. let $G$ be the graph in Fig. 5 obtained from the path $P_3 : u_1, u_2, u_3$ of order 3, by adding $c$ vertices $v_1, v_2, ..., v_{a-2}, w_1, w_2, ..., w_{c-a+2}$ to $P_3$ and joining each vertex $v_i$ $(1 \le i \le a-2)$ to $u_2$; and joining each vertex $w_i(1 \le j \le c - a + 2)$ to $u_1$ and $u_3$. By Theorem 2, Theorem 5 and Corollary 1, every geodetic set, every signal set and every double signal set of $G$ contains the set $S = \{v_1, v_2, ..., v_{a-2}\}$ of all extreme vertices of $G$. Clearly, $S$ is not a geodetic set of $G$. Also, for any $x \in V(G) - S$, $S \cup \{x\}$ is not a geodetic set or a signal set of $G$ and so $g(G) \ge a$. Now it is easy to check that $S_1 = S \cup \{u_1, u_2\}$ is a geodetic set of $G$. Since every vertex in $V(G) - S_1$ lies on the signal path between some vertices from $S_1$, so $S_1$ is the minimum geodetic set as well as signal set of $G$. Thus $g(G) = \operatorname{sn}(G) = a$. It is clear that the pair of vertices $v_i, w_j$ for $(1 \le i \le a - 2)$ and $(1 \le j \le c - a + 2)$ do not lie on any $u$-$v$ geosig path, for any $u, v \in S_1$ and so that $S_1$ is not a double signal set of $G$. It is easy to verify that $S_2 = S \cup \{w_1, w_2, ..., w_{c-a+2}\}$ is a minimum double signal set of $G$ and so $\operatorname{dsn}(G) = c$.

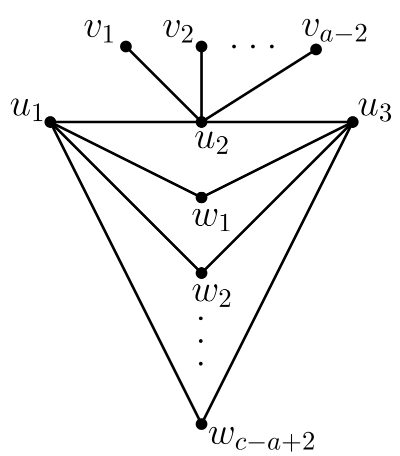



Figure 5. Graph G.

*Case 3.* $a < b = c$. Let $G$ be the graph in Fig. 6 got from the complete graph $K_{b-a+2}$ and the path $P_3 : x, y, z$ of order 3 by joining all the vertices of $K_{b-a+2}$ to $x$ and $y$ and adding $a-2$ new pendant edges $v_1, v_2, ..., v_{a-2}$. By Theorem 2, Theorem 5 and Corollary 1, every geodetic set, every signal set and every double signal set of $G$ contain the set $S = \{v_1, v_2, ..., v_{a-2}\}$ of all extreme vertices of $G$. Clearly, $S$ is not a geodetic set of $G$. Also, for any $u \in V(G) - S$, $S \cup \{u\}$ is not a geodetic set or a signal set of $G$ and so $g(G) \geq a$. Since $S_1 = S \cup \{x, y\}$ is a geodetic set of $G$, it follows that $g(G) = a$. It is clear that the vertices of $K_{b-a+2}$ do not lie on any signal path between vertices from $S_1$, that $S_1$ is not a signal set of $G$. Clearly, every signal set and every double signal set contains every vertices of $K_{b-a+2}$. Now it is easily to verify that $S \cup V(K_{b-a+2})$ is a minimum signal set and minimum double signal of $G$. Hence, $\operatorname{sn}(G) = \operatorname{dsn}(G) = a - 2 + b - a + 2 = b$.

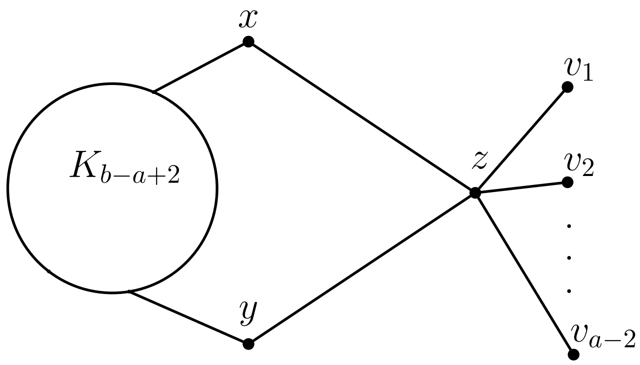



Figure 6. Graph G.

*Case 4.* $a < b < c$. Let $G$ be the graph in Fig. 7 obtained from the path $P_3 : x, y, z$ of order 3 by adding $c$ new vertices $u_1, u_2, ..., u_{a-2}, w_1, w_2, ..., w_{b-a}, v_1, v_2, ..., v_{c-b+2}$ to $P_3$ and joining each vertex $w_i (1 \leq i \leq b-a)$ to the vertices $x, y$ and $z$ ; joining each vertex $v_j (1 \leq j \leq c - b + 2)$ to the vertices $x$ and $z$; joining each vertex $u_k (1 \leq k \leq a-2)$ to the vertex $y$. By Theorems 2, Theorem 5 and Corollary 1, every geodetic set, every signal set and every double signal set of $G$ contain the set $S = \{v_1, v_2, ..., v_{a-2}\}$ of all extreme vertices of $G$. Clearly, $S$ is not a geodetic set of $G$. Also, for any $v \in V(G) - S$, $S \cup \{v\}$ is not a geodetic set of $G$ and so $g(G) \geq a$. Since $S_1 = S \cup \{x, z\}$ is a geodetic set of $G$, it follows that $g(G) = a$. Since each vertex $w_j$ does not lie on any geosig of vertices of $S_1$, that $S_1$ is not a signal set of $G$. It is clear that every signal set of $G$ contains $\{w_1, w_2, ..., w_{b-a}\}$. Then $S_2 = S_1 \cup \{w_1, w_2, ..., w_{b-a}\}$ is a minimum signal set of $G$ and so $\operatorname{sn}(G) = b$. Now, each pair $v_j$ is either an initial vertex or terminal vertex of any signal path containing the vertices $v_j$ and $w_i$. Hence $v_1, v_2, ..., v_{c-b+2}$ are weak extreme vertices. It is easily verified that the set $S_3 = S \cup \{w_1, w_2, ..., w_{b-a}, v_1, v_2, ..., v_{c-b+2}\}$ is the unique minimum double signal set of $G$ and so $\operatorname{dsn}(G) = c$.

**Theorem 15.** *For integers $a, b$ and $c$ with $3 \leq a \leq b \leq c$, there exists a connected graph $G$ with $\operatorname{sn}(G) = a$, $g(G) = b$ and $\operatorname{dsn}(G) = c$.*

P r o o f. This theorem is proved by considering three cases.

*Case 1.* $a = b = c$. Then for the complete graph $K_a$, $g(G) = \operatorname{sn}(G) = \operatorname{dsn}(G) = a$.

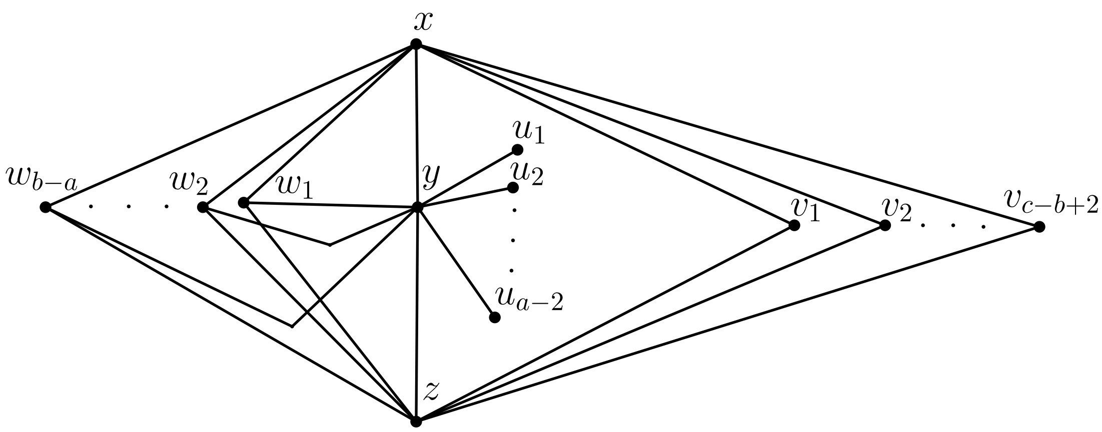



Figure 7. Graph G.

*Case 2.* $a = b < c$. The proof is similar to the proof of case 2 in Theorem 14.

*Case 3.* $a < b \leq c$. Let $H$ be the graph obtained from the path $P_6 : u, v, w, x, y, z$ of order 6, $b - a$ copies of path $P_i : x_i, y_i, z_i$ $(1 \leq i \leq b - a)$ of order 3 and 2 copies of path $P_j : x'_j, y'_j$ $(1 \leq j \leq 2)$ of order 3 by joining each vertex $x_i(1 \leq i \leq b - a)$ to the vertex $u$ of $P_6$, each vertex $z_i(1 \leq i \leq b - a)$ to the vertex $v$ of $P_6$, each vertex $x'_j$ $(1 \leq j \leq 2)$ to the vertex $u$ of $P_6$ and each vertex $y'_j$ $(1 \leq j \leq 2)$ to the vertex $v$ of $P_6$ and add an edge $x'_1 y'_2$. Let $G$ be the graph in Fig. 8 obtained from $H$ by adding the following new vertices to $H$.

(i) Add $a - 1$ new vertices $u_1, u_2, ..., u_{a-1}$ to $H$ and join each $u_i(1 \leq i \leq a - 1)$ to $w$.

(ii) Add $b - a - 1$ new vertices $v_1, v_2, ..., v_{b-a-1}$ to $H$ and join each $v_i(1 \leq i \leq b - a - 1)$ to both $w$ and $y$.

(iii) Add $c - b$ new vertices $w_1, w_2, ..., w_{c-b}$ to $H$ and join each $w_i(1 \leq i \leq c - b)$ to both $v$ and $x$.

Let $S = \{u_1, u_2, ..., u_{a-1}, z\}$ be the set of extreme vertices of $G$. By Theorem 5, Theorem 2 and Corollary 1, every signal set, every geodetic set and every double signal set contains $S$. Clearly $S$ itself is not a signal set of $G$ and so $\operatorname{sn}(G) \geq a$. It is clear that $S_1 = S \cup \{u\}$ is a signal set of $G$ and hence $\operatorname{sn}(G) = a$. Since the vertices $x_i, y_i, z_i(1 \leq i \leq b - a)$ do not lie on any geodesic joining a pair of vertices from $S_1$, $S_1$ is not a geodetic set of $G$. Let $S_2 = S \cup \{y_1, y_2, ..., y_{b-a}\}$. It is easy to verify that $S_2$ is a minimum geodetic set of $G$ and so $g(G) = a + b - a = b$. Since the pair of vertices $w_i(1 \leq i \leq c - b), v_j(1 \leq j \leq b - a - 1)$ do not lie on any signal path between a pair of vertices from $S_1$, $S_1$ is not a double signal set of $G$. Also, $x$ is either an initial vertex or terminal vertex of any geosig containing the vertices $x$ and $v_1$ and so $x$ is a weak extreme vertex. Hence $w_1, w_2, ..., w_{c-b}, v_1, v_2, ..., v_{b-a-1}, x$ are weak extreme vertices.

Let $S'= S_1 \cup \{w_1, w_2, ..., w_{c-b}, v_1, v_2, ..., v_{b-a-1}, x\}$. It is easily verified that $S'$ is the set of all weak extreme vertices of $G$. Since $S'$ is a double signal set of $G$, by Theorem 6 it follows that $\operatorname{dsn}(G) = c$. □

**Theorem 16.** *For every pair $a, b$ of integers with $4 \leq a \leq b$ and $b \neq a + 1$, there exists a connected graph $G$ with $dg(G) = a$ and $\operatorname{dsn}(G) = b$.*

P r o o f. For $4 \leq a = b$, then the complete graph $K_a$ has the desired properties. So, assume that $4 \leq a < b$ and $b \neq a + 1$. Let $G$ be the graph in Fig. 9 formed from the path $P_4 : u, v, w, y$ of

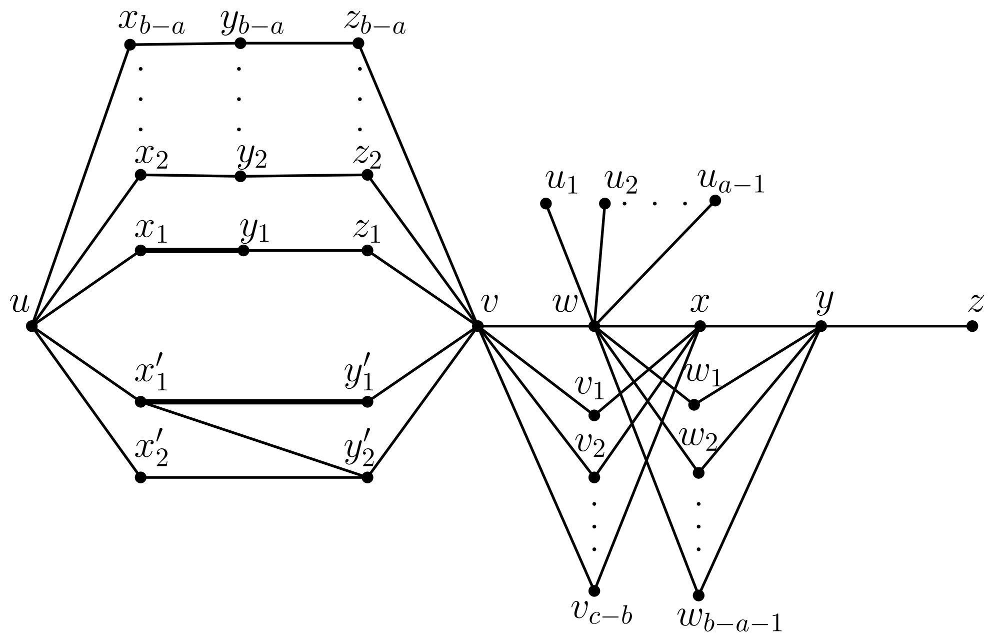



Figure 8. Graph G.

order 4, by adding $b$ new vertices $u_1, u_2, ..., u_{a-3}, v_1, v_2, ..., v_{b-a-1}, x$ to $P_4$ and joining each vertex $u_i (1 \leq i \leq a-3)$ to $v$; and joining each vertex $v_j (1 \leq j \leq b-a-1)$ to both $v$ and $y$; and join the vertex $x$ to $u$ and $w$. Let $S = \{u_1, u_2, ..., u_{a-3}\}$ be the set of extreme vertices of $G$. By Theorem 3 and Corollary 1, every double geodetic and every double signal set contains $S$. Now it is clear that $S_1 = S \cup \{u, x, y\}$ is a minimum double geodetic set of $G$ and so $\operatorname{dg}(G) = a$. Since $L[u, y]$ contains $u, x, w, y$, the pair of vertices $x, v_i (1 \leq i \leq b-a-1)$ do not lie on any geosig of a pair of vertices from $S$. So that $S_1$ is not a double signal set of $G$. It is easy to verify that $S_2 = S \cup \{v_1, v_2, ..., v_{b-a-1}, w\}$ be the unique minimum double signal set of $G$. Hence, $\operatorname{dsn}(G) = b$. $\square$

## 4. Closing open problems

We close with the following list of open problems that we have yet to settle.

**Problem 1.** Determine the class of graphs $G$ for which $g(G) = \operatorname{sn}(G)$.

**Problem 2.** Determine the class of graphs $G$ for which $\operatorname{sn}(G) = \operatorname{dsn}(G)$.

### REFERENCES

1. Buckley F., Harary F. *Distance in Graphs.* Redwood City, Calif: Addison-Wesley Pub. Co., 1990. 335 p.
2. Balamurugan S., Antony Doss R. The edge signal number of a graph. *Discrete Math. Algorithms Appl.*, 2021. Vol. 13, No. 03. Art. no. 2150024. DOI: 10.1142/s1793830921500245
3. Chartrand G., Harary F., Swart H.C., Zhang P. Geodomination in graphs. Bull. Inst. Combin. Appl., 2001. Vol. 31. P. 51–59.
4. Chartrand G., Harary F., Zhang P. On the geodetic number of a graph. *Networks*, 2002. Vol. 39, No. 1. P. 1–6. DOI: 10.1002/net.10007
5. Kathiresan K.M, Sumathi R. A study on signal distance in graphs. *Algebra, Graph Theory, Appl.*, 2009. P. 50–54.

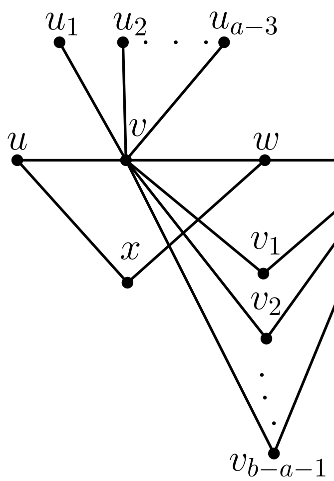



Figure 9. Graph G.


6. Ostrand P.A. Graphs with specified radius and diameter. *Discrete Math.*, 1973. Vol. 4, No. 1. P. 71–75. DOI: 10.1016/0012-365X(73)90116-7
7. Santhakumaran A.P., Jebaraj T. Double geodetic number of a graph. *Discuss. Math. Graph Theory*, 2012. Vol. 32, No. 1. P. 109–119.
8. Sethu Ramalingam S., Balamurugan S. On the signal distance in graphs. *Ars Combinatoria*, (accepted on 2018).

URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 76–89
DOI: 10.15826/umj.2022.1.008

# MATRIX RESOLVING FUNCTIONS IN THE LINEAR GROUP PURSUIT PROBLEM WITH FRACTIONAL DERIVATIVES[1]

**Alena I. Machtakova**†**, Nikolai N. Petrov**††

Udmurt State University,
1 Universitetskaya Str., Izhevsk, 426034, Russian Federation

Krasovskii Institute of Mathematics and Mechanics,
Ural Branch of the Russian Academy of Sciences,
16 S. Kovalevskaya Str., Ekaterinburg, 620108, Russian Federation

†bichurina.alyona@yandex.ru, ††kma3@list.ru

**Abstract:** In finite-dimensional Euclidean space, we analyze the problem of pursuit of a single evader by a group of pursuers, which is described by a system of differential equations with Caputo fractional derivatives of order $\alpha$. The goal of the group of pursuers is the capture of the evader by at least $m$ different pursuers (the instants of capture may or may not coincide). As a mathematical basis, we use matrix resolving functions that are generalizations of scalar resolving functions. We obtain sufficient conditions for multiple capture of a single evader in the class of quasi-strategies. We give examples illustrating the results obtained.

**Keywords:** Differential game, Group pursuit, Pursuer, Evader, Fractional derivatives.

## 1. Introduction

The theory of two-player differential games, originally considered by Isaacs [20], has grown to be a profound and substantial theory that develops various approaches to the analysis of conflict situations [3, 14, 15, 19, 21, 22, 24, 36, 40]. The following methods for solving game problems were developed: the Isaacs method based on the analysis of some partial differential equation and its characteristics, the method of stable bridges, Krasovskii's rule of extremal aiming, Pontryagin's method based on alternating integration of convex sets, etc.

In [6, 7], Chikrii proposed a method of scalar resolving functions using Pontryagin's condition and, based on it, measurable choice theorems.

The method of scalar resolving functions was developed further to investigate linear and quasi-linear group pursuit problems [2, 10, 18, 28–30, 38, 39]. In [8], Chikrii noted that scalar resolving functions attract the terminal set to the images of some multivalued maps. This attraction occurs in the conical hull of this set, which restricts the maneuverability of pursuers.

In [8, 11], for the analysis of two-player pursuit games, matrix resolving functions were proposed. In [26], matrix resolving functions were applied to studying the group pursuit problem described by a linear autonomous system of differential equations.

In the present paper, we consider matrix resolving functions in a linear group pursuit problem described by a system of differential equations with Caputo fractional derivatives. It should be noted that matrix resolving functions for solving group pursuit problems with fractional derivatives are used for the first time. Previously, scalar resolving functions were used in [23, 25, 27] devoted to this class of problems. We obtain sufficient conditions for multiple capture of a single evader.

[1]This work was supported by the Russian Science Foundation, project 21-71-10070.

The multiple capture of a single evader in the simple pursuit problem was considered in [4, 17]; [4] investigated it in a discrete setting. In [31, 32], the problem of multiple capture of a single evader was presented in the example of L.S. Pontryagin, and in [1, 33] it was considered in linear differential games.

## 2. Statement of the problem

**Definition 1** [5]. *Let $f\colon [0,\infty) \to \mathbb{R}^k$ be an absolutely continuous function and $\alpha \in (0,1)$. The Caputo derivative of order $\alpha$ of the function $f$ is defined to be a function $D^{(\alpha)} f$ of the form*

$$\big(D^{(\alpha)} f\big)(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{f'(s)}{(t-s)^\alpha}\, ds, \quad \textit{where} \quad \Gamma(\beta) = \int_0^\infty e^{-s} s^{\beta-1}\, ds.$$

In the space $\mathbb{R}^k$ $(k \geq 2)$, we consider a differential game $G(n+1)$ involving $n+1$ players: $n$ pursuers $P_1, \ldots, P_n$ and an evader $E$, which is described by a system of the form

$$D^{(\alpha)} z_i = A_i z_i + u_i - v, \quad z_i(0) = z_i^0, \quad u_i \in U_i, \quad v \in V. \tag{2.1}$$

Here $i \in I = \{1, \ldots, n\}$, $z_i, u_i, v \in \mathbb{R}^k$, $U_i$ and $V$ are compact sets from $\mathbb{R}^k$, $\alpha \in (0,1)$, $D^{(\alpha)} f$ is the Caputo derivative of order $\alpha$ of the function $f$, and $A_i$ are constant square matrices of order $k \times k$. Assume that $z_i^0 \neq 0$ for all $i \in I$. Define $z^0 = \{z_i^0, i \in I\}$ to be the vector of initial positions.

Let $v : [0,\infty) \to V$ be a measurable function. Let us call the restriction of the function $v$ to $[0,t]$ the prehistory $v_t(\cdot)$ of the function $v$ at time $t$.

**Definition 2.** *We will say that a quasi-strategy $\mathcal{U}_i$ of a pursuer $P_i$ is given if a map $U_i^0$ is defined that associates a measurable function $u_i(t)$ with values in $U_i$ to the initial positions $z^0$, time $t$, and arbitrary prehistory of control $v_t(\cdot)$ of the evader $E$.*

**Definition 3.** *An $m$-fold capture (a capture for $m = 1$) occurs in the game $G(n+1)$ if there exist a time $T > 0$ and quasi-strategies $\mathcal{U}_1, \ldots, \mathcal{U}_n$ of pursuers $P_1, \ldots, P_n$ such that, for any measurable function $v(\cdot)$, $v(t) \in V$, $t \in [0,T]$, there exist times $\tau_1, \ldots, \tau_m \in [0,T]$ and pairwise different indices $i_1, \ldots, i_m \in I$ such that $z_{i_l}(\tau_l) = 0$ for all $l = 1, \ldots, m$.*

The aim of this paper is to obtain conditions for the solvability of the pursuit problem.

**Assumption 1.** *For all $i \in I$, it is true that $0 \in \bigcap\limits_{v \in V} \big(U_i - v\big)$.*

In what follows, we assume that Assumption 1 holds. We introduce the following notation:

$$E_\rho(B,\mu) = \sum_{l=0}^{\infty} \frac{B^l}{\Gamma(l\rho^{-1} + \mu)},$$

which is a generalized Mittag-Leffler function [16], where $B$ is a square matrix of order $k \times k$, $\rho > 0$, and $\mu \in \mathbb{R}^1$; $\Delta = \big\{(t,\tau) \ : \ t \geq 0, \ 0 \leq \tau \leq t\big\}$, $J = \{1, \ldots, k\}$,

$$g_i(t,\tau) = (t-\tau)^{\alpha-1} E_{\frac{1}{\alpha}}(A_i (t-\tau)^\alpha, \alpha), \quad \tau \neq t, \quad g(t,t) = 0,$$
$$f_i(t) = E_{\frac{1}{\alpha}}(A_i t^\alpha, 1) z_i^0, \quad W_i(t,\tau,v) = g_i(t,\tau)\big(U_i - v\big),$$
$$W_i(t,\tau) = \bigcap_{v \in V} W_i(t,\tau,v), \quad i \in I, \quad 0 \leq \tau \leq t,$$

where $(t,\tau)\in\Delta$ and $v\in V$.

Consider an arbitrary diagonal square matrix $L_i$ of order $k\times k$ of the form

$$L_i=\begin{pmatrix}\lambda_{i1} & 0 & \dots & 0\\ 0 & \lambda_{i2} & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & \lambda_{ik}\end{pmatrix}=\operatorname{diag}(\lambda_{i1},\lambda_{i2},\dots,\lambda_{ik}).$$

We identify the matrix $L_i$ with the vector $(\lambda_{i1},\dots,\lambda_{ik})$, understand the inequality $L_i\ge 0$ coordinatewise, and introduce the multivalued maps

$$\mathcal{M}_i(t,\tau,v)=\big\{L_i\ :\ L_i\ge 0, -L_if_i(t)\in W_i(t,\tau,v)\big\},\quad (t,\tau)\in\Delta,\quad v\in V.$$

By Assumption 1, for all $i\in I$, $v\in V$, and $t,\tau$ such that $0\le\tau\le t$, the sets $W_i(t,\tau,v)$ are not empty and $0\in\mathcal{M}_i(t,\tau,v)$. By the properties of the parameters of the conflict-controlled process (2.1), the maps $\mathcal{M}_i(t,\tau,v)$ are measurable in $\tau$ [12]. Then the maps $W_i(t,\tau)$ are measurable in $\tau$ [12].

Define the scalar functions

$$\lambda_i^0(t,\tau,v)=\sup_{L_i\in\mathcal{M}_i(t,\tau,v)}\min_{j\in J}\lambda_{ij}(t,\tau,v),\quad (t,\tau)\in\Delta,\quad v\in V. \tag{2.2}$$

**Assumption 2.** *For all $(t,\tau)\in\Delta$ and $v\in V$, the supremum in (2.2) is attained.*

Assuming that the supremum in (2.2) is attained, we define the sets

$$\mathcal{M}_i^*(t,\tau,v)=\big\{L_i(t,\tau,v)\in\mathcal{M}_i(t,\tau,v)\ :\ \lambda_i^0(t,\tau,v)=\min_j\lambda_{ij}(t,\tau,v)\big\}.$$

It follows from [12] that, under the above assumptions, $\mathcal{M}_i(t,\tau,v)$ and $\mathcal{M}_i^*(t,\tau,v)$ are measurable in $(\tau,v)$ and closed-valued for any $t\ge 0$. By the measurable choice theorem [35, Theorem 20.6], for each $i\in I$ in $\mathcal{M}_i^*(t,\tau,v)$, there exists at least one selector measurable in $(\tau,v)$ for any $t\ge 0$. We fix these selectors $L_i^*(t,\tau,v)$ and define $\lambda_i^*(t,\tau,v)=\min_j\lambda_{ij}^*(t,\tau,v)$. Next, define

$$\Omega(m)=\{(i_1,\dots,i_m)\ :\ i_1,\dots,i_m\in I \text{ and are pairwise different}\},$$
$$\delta(t,\tau)=\inf_{v\in V}\max_{\Lambda\in\Omega(m)}\min_{l\in\Lambda}\lambda_l^*(t,\tau,v).$$

## 3. Sufficient conditions for capture

**Lemma 1.** *Suppose that Assumptions 1 and 2 hold and*

$$\lim_{t\to+\infty}\int_0^t\delta(t,s)ds=+\infty.$$

*Then there exists a time $T>0$ such that, for every measurable function $v(\cdot)$, $v(t)\in V$, $t\in[0,T]$, there is a set $\Lambda\in\Omega(m)$ such that the following inequalities hold for all $l\in\Lambda$, $j\in J$:*

$$\int_0^T\lambda_{lj}^*(T,s,v(s))ds\ge 1.$$

P r o o f. Let $v(\cdot)$ be an arbitrary measurable function, $v\colon [0,\infty) \to V$. Then the inequalities

$$\lambda_{lj}^*(t,s,v(s)) \geq \lambda_l^*(t,s,v(s))$$

hold for all $t > 0$, $s \in [0,t]$, $l \in I$, and $j \in J$. Therefore, the inequalities

$$\int\limits_0^t \lambda_{lj}^*(t,s,v(s))ds \geq \int\limits_0^t \lambda_l^*(t,s,v(s))ds \tag{3.1}$$

hold for all $t \geq 0$, $l \in I$, and $j \in J$. In addition,

$$\max_{\Lambda\in\Omega(m)} \min_{l\in\Lambda} \int_0^t \lambda_l^*(t,s,v(s))ds \geq \max_{\Lambda\in\Omega(m)} \int_0^t \min_{l\in\Lambda} \lambda_l^*(t,s,v(s))ds. \tag{3.2}$$

Since, for any nonnegative numbers $a_\Lambda (\Lambda \in \Omega(m))$, one has

$$\max_{\Lambda\in\Omega(m)} a_\Lambda \geq \frac{1}{C_n^m} \sum_{\Lambda\in\Omega(m)} a_\Lambda, \quad \text{where} \quad C_n^m = \frac{n!}{(n-m)!\,m!},$$

it follows from (3.2) that

$$\max_{\Lambda\in\Omega(m)} \min_{l\in\Lambda} \int\limits_0^t \lambda_l^*(t,s,v(s))ds \geq \frac{1}{C_n^m} \int\limits_0^t \sum_{\Lambda\in\Omega(m)} \min_{l\in\Lambda} \lambda_l^*(t,s,v(s))ds \geq$$
$$\geq \frac{1}{C_n^m} \int\limits_0^t \max_{\Lambda\in\Omega(m)} \min_{l\in\Lambda} \lambda_l^*(t,s,v(s))ds \geq \frac{1}{C_n^m} \int\limits_0^t \delta(t,s)ds.$$

Since

$$\int\limits_0^t \delta(t,s)ds = +\infty,$$

there exists $T > 0$ such that

$$\frac{1}{C_n^m} \int\limits_0^T \delta(T,s)ds \geq 1.$$

Hence,

$$\max_{\Lambda\in\Omega(m)} \min_{l\in\Lambda} \int\limits_0^T \lambda_l^*(T,s,v(s))ds \geq 1.$$

Therefore, there exists $\Lambda \in \Omega(m)$ such that the following inequalities hold for all $l \in \Lambda$:

$$\int\limits_0^T \lambda_l^*(T,s,v(s))ds \geq 1.$$

This inequality and inequality (3.1) imply the validity of the lemma. $\square$

Let $\mathcal{V}$ be the set of all measurable functions $v\colon [0,\infty) \to V$. Let us define the number

$$\hat{T} = \inf\{t \geq 0 \ : \ \inf_{v(\cdot)\in\mathcal{V}} \max_{\Lambda\in\Omega(m)} \min_{l\in\Lambda} \min_{j\in J} \int_0^t \lambda^*_{lj}(t,s,v(s))ds \geq 1\}.$$

Consider the sets $(i \in I, j \in J, v(\cdot) \in \mathcal{V})$

$$T_{ij}(v(\cdot)) = \Big\{t \geq 0 \ : \ \int_0^t \lambda^*_{ij}(\hat{T},s,v(s))ds \geq 1\Big\}.$$

Define the quantities $(i \in I, j \in J, v(\cdot) \in \mathcal{V})$

$$t^*_{ij}(v(\cdot)) = \begin{cases} \inf\{t \ : \ t \in T_{ij}(v(\cdot))\} & \text{if} \quad T_{ij}(v(\cdot)) \neq \emptyset, \\ +\infty & \text{if} \quad T_{ij}(v(\cdot)) = \emptyset. \end{cases} \tag{3.3}$$

**Assumption 3.** *For any $\tau \in [0,\hat{T}]$, $v \in V$, $l \in I$, and $J_0 \subset J$, the selector*

$$B_l(\hat{T},\tau,v) = \operatorname{diag}\big(\beta_{l1}(\hat{T},\tau,v),\ldots,\beta_{lk}(\hat{T},\tau,v)\big),$$

*where*

$$\beta_{lj}(\hat{T},\tau,v) = \begin{cases} \lambda^*_{lj}(\hat{T},\tau,v), & j \in J_0, \\ 0, & j \notin J_0, \end{cases}$$

*satisfies the condition $B_l(\hat{T},\tau,v) \in \mathcal{M}_l(\hat{T},\tau,v)$.*

**Theorem 1.** *Suppose that Assumptions 1, 2, and 3 hold and*

$$\lim_{t\to+\infty} \int_0^t \delta(t,s)ds = +\infty.$$

*Then an $m$-fold capture occurs in the game $G(n+1)$.*

P r o o f. By Lemma 1, $\hat{T} < +\infty$. Let $v : [0,\hat{T}] \to V$ be an arbitrary measurable function and $\tau \in [0,\hat{T}]$. Let us introduce functions $\big(\beta_{i1}(\hat{T},\tau,v),\ldots,\beta_{ik}(\hat{T},\tau,v)\big)$ of the form

$$\beta_{ij}(\hat{T},\tau,v) = \begin{cases} \lambda^*_{ij}(\hat{T},\tau,v), & \tau \in [0,t^*_{ij}(v(\cdot))], \\ 0, & \tau \in (t^*_{ij}(v(\cdot)),\hat{T}], \end{cases}$$

where $t^*_{ij}(v(\cdot))$ are defined by formula (3.3). Let $B^*_i(\hat{T},s,v)$ be a matrix of the form

$$B^*_i(\hat{T},s,v) = \begin{pmatrix} \beta^*_{i1}(\hat{T},s,v) & 0 & \ldots & 0 \\ 0 & \beta^*_{i2}(\hat{T},s,v) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \beta^*_{ik}(\hat{T},s,v) \end{pmatrix}.$$

Consider the multivalued maps $(s \in [0,\hat{T}],\ v \in V)$

$$\tilde{U}_i(\hat{T},s,v) = \big\{u_i \in U_i \ : \ g_i(\hat{T},s)(u_i - v) = -B^*_i(\hat{T},s,v)f_i(\hat{T})\big\}.$$

By Assumption 3, $B_i^*(\hat{T}, s, v)$ is a measurable selector of $\mathcal{M}_i(\hat{T}, s, v)$. Therefore, the sets $\tilde{U}_i(\hat{T}, s, v)$ are nonempty for all $i \in I$, $s \in [0, \hat{T}]$, and $v \in V$. Hence, by the measurable choice theorem [35, Theorem 20.6], there exists at least one measurable selector $u_i^*(\hat{T}, s, v)$. We define the controls of pursuers $P_i$, $i \in I$, assuming

$$u_i(\tau) = u_i^*(\hat{T}, \tau, v(\tau)).$$

By [12], the functions $u_i(\cdot)$ are measurable. We show that these controls of the pursuers guarantee the $m$-fold capture of the evader. The solution of the Cauchy problem for system (2.1) has the form [9]:

$$z_i(t) = f_i(t) + \int_0^t g_i(t, s)(u_i(s) - v(s))ds.$$

By the choice of controls of the pursuers, we obtain

$$z_i(\hat{T}) = f_i(\hat{T}) - \int_0^{\hat{T}} B_i(\hat{T}, s, v(s))f_i(\hat{T})ds = \Big(E - \int_0^{\hat{T}} B_i(\hat{T}, s, v(s))ds\Big)f_i(\hat{T}),$$

where $E$ is an identity matrix. It follows from the definition of $B_i(\hat{T}, s, v(s))$ that there exists $\Lambda \in \Omega(m)$ such that $z_l(\hat{T}) = 0$ for all $l \in \Lambda$. This proves the theorem. $\square$

**Assumption 4.** *The matrices $A_i$ are diagonal matrices of the form*

$$A_i = \begin{pmatrix} a_{i1} & 0 & \dots & 0 \\ 0 & a_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{ik} \end{pmatrix} \quad \textit{with} \;\; a_{ij} \le 0 \;\; \textit{for all} \quad i \in I, \quad j \in J.$$

Let us introduce multivalued maps ($v \in V$)

$$\mathcal{M}_i^0(v) = \big\{L_i \;:\; L_i \ge 0, -L_i z_i^0 \in \big(U_i - v\big)\big\}.$$

By Assumption 1, the sets $\mathcal{M}_i^0(v)$ for all $i \in I$ and $v \in V$ are nonempty and $0 \in \mathcal{M}_i^0(v)$. Next, we define functions $\overline{\lambda}_i(v)$ of the form

$$\overline{\lambda}_i(v) = \sup_{L_i \in \mathcal{M}_i^0(v)} \min_j \lambda_{ij}(v). \tag{3.4}$$

**Assumption 5.** *For all $v \in V$, the supremum in (3.4) is attained.*

Assuming that the supremum in (3.4) is attained, we define the sets ($v \in V$)

$$\overline{\mathcal{M}}_i(v) = \big\{L_i(v) \in \mathcal{M}_i^0(v) \;:\; \overline{\lambda}_i(v) = \min_j \lambda_{ij}(v)\big\}.$$

Next, suppose that $\overline{\lambda}_i^*(v)$ is a measurable selector of $\overline{\mathcal{M}}_i(v)$ and

$$\delta = \inf_{v \in V} \max_{\Lambda \in \Omega(m)} \min_{l \in \Lambda} \overline{\lambda}_l^*(v).$$

Define $((t,\tau)\in\Delta)$

$$\begin{gathered}
a=\max_{i,j}(-a_{ij}),\\
g_{ij}(t,s)=(t-s)^{\alpha-1}E_{\frac{1}{\alpha}}\big(a_{ij}(t-s)^{\alpha},\alpha\big),\quad t\neq s,\\
g(t,s)=(t-s)^{\alpha-1}E_{\frac{1}{\alpha}}\big(-a(t-s)^{\alpha},\alpha\big),\quad t\neq s,\\
g_{ij}(t,t)=g(t,t)=0.
\end{gathered}$$

**Lemma 2.** *Suppose that Assumptions 1, 4, and 5 hold, and $\delta>0$ and $a_{ij}<0$ for all $i\in I$ and $j\in J$. Then there exists $T>0$ such that, for every admissible function $v(\cdot)$, there is a set $\Lambda\in\Omega(m)$ such that the following inequalities hold for all $l\in\Lambda$ and $j\in J$:*

$$E_{\frac{1}{\alpha}}(a_{lj}T^{\alpha},1)-\int\limits_0^T g_{lj}(T,s)\overline{\lambda}^*_{lj}(v(s))\,ds\leq 0.$$

P r o o f. Let $v(\cdot)$ be an admissible function. Then $0<-a_{ij}\leq a$ for all $i$ and $j$. Therefore, the following inequalities hold [34] for all $t\geq 0$, $s\in[0,t]$, $i\in I$, and $i\in J$:

$$E_{\frac{1}{\alpha}}(a_{ij}(t-s)^{\alpha},\alpha)\geq E_{\frac{1}{\alpha}}(-a(t-s)^{\alpha},\alpha).$$

It follows from [37, Theorem 4.1.1] that $E_{\frac{1}{\alpha}}(z,\mu)\geq 0$ for all $z\in\mathbb{R}^1$ and $\mu\in[\alpha,+\infty)$. Hence, the inequalities

$$\int\limits_0^t g_{ij}(t,s)\overline{\lambda}^*_{ij}(v(s))ds\geq\int\limits_0^t g(t,s)\overline{\lambda}^*_{i}(v(s))ds$$

hold for all $t\geq 0$, $i\in I$, and $j\in J$. Next, we have

$$\max_{\Lambda\in\Omega(m)}\min_{l\in\Lambda}\int\limits_0^t g(t,s)\overline{\lambda}^*_{l}(v(s))ds\geq\max_{\Lambda\in\Omega(m)}\int\limits_0^t g(t,s)\min_{l\in\Lambda}\overline{\lambda}^*_{l}(v(s))ds. \tag{3.5}$$

Using inequality (3.5), we obtain

$$\begin{gathered}
\max_{\Lambda\in\Omega(m)}\int\limits_0^t g(t,s)\min_{l\in\Lambda}\overline{\lambda}^*_{l}(v(s))ds\geq\frac{1}{C_n^m}\int\limits_0^t g(t,s)\sum_{\Lambda\in\Omega(m)}\min_{l\in\Lambda}\overline{\lambda}^*_{l}(v(s))ds\geq\\
\geq\frac{1}{C_n^m}\int\limits_0^t g(t,s)\max_{\Lambda\in\Omega(m)}\min_{l\in\Lambda}\overline{\lambda}^*_{l}(v(s))ds\geq\frac{\delta}{C_n^m}\int\limits_0^t g(t,s)ds.
\end{gathered}$$

By [13, Ch. 3, formula (1.15)],

$$\int\limits_0^t g(t,s)ds=t^{\alpha}E_{\frac{1}{\alpha}}(-at^{\alpha},\alpha+1).$$

Consider the functions ($t \in [0, \infty)$)

$$h_{ij}(t) = E_{\frac{1}{\alpha}}(a_{ij}t^{\alpha}, 1) - \frac{\delta}{C_n^m} t^{\alpha} E_{\frac{1}{\alpha}}(-at^{\alpha}, \alpha + 1).$$

Since $a_{ij} < 0, -a < 0$, it follows from [37, Theorem 1.2.1] that the following asymptotic representation holds as $t \to +\infty$:

$$E_{1/\alpha}(a_{ij}t^{\alpha}, 1) = -\frac{1}{a_{ij}t^{\alpha}\Gamma(1-\alpha)} + O\left(\frac{1}{t^{2\alpha}}\right), \quad E_{1/\alpha}(-at^{\alpha}, \alpha+1) = \frac{1}{at^{\alpha}} + O\left(\frac{1}{t^{2\alpha}}\right).$$

Therefore,

$$h_{ij}(t) = \frac{c_{ij}}{t^{\alpha}} - \frac{\delta}{aC_n^m} + O\left(\frac{1}{t^{2\alpha}}\right).$$

Consequently, $\lim\limits_{t\to+\infty} h_{ij}(t) < 0$ for all $i \in I$ and $j \in J$. Hence, there exists $T > 0$ such that $h_{ij}(T) \leq 0$ for all $i \in I$ and $j \in J$. Next, let $\Lambda \in \Omega(m)$ be such that

$$\max_{\Lambda\in\Omega(m)} \min_{l\in\Lambda} \int_0^T g(T,s)\overline{\lambda}_l^*(v(s))ds = \min_{l\in\Lambda} \int_0^T g(T,s)\overline{\lambda}_l^*(v(s))ds.$$

Then, for all $l \in \Lambda$, one has

$$\int_0^T g(T,s)\overline{\lambda}_l^*(v(s))ds \geq \frac{\delta}{C_n^m} T^{\alpha} E_{\frac{1}{\alpha}}(-aT^{\alpha}, \alpha+1).$$

Therefore,

$$-\int_0^T g_{lj}(T,s)\overline{\lambda}_{lj}^*(v(s))ds \leq -\int_0^T g(T,s)\overline{\lambda}_l^*(v(s))ds \leq -\frac{\delta}{C_n^m} T^{\alpha} E_{\frac{1}{\alpha}}(-aT^{\alpha}, \alpha+1).$$

Hence, the inequalities

$$E_{\frac{1}{\alpha}}(a_{lj}T^{\alpha}, 1) - \int_0^T g_{lj}(T,s)\overline{\lambda}_{lj}^*(v(s))\,ds \leq E_{\frac{1}{\alpha}}(a_{lj}T^{\alpha}, 1) - \frac{\delta}{C_n^m} T^{\alpha} E_{\frac{1}{\alpha}}(-aT^{\alpha}, \alpha+1) \leq 0$$

hold for all $l \in \Lambda$ and $j \in J$. This proves the lemma. □

**Lemma 3.** *Suppose that Assumptions 1, 4, and 5 hold, $a_{ij} \leq 0$, and $\delta > 0$. Then there exists $T > 0$ such that, for every admissible function $v(\cdot)$, there is a set $\Lambda \in \Omega(m)$ such that the following inequalities hold for all $l \in \Lambda$ and $j \in J$:*

$$E_{\frac{1}{\alpha}}(a_{lj}T^{\alpha}, 1) - \int_0^T g_{lj}(T,s)\overline{\lambda}_{lj}^*(v(s))\,ds \leq 0.$$

P r o o f. The proof is similar to the proof of Lemma 2.

□

**Assumption 6.** *For all $v \in V$, $l \in I$, and $J_0 \subset J$, the selector*

$$B_l(v) = \operatorname{diag}(\beta_{l1}(v), \ldots, \beta_{lk}(v)),$$

*where*

$$\beta_{lj}(v) = \begin{cases} \lambda^*_{lj}(v), & j \in J_0, \\ 0, & j \notin J_0, \end{cases}$$

*satisfies the condition $B_l(v) \in \mathcal{M}^0_l(v)$.*

*Remark 1.* Note that Assumption 6 does not always hold. Suppose that, in system (2.1), $k = 2$, $n = 1$, $m = 1$, $z^0_1 = (1, 2)$, $A_1$ is a zero matrix, and

$$U_1 = V = \{(u_1, u_2) : \ u_1 = u_2, \ u_2 \in [-1, 1]\}.$$

Let $v = 0$. Then

$$\mathcal{M}^0_1(0) = \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda/2 \end{pmatrix}, \quad \lambda \in [0, 1] \right\}.$$

Therefore,

$$\sup_{L \in \mathcal{M}^0_1(0)} \min_j \lambda_{1j} = \frac{1}{2}.$$

Hence,

$$\overline{\mathcal{M}}_1(0) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \right\}$$

and the extremal selector is $\lambda^*_1(0) = \operatorname{diag}(1, 1/2)$. However, the selector $B_1(0) = \operatorname{diag}(1, 0) \notin \mathcal{M}^0_1(0)$. Similarly, the selector $B_2(0) = \operatorname{diag}(0, 1/2) \notin \mathcal{M}^0_1(0)$.

*Remark 2.* If Assumption 1 holds, in particular, if the sets $U_i$ have the form $U_i = [a_{i1}, b_{i1}] \times [a_{i2}, b_{i2}] \times \ldots \times [a_{ik}, b_{ik}]$ for all $i$, then Assumption 6 also holds.

**Theorem 2.** *Suppose that Assumptions 1, 4, 5, and 6 hold and $\delta > 0$. Then an $m$-fold capture occurs in the game $G(n + 1)$.*

P r o o f. Define the number

$$\hat{T} = \inf\Big\{ t \geq 0 \ : \ \sup_{v(\cdot) \in \mathcal{V}} \min_{\Lambda \in \Omega(m)} \max_{l \in \Lambda} \max_{j \in J} \Big( E_{\frac{1}{\alpha}}(a_{lj} t^\alpha, 1) - \int_0^t g_{lj}(t, s) \overline{\lambda}^*_{lj}(v(s)) ds \Big) \leq 0 \Big\}.$$

Then, by Lemma 3, $\hat{T} < +\infty$. Let $v(\cdot)$ be the admissible control of the evader. Consider the sets $(i \in I, j \in J, v(\cdot) \in \mathcal{V})$

$$T_{ij}(v(\cdot)) = \{ t \ : \ E_{\frac{1}{\alpha}}(a_{lj} \hat{T}^\alpha, 1) - \int_0^t g_{lj}(\hat{T}, s) \overline{\lambda}^*_{lj}(v(s)) ds \leq 0 \}.$$

Next, let

$$t_{ij}^*(v(\cdot)) = \begin{cases} \inf\{t : t \in T_{ij}(v(\cdot))\} & \text{if } \ T_{ij}(v(\cdot)) \neq \emptyset, \\ +\infty & \text{if } \ T_{ij}(v(\cdot)) = \emptyset, \end{cases} \quad \beta_{lj}(t) = \begin{cases} \lambda_{lj}^*(v(t)), & t \in [0, t_{ij}^*(v(\cdot))], \\ 0, & t \in (t_{ij}^*(v(\cdot)), \hat{T}], \end{cases}$$

$$B_i(t) = \mathrm{diag}\,(\beta_{i1}(t), \ldots, \beta_{ik}(t)).$$

Define the controls of pursuers $P_i$, $i \in I$, assuming

$$u_i(t) = v(t) - B_i(t) z_i^0.$$

The solution of the Cauchy problem for system (2.1) has the form [9]

$$z_i(t) = E_{\frac{1}{\alpha}}(A_i t^\alpha, 1) z_i^0 + \int_0^t (t-s)^{\alpha-1} E_{\frac{1}{\alpha}}(A_i (t-s)^{\alpha-1}, \alpha)(u_i(s) - v(s)) ds.$$

Therefore,

$$z_{lj}(\hat{T}) = \Big( E_{\frac{1}{\alpha}}(a_{ij}\hat{T}^\alpha, 1) - \int_0^{\hat{T}} g_{ij}(\hat{T}, s) B_{ij}(s) ds \Big) z_{ij}^0 =$$

$$= \Big( E_{\frac{1}{\alpha}}(a_{ij}\hat{T}^\alpha, 1) - \int_0^{t_{ij}^*(v(\cdot))} g_{ij}(\hat{T}, s) \overline{\lambda}_{ij}^*(v(s)) ds \Big) z_{ij}^0.$$

It follows from the assumptions of the theorem and the definition of $B_i(t)$ ($i \in I$, $t \in [0, \infty)$) that there exists $\Lambda \in \Omega(m)$ such that $z_{lj}(\hat{T}) = 0$ for all $l \in \Lambda$ and $j \in J$, which implies that an $m$-fold capture occurs in the game $G(n+1)$. This proves the theorem. □

*Example 1.* Suppose that, in system (2.1), $k = 2$, $n = 1$, $m = 1$, $z_1^0 = (1, 2)$, $A_1$ is a zero matrix, $V = \{0\}$, and

$$U_1 = \{(u_1, u_2) : u_1{=}0, u_2{\in}[-1, 1]\} \cup \{(u_1, u_2) : u_2{=}0, u_1{\in}[-1, 1]\} \cup \{(u_1, u_2) : u_1{=}u_2{\in}[-1, 1]\}.$$

Then

$$\mathcal{M}_1^0(0) = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix}, \lambda \in [0, 1/2] \right\} \bigcup \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & 0 \end{pmatrix}, \lambda \in [0, 1] \right\} \bigcup \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda/2 \end{pmatrix}, \lambda \in [0, 1] \right\}.$$

Hence,

$$\sup_{L \in \mathcal{M}_1^0(0)} \min_j \lambda_{1j} = 1/2.$$

Consequently,

$$\overline{\mathcal{M}}_1(0) = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \right\}$$

and the extremal selector is $\overline{\lambda}_1^*(0) = \mathrm{diag}\,(1, 1/2)$. Therefore, $\hat{T} = \big(2\alpha\Gamma(\alpha)\big)^{1/\alpha}$, and the control of the pursuer $P_1$ has the form

$$u_1(t) = \begin{cases} (-1, -1), & t \in [0, T_1], \\ (0, -1), & t \in (T_1, \hat{T}], \end{cases}$$

where $T_1 = \hat{T} - \big(\alpha\Gamma(\alpha)\big)^{1/\alpha}$. Then [9]

$$z_1(\hat{T}) = z_1^0 + \frac{1}{\Gamma(\alpha)}\int\limits_0^{\hat{T}} (\hat{T}-s)^{\alpha-1} u_1(s)ds.$$

Therefore,

$$z_{11}(\hat{T}) = z_{11}^0 - \frac{1}{\Gamma(\alpha)}\int\limits_0^{T_1} (\hat{T}-s)^{\alpha-1} ds = 0, \quad z_{12}(\hat{T}) = z_{12}^0 - \frac{1}{\Gamma(\alpha)}\int\limits_0^{\hat{T}} (\hat{T}-s)^{\alpha-1} ds = 0.$$

Note that the use of scalar resolving functions, i.e., functions of the form

$$L = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix},$$

does not allow one to get the capture since, in this case, the condition $-Lz_0 \in U_1 - v$ is satisfied only for the zero matrix $L$.

We now present conditions on the game parameters under which the capture is guaranteed when scalar resolving functions are used.

**Assumption 7.** *In system (2.1), the matrices $A_i$ have the form $A_i = a_i E$, $a_i \leq 0$, $i \in I$, $E$ is an identity matrix, and*

$$\delta_0 = \inf_{v \in V} \max_{\Lambda \in \Omega(m)} \min_{l \in \Lambda} \mu_l(v) > 0,$$

*where $\mu_l(v) = \sup\{\mu \geq 0 \ : -\mu z_l^0 \in U_l - v\}$.*

**Theorem 3.** *Suppose that Assumptions 1 and 7 hold. Then an $m$-fold capture occurs in the game $G(n+1)$.*

P r o o f. It follows from the conditions of the theorem that the following equations hold for all $i \in I$, $j \in J$:

$$g_{ij}(t,s) = (t-s)^{\alpha-1} E_{\frac{1}{\alpha}}\big(a_i(t-s)^{\alpha}, \alpha\big) = g_i(t,s), \quad t \neq s, \quad g_{ij}(t,t) = 0,$$
$$E_{\frac{1}{\alpha}}(a_{ij}t^{\alpha}, 1) = E_{\frac{1}{\alpha}}(a_i t^{\alpha}, 1).$$

Therefore, it follows from Lemma 3 that there exists a time $T > 0$ such that, for every admissible function $v(\cdot) \in \mathcal{V}$, there is a set $\Lambda \in \Omega(m)$ such that the inequalities

$$E_{\frac{1}{\alpha}}(a_l T^{\alpha}, 1) - \int\limits_0^T g_l(T,s)\mu_l(v(s))ds \leq 0$$

hold for all $l \in \Lambda$. Define the number

$$T_0 = \inf\Big\{ t > 0 \ : \ \sup_{v(\cdot)} \min_{\Lambda \in \Omega(m)} \max_{l \in \Lambda} \Big( E_{\frac{1}{\alpha}}(a_l t^{\alpha}, 1) - \int\limits_0^t g_l(t,s)\mu_l(v(s))ds \Big) \leq 0 \Big\}.$$

Next, let $v(\cdot)$ be the admissible control of the evader:

$$\tau_l = \inf\Big\{t > 0 \ : \ E_{\frac{1}{\alpha}}(a_i T_0^\alpha, 1) - \int\limits_0^t g_l(T_0, s)\mu_l(v(s))ds \le 0\Big\}.$$

It follows from the above proof that there exists a set $\Lambda_0 \in \Omega(m)$ such that the inequalities $\tau_l \le T_0$ hold for all $l \in \Lambda_0$. Define the controls of pursuers $P_i$, $i \in I$, assuming

$$u_i(t) = \begin{cases} v(t) - \mu_i(v(t))z_i^0, & t \in [0, \tau_i], \\ v(t), & t \in [\tau_i, T_0]. \end{cases}$$

The solution of the Cauchy problem for system (2.1) has the form [9]

$$z_l(T_0) = \Big(E_{\frac{1}{\alpha}}(a_l T_0^\alpha, 1) - \int\limits_0^{T_0} g_l(T_0, s)\mu_l(v(s))ds\Big) z_l^0.$$

This equation and the definition of $\Lambda_0$ imply that $z_l(T_0) = 0$ for all $l \in \Lambda_0$. This proves the theorem. □

**Corollary 1.** *Suppose that, in system* (2.1)*, the matrices $A_i$ have the form $A_i = a_i E$, $a_i \le 0$, $i \in I$, $E$ is an identity matrix, $U_i = V$ for all $i \in I$, $V$ is a strictly convex compact set with a smooth boundary, and*

$$0 \in \bigcap_{\Lambda \in \Omega(n-m+1)} \operatorname{Int} \operatorname{co} \{z_l^0, l \in \Lambda\}, \tag{3.6}$$

*where* $\operatorname{Int} A$ *and* $\operatorname{co} A$ *denote the interior and the convex hull of the set $A$, respectively. Then an $m$-fold capture occurs in the game $G(n+1)$.*

Indeed, in this case, condition (3.6) implies that $\delta_0 > 0$ [30].

## 4. Conclusion

We obtained new sufficient conditions for multiple capture of the evader in the group pursuit problem with fractional derivatives. To solve the problem, we introduced matrix resolving functions.

### REFERENCES


1. Blagodatskikh A. I. Simultaneous multiple capture in a conflict-controlled process. *J. Appl. Math. Mech.*, 2013. Vol. 77, No. 3. P. 314–320. DOI: 10.1016/j.jappmathmech.2013.09.007
2. Blagodatskikh A. I., Petrov N. N. *Konfliktnoe vzaimodejstvie grupp upravlyaemyh ob"ektov* [Conflict Interaction of Groups of Controlled Objects]. Izhevsk: Udmurt State University, 2009. 266 p. (in Russian)
3. Blaquière A., Gérard F., Leitmann G. *Quantitative and Qualitative Differential Games.* New York, London: Academic Press, 1969. 172 p.
4. Bopardikar S. D., Suri S. $k$-Capture in multiagent pursuit evasion, or the lion and the hyenas. *Theoret. Comput. Sci.*, 2014. Vol. 522. P. 13–23. DOI: 10.1016/j.tcs.2013.12.001
5. Caputo M. Linear model of dissipation whose $Q$ is almost frequency independent-II. *Geophys. R. Astr. Soc.*, 1967. Vol. 13, No. 5. P. 529–539. DOI: 10.1111/j.1365-246X.1967.tb02303.x
6. Chikrii A. A. *Conflict-Controlled Processes.* Dordrecht: Springer, 1997. 404 p. DOI: 10.1007/978-94-017-1135-7

7. Chikrii A. A. Quasilinear controlled processes under conflict. *J. Math. Sci.* 1996. Vol. 80. No 1. P. 1489–1518.
8. Chikrii A. A., Chikrii G. Ts. Matrix resolving functions in game problems of dynamics. *Proc. Steklov Inst. Math.*, 2015. Vol. 291. Suppl. 1. P. 56–65. DOI: 10.1134/S0081543815090047
9. Chikrii A. A., Machikhin I. I. On an analogue of the Cauchy formula for linear systems of any fractional order. *Dopov. Nats. Akad. Nauk Ukr., Mat. Pryr. Tekh. Nauky.* 2007. No. 1. P. 50–55. (in Russian)
10. Chikrii A. A., Rappoport I. S. Method of resolving functions in the theory of conflict-controlled processes. *Cybern. Syst. Anal.*, 2012. Vol. 48, No. 4. P. 512–531. DOI: 10.1007/s10559-012-9430-y
11. Chikrii A. O., Chikrii G. Ts. Matrix resolving functions in dynamic games of approach. *Cybern. Syst. Anal.*, 2014. Vol. 50, No. 2. P. 201–217. DOI: 10.1007/s10559-014-9607-7
12. Chikriy A. A., Rappoport I. S. Measurable many-valued maps and their selectors in dynamic pursuit games. *J. Autom. Inform. Sci.*, 2006. Vol. 38. No. 1. P. 57–67. DOI: 10.1615/J Automat Inf Scien.v38.i1.60
13. Dzhrbashyan M. M. *Integral'nye preobrazovaniya i predstavleniya funkcij v kompleksnoj oblasti* [Integral Transforms and Representations of Functions in the Complex Domain]. Moscow: Nauka, 1966. 672 p. (in Russian)
14. Friedman A. *Differential Games.* New York: John Wiley & Sons. 1971. 350 p.
15. Gomoyunov M. I. Extremal shift to accompanying points in a positional differential game for a fractional-order system. *Proc. Steklov Inst. Math.*, 2020. Vol. 308. P. 83–105. DOI: 10.1134/S0081543820020078
16. Gorenflo R., Kilbas A. A., Mainardi F., Rogosin S. V. *Mittag–Leffler Functions, Related Topics and Applications.* Berlin Heidelberg: Springer-Verlag, 2014. 443 p. DOI: 10.1007/978-3-662-43930-2
17. Grigorenko N. L. Simple pursuit evasion game with a group of pursuers and one evsder. *Vestnik Moskov. Univ. Ser. XV Vychisl. Matematika i Kibernetika.* 1983. No. 1. P. 41–47. (in Russian)
18. Grigorenko N. L. *Matematicheskie metody upravleniya neskol'kimi dinamicheskimi processami* [Mathematical methods of control a few dynamic processes]. Moscow: Moscow State University, 1990. 197 p. (in Russian)
19. Hájek O. *Pursuit Games.* New York: Academic Press, 1975. 265 p.
20. Isaacs R. *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization.* New York: John Wiley and Sons, 1965. 384 p.
21. Krasovskii N. N., Subbotin A. I. *Pozicionnye differencial'nye igry* [Positional Differential Games]. Moscow: Nauka, 1974. 456 p. (in Russian)
22. Leitmann G. *Cooperative and Non-Cooperative Many Players Differential Games.* Wien: Springer-Verlag, 1974. 77 p. DOI: 10.1007/978-3-7091-2914-2
23. Machtakova A. I., Petrov N. N. Pursuit of Rigidly Coordinated Evaders in a Linear Problem with Fractional Derivatives, a Simple Matrix, and Phase Restrictions. In: *Lect. Notes Control Inf. Sci.: Proc. Stability and Control Processes. SCP 2020 (Smirnov N., Golovkina A. (eds.))*, 2022. P. 391–398. DOI: 10.1007/978-3-030-87966-2_43
24. Petrosyan L. A. *Differencial'nye igry presledovaniya* [Differential Games of Pursuit]. Leningrad: Leningrad University Press, 1977. 224 p. (in Russian)
25. Petrov N. N. Group Pursuit Problem in a Differential Game with Fractional Derivatives, State Constraints, and Simple Matrix. *Diff. Equat.*, 2019. Vol. 55. No. 6. P. 841–848. DOI: 10.1134/S0012266119060119
26. Petrov N. N. Matrix resolving functions in a linear problem of group pursuit with multiple capture. *Tr. Inst. Mat. Mekh. UrO RAN*, 2021. Vol. 27. No. 2. P. 185–196. DOI: 10.21538/0134-4889-2021-27-2-185-196
27. Petrov N. N. Multiple capture in a group pursuit problem with fractional derivatives and phase restrictions. *Mathematics*, 2021. Vol. 9. No. 11. Art. no. 1171. 12 p. DOI: 10.3390/math9111171
28. Petrov N. N. The problem of simple group pursuit with phase constraints in time scale. *Vestn. Udmurt. Univ. Mat. Mekh. Komp. Nauki*, 2020. Vol. 30. No. 2. P. 249–258. DOI: 10.35634/vm200208
29. Petrov N. N., Machtakova A. I. Capture of two coordinated evaders in a problem with fractional derivatives, phase restrictions and a simple matrix. *Izv. IMI UdGU*, 2020. Vol. 56. P. 50–62. DOI: 10.35634/2226-3594-2020-56-05
30. Petrov N. N., Solov'eva N. A. Multiple capture of given number of evaders in linear recurrent differential games. *J. Optim. Theory Appl.*, 2019. Vol. 182. No. 1. P. 417–429. DOI: 10.1007/s10957-019-01526-7

31. Petrov N. N., Solov'eva N. A. Multiple capture in Pontryagin's recurrent example with phase constraints. *Proc. Steklov Inst. Math.*, 2016. Vol. 293, Suppl. 1 P. 174–182. DOI: 10.1134/S0081543816050163
32. Petrov N. N., Solov'eva N. A. Multiple capture in Pontryagin's recurrent example. *Autom. Remote Control*, 2016. Vol. 77. No. 5. P. 855–861. DOI: 10.1134/S0005117916050088
33. Petrov N. N., Narmanov A. Ya. Multiple capture of a given number of evaders in a problem with fractional derivatives and a simple matrix. *Proc. Steklov Inst. Math.*, 2020. Vol. 309, Suppl. 1. P. S105–S115. DOI: 10.1134/S0081543820040136
34. Pollard H. The completely monotonic character of the Mittag–Leffler function $E_a(-x)$. *Bull. Amer. Math. Soc.*, 1948. Vol. 54, No. 12. P. 1115–1116. DOI: 10.1090/S0002-9904-1948-09132-7
35. Polovinkin E. S. *Mnogoznachnyj analiz i differencial'nye vklyucheniya* [Multivalued Analysis and Differential Inclusions]. Moscow: Fizmathlit, 2015. 524 p. (in Russian)
36. Pontryagin L. S. *Izbrannye nauchnye trudy* [Selected Scientific Works]. Vol. 2. Moscow: Nauka, 1988. 576 p. (in Russian)
37. Popov A. Yu., Sedletskii A. M. Distribution of roots of Mittag–Leffler functions. *J. Math. Sci.*, 2013. Vol. 190. No. 2. P. 209–409. DOI: 10.1007/s10958-013-1255-3
38. Pshenichnyi B. N. Simple pursuit by several objects. *Kibernetika*, 1976. No. 3. P. 145–146. (in Russian)
39. Rappoport J. S. Strategies of group approach in the method of resolving functions for quasilinear conflict-controlled processes. *Cybern. Syst. Anal.*, 2019. Vol. 55. No. 1. P. 128–140. DOI: 10.1007/s10559-019-00118-7
40. Subbotin A. I., Chentsov A. G. *Optimizaciya garantii v zadachah upravleniya* [Optimization of a Guarantee in Problems of Control]. Moscow: Nauka, 1981. 288 p. (in Russian)

**URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 90–116**

DOI: 10.15826/umj.2022.1.009

# A MARKOVIAN TWO COMMODITY QUEUEING–INVENTORY SYSTEM WITH COMPLIMENT ITEM AND CLASSICAL RETRIAL FACILITY

**M. Nithya**[†], **C. Sugapriya**[†], **S. Selvakumar**[††], **K. Jeganathan**[††,1], **T. Harikrishnan**[†††]

[†]Department of Mathematics, Queen Mary's College,
Tamil Nadu, Chennai – 600004, India

[††]Ramanujan Institute for Advanced Study in Mathematics, University of Madras,
Chepauk, Tamil Nadu, Chennai – 600005, India

[†††]Department of Mathematics, Guru Nanak College (Autonomous),
Tamil Nadu, Chennai – 600042, India

[1]kjeganathan84@gmail.com

**Abstract:** This paper explores the two-commodity (TC) inventory system in which commodities are classified as major and complementary items. The system allows a customer who has purchased a free product to conduct Bernoulli trials at will. Under the Bernoulli schedule, any entering customer will quickly enter an orbit of infinite capability during the stock-out time of the major item. The arrival of a retrial customer in the system follows a classical retrial policy. These two products' re-ordering process occurs under the $(s, Q)$ and instantaneous ordering policies for the major and complimentary items, respectively. A comprehensive analysis of the retrial queue, including the system's stability and the steady-state distribution of the retrial queue with the stock levels of two commodities, is carried out. The various system operations are measured under the stability condition. Finally, numerical evidence has shown the benefits of the proposed model under different random situations.

**Keywords:** Markov process, Compliment item, Infinite orbit, Waiting time.

## 1. Introduction

In an inventory business, substitution schemes play a key role in reducing customer losses. During the stock-out period of a demanded item, the flexible item can be used. For example, any product from dairy inventories has stochastic demand, and some of them can be used as substitutions for others. As sales and demand are directly proportional to the complement items, it is convenient to calculate the relevant product's future sales and demand. Besides, the business plans to launch a new product as a supplement to another product that will help the new product understand consumers' needs. Also, the company initially offers a supplement that eventually encourages a customer to make a purchase and continue to purchase the same product without a compliment in the future due to the market's product rivalry. Everyone in a mobile store, supermarket, car showroom, electrical equipment shop, etc., will experience these kinds of situations. We are motivated by these products to evaluate a TC inventory system, including complimentary items. The different varieties of this complimentary technique shall be defined as follows:

- Discounting is the crucial factor used to market tomorrow's cash flow and how the seller uses pricing to accomplish a particular market goal. More precisely, it deals with the customer's psychological reaction to those types of prices. For example, a textile showroom offers a 20%

discount or 30% discount. Online shopping platforms like Amazon, Flipkart and Myntra, etc., offer a discount of 20% on all domestic products, etc.

- The company increases the quantity of product to a certain percentage as an offer rather than giving a discount in price. For example, the Colgate company increases its quantity by 20% for the same price to attract more customers. A detergent, soap, or powder company increases a considerable percentage to sell the product effectively.

- Similarly, few firms provide buy one get one offer (may be the same item or a different item) to increase sales and profit. For example, buy one shirt and get the same shirt free, a mobile phone with a memory card, a laptop with a pen drive, etc.

So far, several researchers have studied two product inventory structures with a finite orbit. Since multi-commodity is more challenging than the single commodity system, most researchers have studied the single commodity queuing inventory model. For more details about a single commodity, one may refer [6, 11–15, 17, 20, 21, 26].

Nowadays, the modeling of multi-product inventory systems receives more attention. Many businesses and firms have gradually begun to use multi-commodity inventory systems in modern computer technologies. Kalpakam and Arivarignan [16] analyzed a joint reordering policy of a multi-item inventory such that the replenishment duration of every new order is zero. In [10], Goyal and Satir explored inventory control models where a range of items are jointly replenished. Under deterministic and stochastic demand conditions, the approaches available for determining the economic operating strategy for jointly replenished items are reviewed. Sivazlian [27] looks at the stationary properties of a multi-commodity inventory analysis periodically. The stochastic model assumes a dyadic replenishment strategy with proportional costs and a single set-up cost. The optimal operating costs for individual and joint ordering policies were compared in numerical scenarios.

Yadavalli et al. [36] assumed a Markovian arrival process (MAP) for a TC continuous review inventory system with three categories of arriving customers. Sivakumar et al. [28] suggested a TC perishable stochastic inventory method under continuous study at a service facility with a finite waiting hall. TC's are supposed to be interchangeable. That is, if either of the stocks is empty instantly, the other commodity may be used to meet the demand. Sivakumar et al. [29] assumed an inventory system with a loss in sales where each commodity's lifetime and lead time of a joint reorder of two commodities are all independent exponential distributions. Serife Ozkar and Umay Uzunoglu Kocerciteser investigated a TC queueing-inventory model with an individual ordering policy in which there are two types of customers: priority (Type-1) and ordinary (Type-2). Customers of Type-1 request commodity-1 only, while customers of Type-2 demand commodity-2. Each customer's arrival is subjected to an independent Poisson process of varying rates. Senthil Kumar [25] studied a TC inventory system that was subjected to discrete-time review, with each commodity's demand determined by an independent Bernoulli process. Anbazhagan and Arivarignan [2] elaborately studied independent Poisson demand processes of the TC continuous review inventory system. They made coordinated reorders whenever each commodity's stock level is less than or equal to its reorder level. TC inventory schemes were studied under separate ordering policies by Anbazhagan and Arivarignan [3, 4].

Sivakumar [31] initiated a retrial policy on the TC inventory system where the demand for any commodity is substitutable with others if the demanded commodity is not available. However, Yadavalli et al. [33] made the ordering quantity of each up to its maximum stock level. A TC inventory system with a single server was considered by Binitha Benny et al. [8]. The buffer capacity in this paper will be finite. Customers arrive in a Poisson process, and the demand for each type of commodity, or both types of commodities, is defined using specific probabilities.

Krishnamoorthy and Merlymol Joseph [19] discussed a continuous review of TC inventory problems with bulk demand. Assume that the model has both the commodities' probability of a demand equal to zero. Instantaneous replenishment and no shortages are permitted. Sivakumar et al. [30] studied a TC continuous review inventory system with a renewal of demand and ordering policy, a combination of policies referred to as the ordering of individual commodities and the ordering of both commodities jointly. Krishnamoorthy et al. [18] dealt with a TC inventory system with zero lead time in which the Poisson arrival of any customer is satisfied either with one or both commodities under a prefixed probability distribution. Yadavalli et al. [34] studied two commodity inventory systems in which the customers' demand patterns for each commodity follow Poisson and the demand for each commodity is fulfilled with another commodity with distinct probabilities.

Artaljeo et al. [7] began researching a classical retrial strategy in an inventory system. The reality of the classical nature of retrial policy in an inventory system has also been determined by Ushakumari [32]. A classical retrial queue with a single server with phase-type service facilities was addressed by Krishnamoorthy and Dhanya Shajin [22]. A classical retrial policy on an $(s, Q)$ inventory system with a finite customer source in which multi-servers provide homogeneous services was extensively studied by Yadavalli et al. [35]. Srinivas R. Chakravarthy et al. [9] studied a multi-server, infinite-orbit retrial method that considers a classic customer retrial pattern.

Anbazhagan and Jeganathan [5] addressed a two-commodity system with a complimenting item, where primary demand for the first commodity enters a finite orbit size $N$. Lakshmanan et al. [23] examine a two-commodity situation with a complement and frequent working holidays. Both commodities are independent of their ordering policies, and each customer orders service at a positive time. When the requested item is out of stock or the server is busy, each consumer is allowed to a given finite retrial orbit.

From the above studies, it is noticed that there is no work on TC inventory systems with complement items and an infinite orbit under the classical retrial policy. We consider that this concept is a gap in the inventory system until now. To fill such a gap in this field, we propose the model as a two-commodity inventory system involving an infinite orbit in which the orbital customers approach the system under the classical retrial policy. Sections 2, 3, 4, 5, 6, and 7 of the paper presented a model description, system analysis, waiting time analysis, measures of different system performances, cost analysis, and numerical illustration and conclusion, respectively.

### 1.1. Notation

We will use the following notation. Let the symbol $\mathbf{0}$ denotes the matrix with zero entries, let $\mathbf{e}$ be a convenient-sized column vector with one in each of the co-ordinates, a $I_n$ be an $n$th order identity matrix. Let

$$\begin{aligned}
\delta_{ij} &:= \begin{cases} 1 & \text{if } j = i, \\ 0, & \text{otherwise,} \end{cases} \\
\bar{\delta}_{ij} &:= 1 - \delta_{ij}, \\
H(x) &:= \begin{cases} 1 & \text{if} \quad x \geq 0, \\ 0, & \text{otherwise;} \end{cases} \\
[B_n]_{a,b} &:= \begin{cases} 1 & \text{if} \quad a = 2, \cdots, n, \quad b = a-1, \\ 0, & \text{otherwise;} \end{cases} \\
[C_{S_2}]_{a,b} &:= \begin{cases} 1 & \text{if} \quad a = S_2, \quad b = S_2, \\ 0, & \text{otherwise;} \end{cases} \\
H &:= \{1 \leq u \leq L, \quad 0 \leq v \leq S_1, \quad 1 \leq w \leq S_2\}.
\end{aligned}$$

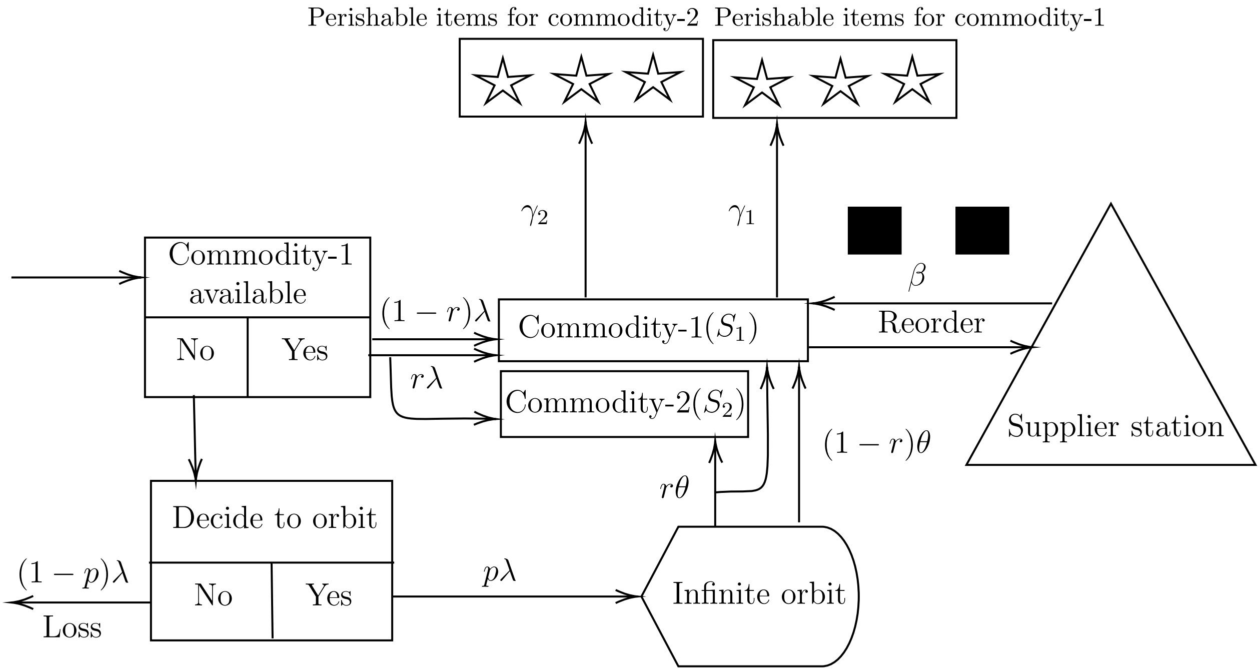



Figure 1. Graphical illustration of the model

## 2. Explanation of model

A system of two commodities is considered, in which one commodity is labelled as a main commodity and the other one as a complementary product. Any customer's timely arrival follows an exponential distribution at a rate of $\lambda$; $S_1$ and $S_2$ refer to the maximum stock level of main and complimentary products, respectively. Any order quantity of the main commodity follows $(s, Q)$ whereas a $(0, S_2)$ ordering policy is used for the complimentary product. The inter-arrival times between successive reorders of the main commodity follow an exponential distribution with rate $\beta$. However, the system receives any reorder of the complimentary product instantly when it is stocked out. Also, the main commodity and complimentary product are both independently perishable with rates $i\gamma_1$ and $j\gamma_2$ $(1 \leq i \leq S_1, 1 \leq j \leq S_2)$, respectively, and their deteriorating times follow exponential distributions where $i$ and $j$ denote their corresponding stock levels at that time. If there is no main commodity in the stock, any arrival customer enters into an infinite orbit at the rate of $\lambda$ with probability $p$. According to a purchase study, any new customer demands either the main commodity only or both commodities, at the rate of $\lambda$ with probability $1-r$ or $r$, respectively $(0 \leq r \leq 1)$. According to a classical retrial policy, any retrial customer demands the main commodity only or both commodities at the rate of $k\theta$ with probability $1-r$ or $r$ where $k$ $(k > 0)$ is the size of the retrial queue at that time. The proposed model of the two commodity queueing-inventory system is given in Fig. 1.

## 3. Analysis of the system

Consider a triplet $(A(t), B(t), C(t))$, where $A(t), B(t),$ and $C(t)$ denote the size of orbital customers, first commodity level, and second commodity level, respectively. According to the Markov property and the assumptions of the given model, the continuous-time discrete-state random process

$$X(t) = \{(A(t), B(t), C(t)), \quad t \geq 0\}$$

is said to be a Markov chain and its state space $D$ is defined as

$$D = \{(u, v, w)|\ u = 0, 1, 2, ...; \quad v = 0, 1, 2, ..., S_1; \quad w = 1, 2, ..., S_2\}.$$

### 3.1. Construction of infinitesimal generator matrix

The rate matrix of stochastic queueing inventory system (SQIS) $X(t)$ is given by

$$U = \begin{pmatrix} \mathbb{U}_{00} & \mathbb{U}_{01} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbb{U}_{10} & \mathbb{U}_{11} & \mathbb{U}_{01} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbb{U}_{20} & \mathbb{U}_{21} & \mathbb{U}_{01} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbb{U}_{N0} & \mathbb{U}_{N1} & \mathbb{U}_{01} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbb{U}_{(N+1)0} & \mathbb{U}_{(N+1)1} & \mathbb{U}_{01} & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{3.1}$$

where

$$[\mathbb{U}_{01}]_{(vw)(v\prime w\prime)} = \begin{cases} p\lambda, & v\prime = v, \quad v = 0, \quad w\prime = w, \quad w = 1, 2, ..., S_2; \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

for $u = 1, 2, 3, ...,$

$$[\mathbb{U}_{u0}]_{(vw)(v\prime w\prime)} = \begin{cases} ur\theta, & v\prime = v-1, \quad v = 1, 2, ..., S_1, \quad w\prime = S_2, \quad w = 1; \\ ur\theta & v\prime = v-1, \quad v = 1, 2, ..., S_1, \quad w\prime = w-1, \quad w = 2, 3, ..., S_2; \\ u(1-r)\theta, & v\prime = v-1, \quad v = 1, 2, ..., S_1, \quad w\prime = w, \quad w = 1, 2, ..., S_2; \\ 0, & \text{otherwise} \end{cases}$$

for $u = 0, 1, 2, ...,$

$$[\mathbb{U}_{uu}]_{(vw)(v\prime w\prime)} = \begin{cases} \beta, & v\prime = v+Q, & v = 0, 1, ..., s, \\ & w\prime = w, & w = 1, 2, ..., S_2; \\ \gamma_2, & v\prime = v, & v = 0, 1, 2, ..., S_1, \\ & w\prime = S_2, & w = 1; \\ w\gamma_2, & v\prime = v, & v = 0, 1, 2, ..., S_1, \\ & w\prime = w-1, & w = 2, 3, ..., S_2; \\ r\lambda, & v\prime = v-1, & v = 1, 2, ..., S_1, \\ & w\prime = S_2, & w = 1; \\ r\lambda, & v\prime = v-1, & v = 1, 2, ..., S_1, \\ & w\prime = w-1, & w = 2, 3, ..., S_2; \\ v\gamma_1 + (1-r)\lambda, & v\prime = v-1, & v = 1, 2, ..., S_1 \\ & w\prime = w, & w = 1, 2, ..., S_2; \\ -(p\lambda + v\gamma_1 + w\gamma_2 + u\theta + \beta), & v\prime = v, & v = 0, \\ & w\prime = w, & w = 1, 2, ..., S_2; \\ -(\lambda + v\gamma_1 + w\gamma_2 + u\theta + \beta), & v\prime = v, & v = 1, 2, ..., s, \\ & w\prime = w, & w = 1, 2, ..., S_2; \\ -(\lambda + v\gamma_1 + w\gamma_2 + u\theta), & v\prime = v, & v = s+1, s+2, ..., S_1, \\ & w\prime = w, & w = 1, 2, ..., S_2; \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

**Explanation of the above matrix structure.** By the assumption of the proposed model, the submatrices in (3.1) are square matrices of order $(S_1+1)S_2$. First, consider the matrix $\mathbb{U}_{01}$. It contains the submatrix $p\lambda I_{S2}$ whose entries are nothing but the transition rate of arrival $\lambda$ enter into the orbit under the Bernoulli schedule as follows:

$$(u,0,w) \xrightarrow{p\lambda} (u+1,0,w), \quad u=0,1,2,\dots, \quad w=1,2,\dots,S_2,$$

we get equation (3.2).

Next let us discuss the matrix $\mathbb{U}_{u0}$, where $u=1,2,\dots$. This matrix is equal to $[B_{(S_1+1)(S_2)}]A$, where $A$ is the submatrix of dimension $S_2$, diagonal entries of $A$ are the transition rate of retrial arrivals demand for the main commodity $\theta$ with probability $(1-r)$ and the remaining positive entries of $A$ defined as $[B_{S_2}+C_{S_2}]\theta$ with probability $r$ means transition rate of retrial arrivals demand for both commodity, each retrials based on the classical retrial policy enter into getting service, it gives equation (3.1). Then transition rates are as follows:

$$(u,v,1) \xrightarrow{ur\theta} (u-1,v-1,S_2), \quad u=1,2,\dots, \quad v=1,2,\dots,S_1,$$
$$(u,v,w) \xrightarrow{ur\theta} (u-1,v-1,w-1), \quad u=1,2,\dots, \quad v=0,1,2,\dots S_1, \quad w=2,\dots,S_2,$$
$$(u,v,w) \xrightarrow{u(1-r)\theta} (u-1,v-1,w), \quad u=1,2,\dots, \quad v=1,2,\dots S_1, \quad w=1,2,\dots,S_2.$$

Then consider the diagonal matrix $\mathbb{U}_{uu}$, where $u=0,1,2,\dots$, whose elements are of the transition rates as follows:

(1) $\beta$ denotes the rate of reorder transition which follows the $(s,Q)$ ordering policy,

$$(u,v,w) \xrightarrow{\beta} (u,v+Q,w), \quad u=0,1,2,\dots, \quad v=0,1,2,\dots,s, \quad w=1,2,\dots,S_2;$$

(2) $\lambda$ is the rate of arrival transition enter into the service for demanding both commodity with probability $r$,

$$(u,v,1) \xrightarrow{r\lambda} (u,v-1,S_2), \quad u=0,1,2,\dots, \quad v=1,2,\dots,S_1;$$
$$(u,v,w) \xrightarrow{r\lambda} (u,v-1,w-1), \quad u=0,1,2,\dots, \quad v=1,2,\dots,S_1, \quad w=2,\dots,S_2;$$

(3) $\lambda$ is the rate of arrival transition enter into the service for demanding the main commodity with probability $(1-r)$,

$$(u,v,w) \xrightarrow{(1-r)\lambda} (u,v-1,w), \quad u=0,1,2,\dots, \quad v=1,2,\dots S_1, \quad w=1,2,\dots,S_2;$$

(4) $\gamma_1$ indicates the perishable transition rates for the first commodity which depends on the number of present inventory level for the first commodity,

$$(u,v,w) \xrightarrow{v\gamma_1} (u,v-1,w), \quad u=0,1,2,\dots, \quad v=1,2,\dots S_1, \quad w=1,2,\dots,S_2;$$

(5) $\gamma_2$ indicates the perishable transition rates for the second commodity which depends on the number of present inventory level for the second commodity,

$$(u,v,w) \xrightarrow{w\gamma_2} (u,v,w-1), \quad u=0,1,2,\dots, \quad v=0,1,2,\dots,S_1, \quad w=2,\dots,S_2;$$
$$(u,v,1) \xrightarrow{\gamma_2} (u,v,S_2), \quad u=0,1,2,\dots, \quad v=0,1,2,\dots,S_1;$$

(6) then the diagonal element is filled by the sum of all the entries in the corresponding rows with the negative sign to satisfy the sum of all entries in each row yield zero; we obtain equation (3.3). Hence all the submatrices obtained through the respective transitions give the infinitesimal generator matrix $U$ as in equation (3.1).

### 3.2. Matrix geometric approximation

In this section, we find the steady-state probability vector $\Phi$ and the system's stability condition.

#### 3.2.1. Steady state analysis

Consider $K$ to be the truncation process's cutoff point for the matrix-geometric approximation. To find the steady-state of the considered system using Neuts–Rao truncation method, we assume that $\mathbb{U}_{u0} = \mathbb{U}_{K0}$ and $\mathbb{U}_{u1} = \mathbb{U}_{K1}$ for all $u \geq K$. The truncated system $X(t)$'s modified generator matrix is

$$\hat{U} = \begin{pmatrix} \mathbb{U}_{00} & \mathbb{U}_{01} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbb{U}_{10} & \mathbb{U}_{11} & \mathbb{U}_{01} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbb{U}_{20} & \mathbb{U}_{21} & \mathbb{U}_{01} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbb{U}_{K0} & \mathbb{U}_{K1} & \mathbb{U}_{01} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbb{U}_{K0} & \mathbb{U}_{K1} & \mathbb{U}_{01} & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

**Theorem 1.** *The steady-state probability vector $\Phi$ corresponds to the generator matrix $\mathbb{U}_K$, where $\mathbb{U}_K = \mathbb{U}_{K0} + \mathbb{U}_{K1} + \mathbb{U}_{01}$ is given by*

$$\Phi^{(v)} = \Phi^{(Q)} e_v, \quad v = 0, 1, \ldots, S_1, \tag{3.4}$$

*where*

$$e_v = \begin{cases} (-1)^{Q-v} F_Q E_{Q-1}^{-1} F_{Q-1} \ldots F_{v+1} E_v^{-1}, & v = 0, 1, \ldots, Q-1, \\ I, & v = Q, \\ (-1)^{2Q-v+1} \sum\limits_{v\prime=0}^{S-v} [(F_Q E_{Q-1}^{-1} F_{Q-1} \ldots F_{s+1-v\prime} E_{s-v\prime}^{-1}) \times & \\ \qquad G E_{S-v\prime}^{-1} (F_{S-v\prime} E_{S-v\prime-1}^{-1} F_{S-v\prime-1} \ldots F_{v+1} E_v^{-1})], & v = Q+1, Q+2, \ldots, S_1, \end{cases}$$

*and $\Phi^{(Q)}$ is obtained by solving*

$$\Phi^{(Q)} \Big[ (-1)^Q \sum_{v\prime=0}^{s-1} \left[ (F_Q E_{Q-1}^{-1} F_{Q-1} \ldots F_{s+1-v\prime} E_{s-v\prime}^{-1}) G E_{S-v\prime}^{-1} (F_{S-v\prime} E_{S-v\prime-1}^{-1} F_{S-v\prime-1} \ldots F_{v+1} E_v^{-1}) \right]$$
$$F_{Q+1} + E_Q + (-1)^Q F_Q E_{Q-1}^{-1} F_{Q-1} \ldots F_1 E_0^{-1} G \Big] = \mathbf{0}$$

*and*

$$\sum_{v\prime=1}^{S_1} \Phi^{(v)} \mathbf{e} = 1.$$

P r o o f. We have

$$\Phi \mathbb{U}_K = \mathbf{0} \quad \text{and} \quad \Phi \mathbf{e} = 1,$$

where

$$[U_K]_{vv\prime} = \begin{cases} E_v, & v\prime = v, \quad v = 0, 1, 2, \ldots, S_1; \\ F_v, & v\prime = v-1, \quad v = 1, 2, \ldots, S_1; \\ G, & v\prime = v+Q, \quad v = 0, 1, 2, \ldots, s; \\ 0, & \textit{otherwise}. \end{cases}$$

The first equation of the above framework yields the following set of equations:

$$\begin{gathered}\Phi^{v+1}F_{v+1}+\Phi^{v}E_{v}=\mathbf{0}, \quad v=0,1,\ldots,Q-1,\\ \Phi^{v+1}F_{v+1}+\Phi^{v}E_{v}+\Phi^{v-Q}G=\mathbf{0}, \quad v=Q,Q+1,\ldots,S_1-1,\\ \Phi^{v}E_{v}+\Phi^{v-Q}G=\mathbf{0}, \quad v=S_1.\end{gathered} \tag{3.5}$$

We get equation (3.4) by recursively solving the set of equations (3.5) and using the normalising condition. □

Next, the stability condition in which the framework is stable is then determined.

**Theorem 2.** *The system's stability condition at the truncation point $K$ is given by*

$$r_1 p\lambda \boldsymbol{e} < r_2 K\theta \boldsymbol{e},$$

*where*

$$r_1=\sum_{w=1}^{S_2}\Phi^{(0,w)}, \quad r_2=\sum_{v=1}^{S_1}\sum_{w=1}^{S_2}\Phi^{(v,w)}.$$

P r o o f. From the well known-result of Neuts [24] on the positive recurrence of $\mathbb{U}_K$, we have

$$\Phi^{(K)}\mathbb{U}_{01}\mathbf{e} < \Phi^{(K)}\mathbb{U}_{K0}\mathbf{e},$$

and, by exploiting the structure of the matrices $\mathbb{U}_{01}$ and $\mathbb{U}_{K0}$, we get, for $v=0,1,2,\ldots,S_1$ and $w=1,2,\ldots,S_2$,

$$\Phi^{K}(v,w)\mathbb{U}_{01}\mathbf{e} < \Phi^{K}(v,w)\mathbb{U}_{K0}\mathbf{e}.$$

First,

$$[\Phi^{K}(0),\Phi^{K}(1),\ldots,\Phi^{K}(S_1)]\mathbb{U}_{01}\mathbf{e} < [\Phi^{K}(0),\Phi^{K}(1),\ldots,\Phi^{K}(S_1)]\mathbb{U}_{K0}\mathbf{e},$$

where $\Phi^{K}(i)=\phi^{K}(v,w)$ and

$$[\Phi^{K}(0)\lambda I_{S_2},\Phi^{K}(1)\mathbf{0},\ldots,\Phi^{K}(S)\mathbf{0}]\mathbf{e} < [\Phi^{K}(0)A,\Phi^{K}(1)A,\ldots,\Phi^{K}(S)A]\mathbf{e}.$$

The left-hand side becomes

$$\Phi^{K}(0)\lambda I_{S_2}=\Phi^{K}(0,w)p\lambda.$$

On the other hand, due to the structure of $A$, the right-hand side becomes

$$\Phi^{K}(v)A=[\Phi^{K}(v,w),\Phi^{K}(v,w),\ldots,\Phi^{K}(v,w)]K\theta.$$

Therefore, the last inequality becomes

$$\sum_{w=1}^{S_2}\Phi^{K}(0,w)p\lambda\mathbf{e} < \sum_{v=1}^{S_1}\sum_{w=1}^{S_2}\Phi^{K}(v,w)K\theta\mathbf{e}.$$

Hence,

$$r_1 p\lambda\mathbf{e} < r_2 K\theta\mathbf{e},$$

where

$$r_1=\sum_{w=1}^{S_2}\Phi^{(0,w)}, \quad r_2=\sum_{v=1}^{S_1}\sum_{w=1}^{S_2}\Phi^{(v,w)},$$

as desired. □

### 3.3. Stationary probability vector

The regularity of the Markov process $X(t)$ with the state space $D$ can be seen from the structure of the rate matrix $U$ and Theorem 2. Henceforth, the limiting probability distribution defined as

$$\boldsymbol{\chi}^{(u,v,w)} = \lim_{t\to\infty} \Pr\left[A(t)=u,\ B(t)=v,\ C(t)=w \mid A(0),B(0),C(0)\right]$$

exists and is independent of the initial state. Let $\boldsymbol{\chi} = \left(\chi^{(0)}, \chi^{(1)}, \dots\right)$ satisfy

$$\boldsymbol{\chi}U = \mathbf{0}, \quad \boldsymbol{\chi}\mathbf{e} = 1.$$

We can partition the vector $\boldsymbol{\chi}^{(u)}$ as

$$\boldsymbol{\chi}^{(u)} = \left(\chi^{(u,0)}, \chi^{(u,1)}, \dots, \chi^{(u,S_1)}\right), \quad u \geq 0,$$

and

$$\boldsymbol{\chi}^{(u,v)} = \left(\chi^{(u,v,1)}, \chi^{(u,v,2)}, \dots, \chi^{(u,v,S_2)}\right), \quad u \geq 0, \quad 0 \leq v \leq S_1.$$

#### 3.3.1. Computation of the matrix $R$

**Theorem 3.** *Utilizing the vector $\boldsymbol{\chi}$ and the specific structure of $U$, $R$ can be determined by*

$$R^2\mathbb{U}_{K0} + R\mathbb{U}_{K1} + \mathbb{U}_{01} = \mathbf{0},$$

*where $R$ is the minimal nonnegative solution of the matrix quadratic equation (MNSMQE).*

P r o o f. Since the Markov process is a regular, the stationary probability distribution exists and is given by

$$\boldsymbol{\chi}U = \mathbf{0}, \quad \boldsymbol{\chi}\mathbf{e} = 1.$$

In order to express the solution in a recursive form, we assume that

$$\boldsymbol{\chi}^{(u)} = \boldsymbol{\chi}^{(K)}R^u, \quad u \geq K,$$

where the spectrum of $R$ is less than 1, which is ensured by the stability condition. Then, we get

$$\boldsymbol{\chi}^{(m)}(R^2\mathbb{U}_{K0} + R\mathbb{U}_{K1} + \mathbb{U}_{01}) = 0, \quad m = K, K+1, K+2, \dots .$$

Since the above equation is true for all $m = K, K+1, N+2, \dots$, we get

$$(R^2\mathbb{U}_{K0} + R\mathbb{U}_{K1} + \mathbb{U}_{01}) = 0, \quad m = K, K+1, K+2, \dots .$$

Then $R$ is the MNSMQE, and let us assume that the matrix $R$ is of the form

$$R = \begin{pmatrix} R_{00} & R_{01} & \cdots & R_{0S_1} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}.$$

This matrix $R$ has only $S_2$ nonzero rows of dimension $(S_1+1)(S_2)$. The structure of the block matrix $R_{0v\prime}$, where $v\prime \in \{0,1,\cdots,S_1\}$, is of the form

$$R_{0v\prime} = \begin{pmatrix} l_{0v\prime}^{11} & l_{0v\prime}^{12} & l_{0v\prime}^{13} & l_{0v\prime}^{14} & l_{0v\prime}^{15} & \cdots & l_{0v\prime}^{1S_2} \\ l_{0v\prime}^{21} & l_{0v\prime}^{22} & l_{0v\prime}^{23} & l_{0v\prime}^{24} & l_{0v\prime}^{25} & \cdots & l_{0v\prime}^{2S_2} \\ l_{0v\prime}^{31} & l_{0v\prime}^{32} & l_{0v\prime}^{33} & l_{0v\prime}^{34} & l_{0v\prime}^{35} & \cdots & l_{0v\prime}^{3S_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{0v\prime}^{S_21} & l_{0v\prime}^{S_22} & l_{0v\prime}^{S_23} & l_{0v\prime}^{S_24} & l_{0v\prime}^{S_25} & \cdots & l_{0v\prime}^{S_2S_2} \end{pmatrix}.$$

This is also a square matrix of dimension $S_2$. Now, exploiting the coefficient matrices $\mathbb{U}_{K0}, \mathbb{U}_{K1}$, and $\mathbb{U}_{01}$ with $R^2$ and $R$ equal to $\mathbf{0}$, we obtain a system of $S_2$-dimensional equations as follows (for $v=0$):

- for $v\prime = 0,1,2,\ldots,S_1-1,\ \ w=1,2,\ldots,S_2$, and $w\prime = 1,2,\ldots,S_2-1$,

$$\Big(\sum_{x=1}^{S_2} l_{vv\prime}^{wx} l_{v(v\prime+1)}^{xw\prime} K(1-r)\theta + \sum_{x=1}^{S_2} l_{vv\prime}^{wx} l_{v(v\prime+1)}^{x(w\prime+1)} Kr\theta + l_{vv\prime}^{ww\prime} C_{w\prime}^{(v\prime)} + l_{vv\prime}^{w(w\prime+1)}(w\prime+1)\gamma_2$$
$$+ l_{v(v\prime+1)}^{ww\prime}((v\prime+1)\gamma_1 + (1-r)\lambda) + l_{v(v\prime+1)}^{w(w\prime+1)} r\lambda + \delta_{v\prime 0}\delta_{ww\prime}p\lambda + H(s-v\prime) l_{v(v\prime-s)}^{ww}\beta\Big) = \mathbf{0};$$

- for $v\prime = 0,1,2,\cdots,S_1-1,\ \ w=1,2,\cdots,S_2$, and $w\prime = S_2$,

$$\Big(\sum_{x=1}^{S_2} l_{vv\prime}^{wx} l_{v(v\prime+1)}^{xw\prime} K(1-r)\theta + \sum_{x=1}^{S_2} l_{vv\prime}^{wx} l_{v(v\prime+1)}^{x1} Kr\theta + l_{vv\prime}^{ww\prime} C_{w\prime}^{(v\prime)} + l_{vv\prime}^{w1}\gamma_2 + l_{v(v\prime+1)}^{ww\prime}((v\prime+1)\gamma_1$$
$$+(1-r)\lambda) + l_{v(v\prime+1)}^{w1} r\lambda + \delta_{v\prime 0}\delta_{ww\prime}p\lambda + H(s-v\prime) l_{v(v\prime-s)}^{v_1v_1}\beta\Big) = \mathbf{0};$$

- for $v\prime = S_1,\ \ w=1,2,\cdots,S_2$, and $w\prime = 1,2,\cdots,S_2-1$,

$$\big(l_{vv\prime}^{ww\prime} C_{w\prime}^{(v\prime)} + l_{vv\prime}^{w(w\prime+1)}(w\prime+1)\gamma_2 + l_{vs}^{ww}\beta\big) = \mathbf{0};$$

- for $v\prime = S_1,\ \ w=1,2,\cdots,S_2$, and $w\prime = S_2$,

$$\big(l_{vv\prime}^{ww\prime} C_{w\prime}^{(v\prime)} + l_{vv\prime}^{w1}\gamma_2 + l_{vs}^{ww}\beta\big) = \mathbf{0}.$$

After solving all such equations, one can obtain the elements of the matrix $R$. In this case, $C_{w\prime}^{(v\prime)}$ are the diagonal elements of the $v\prime$th diagonal submatrix of $\mathbb{U}_{K1}$. □

**Theorem 4.** *The vector* $\boldsymbol{\chi}$ *can be determined by*

$$\chi^{(i+K-1)} = \chi^{(K-1)} R^i, \quad i \geq 0,$$

*due to the special structure of* $U$*, the fact that* $R$ *is the MNSMQE*

$$R^2\mathbb{U}_{K0} + R\mathbb{U}_{K1} + \mathbb{U}_{01} = \mathbf{0},$$

*and the vector* $\chi^{(i)}$*,* $i \geq 0$*,*

$$\chi^{(i)} = \begin{cases} \sigma X^{(0)} \prod_{j=i+1}^{K} \mathbb{U}_{j0}(-\mathbb{U}_{j-1}), & 0 \leq i \leq K-1, \\ \sigma X^{(0)} R^{(i-K)}, & i \geq K, \end{cases} \tag{3.6}$$

*where*

$$\sigma = [1 + X^{(0)} \sum_{i=0}^{K-1} \prod_{j=i+1}^{K} \mathbb{U}_{j0}(-\mathbb{U}_{j-1})\mathbf{e}]^{-1}, \tag{3.7}$$

*and $X(0)$ can be computed by using the normalising condition*

$$X^{(0)}(I - R)^{-1}\mathbf{e} = 1.$$

P r o o f. The subvector $\chi^{(0)}, \chi^{(1)}, \ldots, \chi^{(K-1)}$ and the block partitioned matrix of $\hat{U}$ give the set of equations $(1 \le i \le K-1)$

$$\begin{aligned} \chi^{(0)}\mathbb{U}_{00} + \chi^{(1)}\mathbb{U}_{10} &= \mathbf{0}, \\ \chi^{(i-1)}\mathbb{U}_{01} + \chi^{(i)}\mathbb{U}_{i1} + \chi^{(i+1)}\mathbb{U}_{(i+1)0} &= \mathbf{0}, \\ \chi^{(K-2)}\mathbb{U}_{01} + \chi^{(K-1)}(\mathbb{U}_{(K-1)1} + R\mathbb{U}_{K0}) &= \mathbf{0}. \end{aligned} \tag{3.8}$$

Using (3.8) repeatedly, we find

$$\chi^{(0)} = \chi^{(1)}\mathbb{U}_{10}(-\mathbb{U}_0)^{-1}$$

and

$$\chi^{(1)} = \chi^{(2)}\mathbb{U}_{20}(-\mathbb{U}_1)^{-1},$$

where

$$\mathbb{U}_1 = (\mathbb{U}_{11} + \mathbb{U}_{10}(-\mathbb{U}_0)^{-1}\mathbb{U}_{01})\mathbb{U}_0 = \mathbb{U}_{00}.$$

Next,

$$\chi^{(2)} = \chi^{(3)}\mathbb{U}_{30}(-\mathbb{U}_2)^{-1},$$

where

$$\mathbb{U}_2 = (\mathbb{U}_{21} + \mathbb{U}_{20}(-\mathbb{U}_1)^{-1}\mathbb{U}_{01}).$$

On continuing this procedure up to $K-1$ times, we get

$$\chi^{(i)} = \chi^{(i+1)}\mathbb{U}_{(i+1)0}(-\mathbb{U}_i)^{-1}, \quad 0 \le i \le K-1, \tag{3.9}$$

where

$$\mathbb{U}_i = \begin{cases} \mathbb{U}_{i0}, & i = 0, \\ (\mathbb{U}_{i1} - \mathbb{U}_{i0}(-\mathbb{U}_{i-1})^{-1}\mathbb{U}_{01}), & 1 \le i \le K. \end{cases}$$

For the next, we use the block Gaussian elimination method to find the vectors $(\chi^{(K)}, \chi^{(K+1)}, \chi^{(K+2)} \ldots)$. The nonboundary states subvector $(\chi^{(K)}, \chi^{(K+1)}, \chi^{(K+2)}, \ldots)$ satisfies the relation

$$(\chi^{(K)}, \chi^{(K+1)}, \chi^{(K+2)} \ldots) \begin{pmatrix} \mathbb{U}_K & \mathbb{U}_{01} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbb{U}_{K0} & \mathbb{U}_{K1} & \mathbb{U}_{01} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbb{U}_{K0} & \mathbb{U}_{K1} & \mathbb{U}_{01} & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \mathbf{0}. \tag{3.10}$$

Let us assume that

$$\sigma = \sum_{i=K}^{\infty} \chi^{(i)}\mathbf{e}, \quad X^{(i)} = \sigma^{-1}\chi^{(K+i)}, \quad i \ge 0.$$

From (3.10), we get

$$\chi^{(K)}\mathbb{U}_K + \chi^{(K+1)}\mathbb{U}_{K0} = \mathbf{0}, \quad \chi^{(K+i)} = \chi^{(K+i-1)}R, \quad i \geq 1,$$

which implies that

$$X^{(0)}\mathbb{U}_K + X^{(1)}\mathbb{U}_{K0} = \mathbf{0} \quad X^{(i)} = X^{(i-1)}R, \quad i \geq 1,$$

that is,

$$X^{(0)}[\mathbb{U}_K + R\mathbb{U}_{K0}] = \mathbf{0}. \tag{3.11}$$

Since

$$\sum_{i=0}^{\infty} X^{(i)}\mathbf{e} = 1,$$

we have

$$X^{(0)}(I - R)^{-1}\mathbf{e} = 1. \tag{3.12}$$

As a result, $X^{(0)}$ is the only solution to equations (3.11) and (3.12). Hence,

$$\chi^{(i)} = \sigma X^{(0)} R^{(i-K)}, \quad i \geq K. \tag{3.13}$$

Again, by (3.9) and (3.13), we get (3.6). Using

$$\sum_{i=0}^{\infty} \chi^{(i)}\mathbf{e} = 1$$

and (3.6), we get

$$\sigma X^{(0)} \sum_{i=0}^{K-1} \prod_{j=i+1}^{K} \mathbb{U}_{j0}(-\mathbb{U}_{j-1})\mathbf{e} + \sigma X^{(0)} \sum_{K}^{\infty} R^{(i-K)}\mathbf{e} = 1,$$

which gives $\sigma$ as in (3.7). □

## 4. Waiting time analysis

Waiting time ($WT$) is the time interval between an epoch when a demand approaches the orbit and the moment when its time of operation completion occurs. Using the Laplace–Stieltjes transform (LST), we look at the $WT$ of demand in orbit. To find the orbital demand's waiting period, we naturally limit the orbit size to a finite size. The continuous random variable $W_o$ represents the waiting time distribution of an orbit customer.

### 4.1. $WT$ of orbital customers

**Theorem 5.** *The probability that an orbital demand will not wait in the orbit is determined as follows*:

$$P\{W_o = 0\} = 1 - \eta_o, \tag{4.1}$$

*where*

$$\eta_o = \sum_{u=1}^{L-1} \sum_{w=1}^{S_2} \chi^{(u,0,w)}.$$

P r o o f. Since the zero and positive waiting time probability sum is 1, we have

$$P\{W_o = 0\} + P\{W_o > 0\} = 1. \tag{4.2}$$

Clearly, the probability of positive $WT$ of orbital demand can be determined as

$$P\{W_o > 0\} = \sum_{u=1}^{L-1} \sum_{w=1}^{S_2} \chi^{(u,0,w)}. \tag{4.3}$$

Equation (4.3) can be found easily using Theorem 4. Substituting it into equation (4.2), we get the stated result as desired in (4.1). □

To enable the distribution of $W_o$, we define some complimentary variables. Suppose that the queueing inventory system is at state $(u, v, w)$, $u > 0$ at an arbitrary time $t$, and

(1) $W_o(u, v, w)$ is the time until chosen demand becomes satisfied;

(2) the LST of $W_o(u, v, w)$ is ${}^*W_o(u, v, w)(y)$ and we denote $W_o$ by ${}^*W_o(y)$;

(3) ${}^*W_o(y) = E[e^{yW_o}]$ is the LST of unconditional waiting time (UWT);

(4) ${}^*W_o(u, v, w)(y) = E[e^{yW_o(u,v,w)}]$ is the LST of conditional waiting time (CWT).

**Theorem 6.** *The LST*

$$\left\{{}^*W_o(u, v, w)(y),\ (u, v, w) \in H^*,\ \textit{where } H^* = H \cup \{*\}\right\}$$

*satisfies the system*

$$\begin{gathered} Z_o(y)^*W_o(y) = -\theta \mathbf{e}(u, v, w), (u, v, w) \in H, \\ Z_o(y) = (P - yI), \end{gathered} \tag{4.4}$$

*the matrix $P$ is derived from $U$ by deleting the state $(0, v, w)$, $0 \le v \le S_1$, $1 \le w \le S_2$, $\{*\}$ is the absorbing state, and the absorption appears if the orbital demand finds the positive commodities.*

P r o o f. To analyse the CWT, we apply the first step analysis as follows:

$$\begin{aligned} {}^*W_o(u, 0, w)(y) = \frac{p\lambda}{a}\ {}^*W_o(u + 1, 0, w)(y) + \delta_{w1}\frac{w\gamma_2}{a}\ {}^*W_o(u, 0, S_2)(y) \\ +\bar{\delta}_{w1}\frac{w\gamma_2}{a}\ {}^*W_o(u, 0, w - 1)(y) + \frac{\beta}{a}\ {}^*W_o(u, Q, w)(y) \end{aligned} \tag{4.5}$$

for

$$1 \le u \le L, \quad v = 0, \quad 1 \le w \le S_2$$

and

$$a = (y + p\lambda + \delta_{w1}w\gamma_2 + \bar{\delta}_{w1}w\gamma_2 + \beta).$$

Next, for

$$1 \le u \le L, \quad 1 \le v \le S_1, \quad 1 \le w \le S_2$$

and

$$b = \big(y + \delta_{w1}r\lambda + (1-r)\lambda + H(s-v)\beta + v\gamma_1 + \delta_{w1}w\gamma_2 + \bar{\delta}_{w1}w\gamma_2 + \delta_{w1}(u-1)r\theta + \bar{\delta}_{w1}(u-1)r\theta\big),$$

we get

$$\begin{aligned}
{}^{*}W_o(u,v,w)(y) &= \delta_{w1}\frac{r\lambda}{b}\,{}^{*}W_o(u,v-1,S_2)(y) + \frac{(1-r)\lambda}{b}\,{}^{*}W_o(u,v-1,w)(y)\\
&+\frac{H(s-v)\beta}{b}\,{}^{*}W_o(u,v+Q,w)(y) + \frac{v\gamma_1}{b}\,{}^{*}W_o(u,v-1,w)(y) + \delta_{w1}\frac{w\gamma_2}{b}\,{}^{*}W_o(u,v,S_2)(y)\\
&+\bar{\delta}_{w1}\frac{w\gamma_2}{b}\,{}^{*}W_o(u,v,w-1)(y) + \delta_{w1}\frac{(u-1)r\theta}{b}\,{}^{*}W_o(u-1,v-1,S_2)(y)\\
&+\bar{\delta}_{w1}\frac{(u-1)r\theta}{b}\,{}^{*}W_o(u-1,v-1,w-1)(y) + \frac{\theta}{b}.
\end{aligned} \tag{4.6}$$

From equations (4.5) and (4.6), we attain a coefficient matrix of the unknowns as a block tridiagonal, which yields the stated result as in (4.4). □

**Theorem 7.** *The nth moments of conditional waiting time is given by*

$$Z_o(y)\frac{d^{n+1}}{dy^{n+1}}\,{}^{*}W_o(y) - (n+1)\frac{d^{n+1}}{dy^{n+1}}\,{}^{*}W_o(y) = 0$$

*and*

$$\frac{d^{n+1}}{dy^{n+1}}\,{}^{*}W_o(y)|_{y=0} = E[W_o^{n+1}(u,v,w)(y)], (u,v,w) \in H^{*}.$$

P r o o f. Using linear equations obtained in Theorem 6, we get a recursive algorithm for finding a conditional and unconditional waiting times. Now, differentiating equations (4.5) and (4.6) $(n+1)$ times and setting $y=0$, we obtain

$$\begin{aligned}
E[W_o^{n+1}(u,0,w)] &= \frac{p\lambda}{a}E[W_o^{n+1}(u+1,0,w)] + \delta_{w1}\frac{w\gamma_2}{a}E[W_o^{n+1}(u,0,S_2)]\\
&+\bar{\delta}_{w1}\frac{w\gamma_2}{a}E[W_o^{n+1}(u,0,w-1)] + \frac{\beta}{a}E[W_o^{n+1}(u,Q,w)]
\end{aligned} \tag{4.7}$$

for

$$1 \le u \le L, \quad v=0, \quad 1 \le w \le S_2$$

and

$$a = (y + p\lambda + \delta_{w1}w\gamma_2 + \bar{\delta}_{w1}w\gamma_2 + \beta).$$

Next, for

$$1 \le u \le L, \quad 1 \le v \le S_1, \quad 1 \le w \le S_2$$

and

$$b=\big(y+\delta_{w1}r\lambda+(1-r)\lambda+H(s-v)\beta+v\gamma_1+\delta_{w1}w\gamma_2+\bar{\delta}_{w1}w\gamma_2+\delta_{w1}(u-1)r\theta+\bar{\delta}_{w1}(u-1)r\theta\big),$$

we get

$$\begin{aligned}
E[W_o^{n+1}(u,v,w)] &= \delta_{w1}\frac{r\lambda}{b}E[W_o^{n+1}(u,v-1,S_2)] + \frac{(1-r)\lambda}{b}E[W_o^{n+1}(u,v-1,w)]\\
&+\frac{H(s-v)\beta}{b}E[W_o^{n+1}(u,v+Q,w)] + \frac{v\gamma_1}{b}E[W_o^{n+1}(u,v-1,w)] + \delta_{w1}\frac{w\gamma_2}{b}E[W_o^{n+1}(u,v,S_2)]\\
&+\bar{\delta}_{w1}\frac{w\gamma_2}{b}E[W_o^{n+1}(u,v,w-1)] + \delta_{w1}\frac{(u-1)r\theta}{b}E[W_o^{n+1}(u-1,v-1,S_2)]\\
&+\bar{\delta}_{w1}\frac{(u-1)r\theta}{b}E[W_o^{n+1}(u-1,v-1,w-1)] + \frac{\theta}{b}.
\end{aligned} \tag{4.8}$$

With reference to equations (4.7) and (4.8), one can determine the unknowns $E[W_p^{n+1}(u,v,w,x)]$ in terms of moments of one order less. Setting $n=0$, we obtain the desired moments of particular order in an algorithmic way. □

**Theorem 8.** *The LST of UWT of orbital demand is given by*

$$^{*}W_o(y) = 1 - \eta_o + \eta_o\, {}^{*}W_o(u+1,v,w)(y). \tag{4.9}$$

P r o o f. Using Poisson arrival see time averages (PASTA) property, one can obtain the LST of $W_o$ as follows:

$$^{*}W_o(y) = \chi^{(i)}\, {}^{*}W_o(u,v,w)(y), \quad 0 \le u \le L, \quad 0 \le v \le S_1, \quad 0 \le w \le S_2. \tag{4.10}$$

Using the expressions (4.10), we get the stated result. Considering the Euler and Post–Widder algorithms in [1] for the numerical inversion of (4.9), we obtain the desired result. □

**Theorem 9.** *The nth moment of UWT, by the above theorem, is given by*

$$E[W_o^n] = \delta_{0n} + (1-\delta_{0n}) \sum_{u=0}^{L-1} \sum_{v=0}^{S_1} \sum_{w=1}^{S_2} \chi^{(u,v,w)} E[W_o^n(u+1,v,w)]. \tag{4.11}$$

P r o o f. To determine the moments of $W_o$, we differentiate the equation in Theorem 8 $n$ times and calculate at $y=0$ to obtain the desired result, which gives the $n$th moment of UWT in terms of the CWT of the same order. □

**Theorem 10.** *The expected waiting time of an orbital demand is defined by*

$$E[W_o] = \sum_{u=0}^{L-1} \sum_{v=0}^{S_1} \sum_{w=1}^{S_2} \chi^{(u,v,w)} E[W_o(u+1,v,w)]. \tag{4.12}$$

P r o o f. Using equation (4.11) in Theorem 9 and substituting $n=1$, we get the desired result as in (4.12). □

## 5. Measures of various performances of the system

In this section, the following measures of corresponding performance are used to obtain the expected total cost under the steady state probability vector.

1. $EIC1$ denotes the expected level of commodity 1. Using the steady state probability vector $\boldsymbol{\chi}$, we define $EIC1$ as

$$EIC1 = \sum_{u=0}^{\infty} \sum_{v=1}^{S1} \sum_{w=1}^{S2} v\chi^{(u,v,w)}.$$

2. $EIC2$ denotes the expected level of commodity 2 (the compliment item). Using the steady state probability vector $\boldsymbol{\chi}$, we define $EIC2$ by

$$EIC2 = \sum_{u=0}^{\infty} \sum_{v=0}^{S1} \sum_{w=1}^{S2} w\chi^{(u,v,w)}.$$

3. $ERC1$ denotes the expected reorder rate of commodity 1. Reorder for $Q$ items is placed whenever the system reaches to $s$ from $s+1$ under the $(s, Q)$ reordering policy. Therefore, $ERC1$ is given by

$$ERC1 = \sum_{u=0}^{\infty} \sum_{w=1}^{S2} [u\theta + (s+1)\gamma_1 + \lambda] \chi^{(u,s+1,w)}.$$

4. $ERC2$ denotes the expected reorder rate of commodity 2. As instantaneous reordering policy is considered for the commodity 2, $S_2$ items are replenished immediately whenever system drops from 1. Then $ERC2$ is defined by

$$ERC2 = \sum_{u=0}^{\infty} \sum_{v=1}^{S1} [ur\theta + \gamma_2 + r\lambda] \chi^{(u,v,1)} + \sum_{u=0}^{\infty} \gamma_2 \chi^{(u,0,1)}.$$

5. $ECRO$ denotes the expected number of customers in the orbit. Therefore, it is defined by

$$ECRO = \sum_{u=1}^{\infty} \sum_{v=0}^{S1} \sum_{w=1}^{S2} u\chi^{(u,v,w)}.$$

6. $EORC$ denotes the overall rate of retrial customers. The customer from the orbit can try to buy the product in the system irrespective of the product's availability. Then we have

$$EORC = \sum_{u=1}^{\infty} \sum_{v=0}^{S1} \sum_{w=1}^{S2} u\theta\chi^{(u,v,w)}.$$

7. $ESRC$ denotes the successful rate of retrial customers. Whenever the orbit customer finds that there exists a positive commodity 1, then their retrial process will be successful. It is given by

$$ESRC = \sum_{u=1}^{\infty} \sum_{v=1}^{S1} \sum_{w=1}^{S2} u\theta\chi^{(u,v,w)}.$$

8. $EPC1$ denotes the expected number of perishable commodity 1. Due to the life time, commodity 1 can be perishable at anytime. The mean number of perishable commodity 1 is defined as

$$EPC1 = \sum_{u=0}^{\infty} \sum_{v=1}^{S1} \sum_{w=1}^{S2} v\gamma_1\chi^{(u,v,w)}.$$

9. The expected number of perishable commodity 2 denoted by $EPC2$ is given by

$$EPC2 = \sum_{u=0}^{\infty} \sum_{v=0}^{S1} \sum_{w=1}^{S2} w\gamma_2\chi^{(u,v,w)}.$$

10. When the customer finds that commodity 1 is empty, they leave the system with a probability of $(1-p)$. Therefore, an expected customer lost in the system is defined as an $ECL$ at any time by

$$ECL = \sum_{u=0}^{\infty} \sum_{w=1}^{S2} (1-p)\lambda\chi^{(u,0,w)}.$$

## 6. Cost analysis and numerical illustration

Here, we discuss the feasibility of a proposed model through the system characteristics and sufficient economic illustrations. The expected total cost (ETC) is given by

$$ETC(S1, S2) = C_{h1}EIC1 + C_{h2}EIC2 + C_{s1}ERC1 + C_{s2}ERC2 + C_{p1}EPC1 \\ + C_{p2}EPC2 + C_w E[W_o] + C_l ECL.$$

To compute the ETC per unit time, the following costs are considered.
$C_{h1}$: Carrying cost of commodity 1/unit item.
$C_{h2}$: Carrying cost of commodity 2/unit item.
$C_{s1}$: Ordering cost of commodity 1/order.
$C_{s2}$: Ordering cost of commodity 2/order.
$C_{p1}$: Perishable cost of commodity 1/unit item.
$C_{p2}$: Perishable cost of commodity 2/unit item.
$C_w$: Waiting cost of an orbiting customer/unit customer.
$C_l$: Cost of a customer lost/unit customer.

### 6.1. Numerical illustration

Numerical analysis is an applied mathematical technique that allows a staggeringly large amount of data to be processed and analyzed for trends, thereby aiding in forming conclusions. To do such numerical illustrations, we fix the cost values as

$$C_{h1} = 0.45, \quad C_{h2} = 0.15, \quad C_{s1} = 20.5, \quad C_{s2} = 5.5, \\ C_{p1} = 2, \quad C_{p2} = 0.8, \quad C_w = 2.6, \quad C_l = 6.6,$$

and the parameters are $\lambda = 1.6$, $p = 0.94$, $r = 0.92$, $\beta = 1.23$, and $\theta = 0.65$.

**Case (i).** Under the following assumptions on various costs and parameters, we illustrate the optimality and convexity of the cost function with independent ordering quantities of both commodities according to the following given range of $S_1$ and $S_2$ (Table 1). Let $S_1 = 15, 16, 17, 18, 19$ and $S_2 = 2, 3, 4, 5, 6$. Using Table 1, under the range of $S_1$, we obtain the minimum expected total cost for each $S_2$ which is noted in bold script. Similarly, under the given range of $S_2$, the minimum expected total cost for each $S_1$ is noted with an underline. From these discussions, the least optimum expected total cost exists at $S_1 = 17$ and $S_2 = 4$, which is given in bold script with an underline (Table 1). Since the proposed model holds the convex properties on the expected

Table 1. ETC rate as a function of $S_1$ and $S_2$

| $S_1 S_2$ | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 15 | 7.52903 | 7.12110 | **7.08464** | 7.11847 | 7.18228 |
| 16 | 7.34905 | 7.02337 | **6.99638** | 7.03800 | 7.10035 |
| 17 | <u>7.27243</u> | <u>7.00295</u> | <u>**6.98645**</u> | <u>7.02999</u> | <u>7.09535</u> |
| 18 | 7.28788 | 7.02966 | **7.02417** | 7.07014 | 7.13718 |
| 19 | 7.30836 | 7.09246 | **7.09167** | 7.14019 | 7.20857 |

total cost under the variation of pair of parameters $(S_1, S_2)$, it can be apply to the real-life product sales business. This model gives the optimal total cost of the entire system for the fixed set of parameters. In a business, one can run it successfully if the entire business process is balanced. That

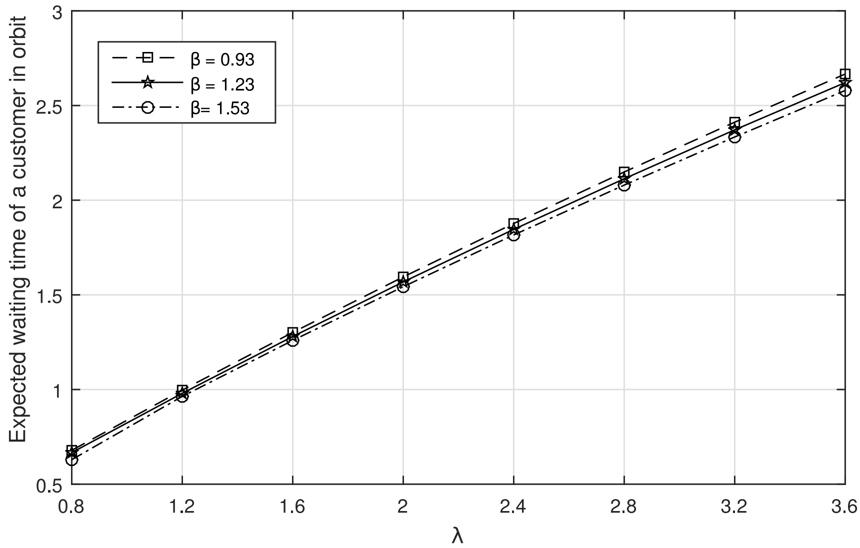



Figure 2. $E[W_o]$ vs $\lambda$ and $\beta$

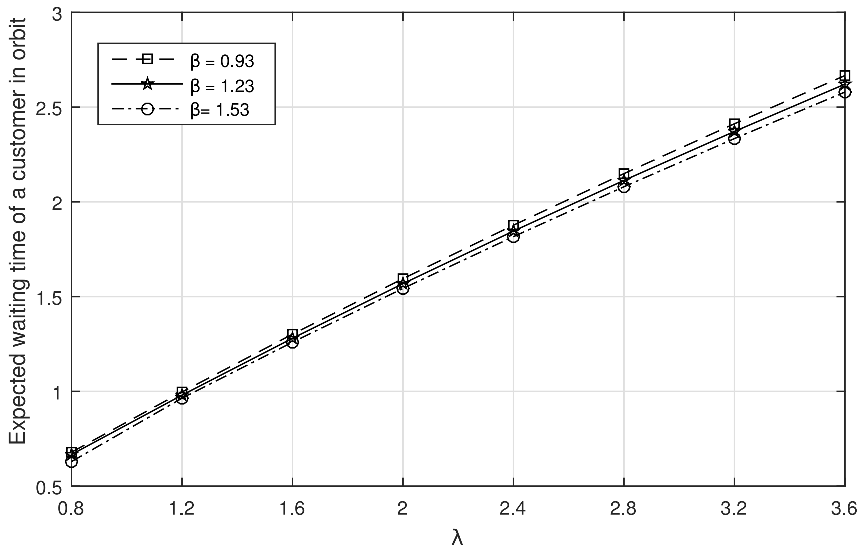



Figure 3. $E[W_o]$ vs $\lambda$ and $\gamma_1$

is, a balanced business means that it will maintain a good relationship between the customer and the system operator. To run a successful business, the customer-owner relationship is important, but at the same time, our business does not fall down. This aspect is determined by the optimal results of the system in a business. In such a way, the proposed model will give the assurance of providing an efficient business.

**Case (ii).** In this study, various parameters influencing the customers lost and the expected total cost rate (from Figs. 2–7) are discussed.

1. Under a Bernoulli's schedule, the rate of customers entering into the orbit increases as $\lambda$ increases, and so both the expected waiting time and expected total cost also increase (Fig. 2–Fig. 7).

2. According to any given increment of mean replenishment time of a commodity 1 $(1/\beta)$, expected waiting time and expected total cost are found to be more sensitive at the higher value of $\lambda$. Nevertheless, we also notice that both measures are poorly sensitive at the optimum value of $\lambda$ (Fig. 5–Fig. 8).

3. According to any given increment of mean spoilage time of a commodity 1, the rate of the expected waiting time is not significant at any value of $\lambda$ (Fig. 4–Fig. 7).

4. Also $\theta$ makes a significant effect on expected total cost, and further it is highly significant at the higher value of $\lambda$, expected waiting time increases as $\lambda$ increases (Fig. 4 and Fig. 7).

According to this analysis, we can relate this numerical to real-life phenomena. In a business, the system manager will pay attention to controlling the average waiting time of a customer and the total cost of the system through the balanced mechanism. Here, the probability $p$ plays such a role in the system. Then the reorder process of a business also plays an important role in reducing the actual waiting time of a customer and the expected total cost of the system. Whenever the system manager controls the average replenishment time, which does not exceed its limit as much as possible, this analysis can be made easier by the case (ii). Similarly, the arrival rate influences those system metrics.

**Case (iii).** The choice of a customer's demand is for either a single commodity (commodity 1) or both (commodity 1 and commodity 2). Hence, all the measures that are relevant to commodity 1 do not make any significant differences. But the rate of choice of demand for both commodities makes a significant change in the measures relevant to commodity 2. Fig. 8 to Fig. 13 show that some significant effects on expected total cost with the expected reorder rate of commodity 2 are noticed according to the independent decision to make the amount of commodity 1 and commodity 2.

1. Under Bernoulli's theory, if the probability of demand for both commodities $(r)$ increases, then the expected reorder rate of commodity 2 and the expected total cost increase (Fig. 8 and Fig. 13).

2. Also, we determine that each measure of this study depends linearly on both commodities demand.

3. When comparing $S_1, s$, and $S_2$, the value of $S_1$ is more sensitive on expected total cost with expected reorder rate of commodity 2 (Fig. 8–Fig. 10).

4. Also we notice that both the measures are highly sensitive at the higher value of $s$ and the lower value of $S_1$ and $S_2$.

5. From this study, it is clear that both commodities' ordered quantities have a significant effect when the demand rate for both commodities increases.

The purchase of both commodities by a customer is decided by the probability $r$ and its complement. This concept will be very helpful to analyze how many customers can buy both products and single products in the inventory sales business. Apart from that, an efficient businessman must have enough knowledge about the available stocking positions in the system, then only they can make a plan to place a reorder of the required items. If the requirement for both commodities increases, the system manager will pay attention to observing the reorder point of the commodity and the available stocks. This analysis will be the required analysis to have such knowledge.

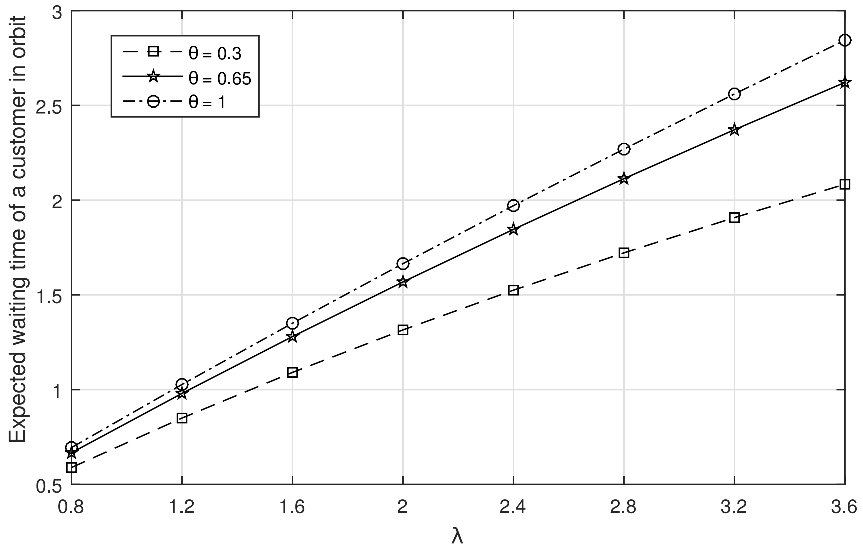



Figure 4. $E[W_o]$ vs $\lambda$ and $\theta$

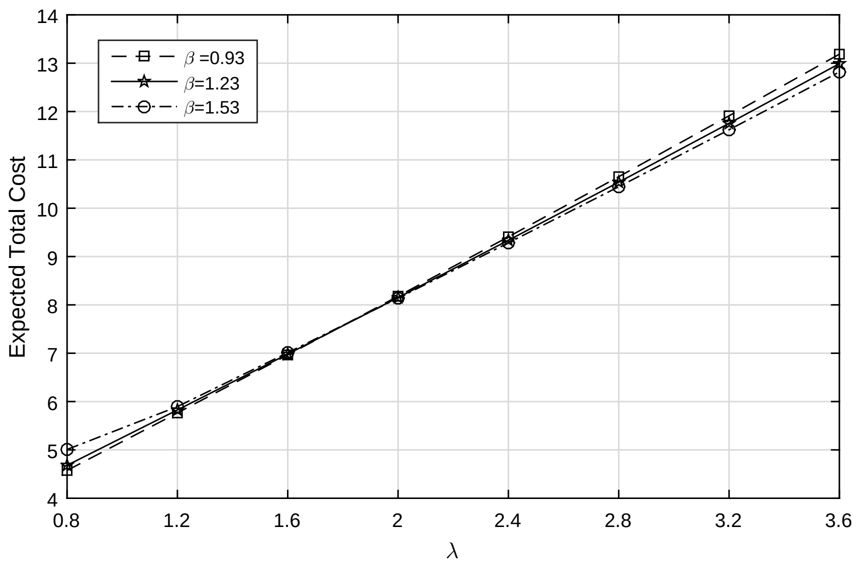



Figure 5. $ETC$ vs $\lambda$ and $\beta$

**Case (iv).** Suppose there is no stock in the system, the preference of a customer enters into the orbit under a Bernoulli's schedule. Using Tables 2–4, we determine the following merits of the proposed model with various measures according to the decision for making the number of orders/production.

1. For every cycle, if the production or order quantity of commodity 1 increases, the expected number of customers in orbit, the expected reorder rate of commodity 1, and the expected number of customers lost decrease significantly.

2. Also, we determine that the expected inventory of commodity 1 and its expected perishable quantity will significantly increase. But we can obtain the minimum expected total cost with a suitable order quantity.

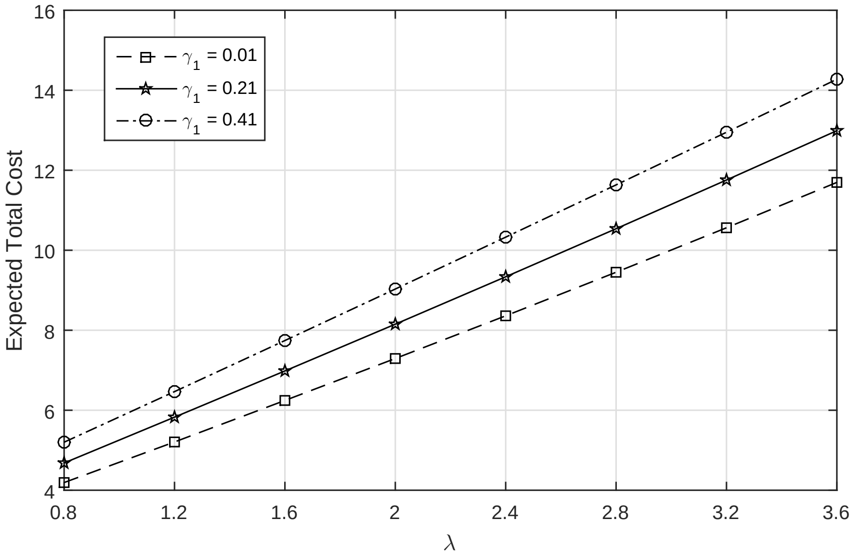



Figure 6. $ETC$ vs $\lambda$ and $\gamma_1$

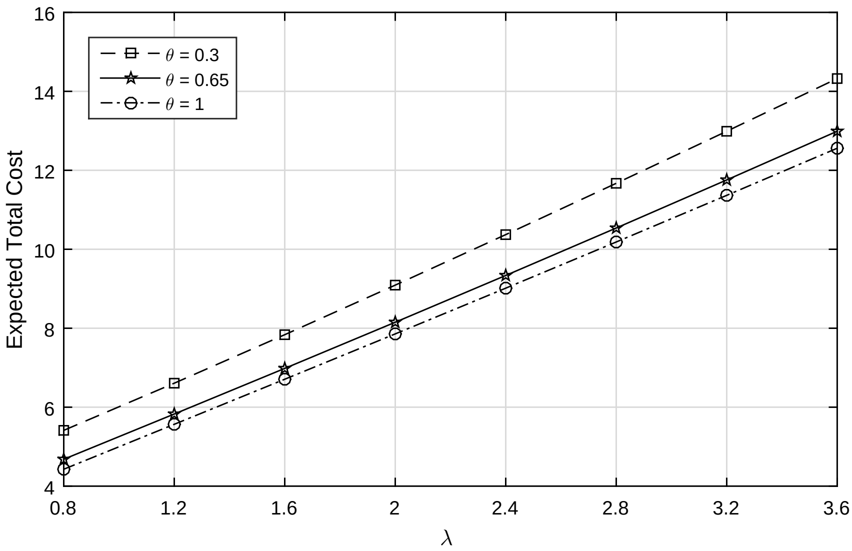



Figure 7. $ETC$ vs $\lambda$ and $\theta$

3. Further, if $p$ increases, then the rate of customers entering the orbit increases and the expected number of customers lost decreases. If $p = 1$, then the system cannot find any customers lost.

4. Also, we notice that the expected number of customers in orbit, the expected reorder rate, the expected inventory level of commodity 1, and its expected perishable quantity are all increasing at a slower rate due to the classical retrial policy.

Table 2. Response of $S_1$ vs $p$ on various measures of the system

| $S_1$ | $p$ | $EIC1$ | $ERC1$ | $EPC1$ | $ECRO$ | $ECL$ | $ETC$ |
|---|---|---|---|---|---|---|---|
| | 0.88 | 0.74143 | 0.16624 | 0.15570 | 1.20715 | 0.16461 | 9.97351 |
| 11 | 0.94 | 0.75841 | 0.16587 | 0.15926 | 1.28318 | 0.08190 | 9.64499 |
| | 1.00 | 0.77503 | 0.16554 | 0.16275 | 1.35851 | 0.00000 | 9.32038 |
| | 0.88 | 1.50279 | 0.00958 | 0.31558 | 1.20451 | 0.16425 | 7.27306 |
| 17 | 0.94 | 1.53739 | 0.01081 | 0.32285 | 1.27996 | 0.08169 | 6.98229 |
| | 1.00 | 1.57109 | 0.01207 | 0.32992 | 1.35463 | 0.00000 | 6.69543 |
| | 0.88 | 2.33520 | 0.00081 | 0.49039 | 1.20436 | 0.16423 | 7.79526 |
| 23 | 0.94 | 2.39262 | 0.00100 | 0.50245 | 1.27974 | 0.08168 | 7.49968 |
| | 1.00 | 2.44854 | 0.00123 | 0.51419 | 1.35435 | 0.00000 | 7.20727 |

Table 3. Response of $s$ vs $p$ on various measures of the system

| $s$ | $p$ | $EIC1$ | $ERC1$ | $EPC1$ | $ECRO$ | $ECL$ | $ETC$ |
|---|---|---|---|---|---|---|---|
| | 0.88 | 1.77499 | 0.00136 | 0.37274 | 1.20443 | 0.16424 | 7.32728 |
| 3 | 0.94 | 1.81612 | 0.00166 | 0.38138 | 1.27984 | 0.08169 | 7.02110 |
| | 1.00 | 1.85604 | 0.00199 | 0.38977 | 1.35449 | 0.00000 | 6.71855 |
| | 0.88 | 1.50279 | 0.00958 | 0.31558 | 1.20451 | 0.16425 | 7.27306 |
| 5 | 0.94 | 1.53739 | 0.01081 | 0.32285 | 1.27996 | 0.08170 | 6.98229 |
| | 1.00 | 1.57109 | 0.01207 | 0.32992 | 1.35463 | 0.00000 | 6.69543 |
| | 0.88 | 1.26500 | 0.07046 | 0.26565 | 1.20469 | 0.16427 | 8.34847 |
| 7 | 0.94 | 1.29626 | 0.07324 | 0.27221 | 1.28018 | 0.08171 | 8.09013 |
| | 1.00 | 1.32691 | 0.07585 | 0.27865 | 1.35492 | 0.00000 | 7.83216 |

Table 4. Response of $S_2$ vs $p$ on various measures of the system

| $S_2$ | $p$ | $EIC2$ | $ERC2$ | $EPC2$ | $ECRO$ | $ECL$ | $ETC$ |
|---|---|---|---|---|---|---|---|
| | 0.88 | 1.35030 | 0.23295 | 0.43209 | 1.20451 | 0.16425 | 7.54911 |
| 3 | 0.94 | 1.35122 | 0.23706 | 0.43239 | 1.27996 | 0.08170 | 7.27871 |
| | 1.00 | 1.35215 | 0.24126 | 0.43268 | 1.35463 | 0.00000 | 7.01223 |
| | 0.88 | 1.98000 | 0.13627 | 0.63360 | 1.20451 | 0.16425 | 7.27306 |
| 5 | 0.94 | 1.98326 | 0.13651 | 0.63464 | 1.27996 | 0.08170 | 6.98229 |
| | 1.00 | 1.98652 | 0.13683 | 0.63568 | 1.35463 | 0.00000 | 6.69543 |
| | 0.88 | 2.55762 | 0.11346 | 0.81844 | 1.20451 | 0.16425 | 7.38211 |
| 7 | 0.94 | 2.56362 | 0.11326 | 0.82036 | 1.27996 | 0.08170 | 7.09008 |
| | 1.00 | 2.56954 | 0.11311 | 0.82225 | 1.35463 | 0.00000 | 6.80177 |

**Case (v).** Suppose that any corresponding cost rates of both commodities may be changed, then we notice some significant effects of the proposed model according to the given parameters' values (from Table 5).

1. For any holding cost and perishable cost, the expected total cost rate decreases as the setup cost decreases.

Table 5. Expected Total cost with different combinations of various cost

| $C_{p1}$ | $C_{p2}$ | $C_{h1}$ | $C_{h1}$ | $C_{s1} = 5.5$ | | $C_{s1} = 20.5$ | |
|---|---|---|---|---|---|---|---|
| | | | | $C_{s2} = 5.5$ | $C_{s2} = 20.5$ | $C_{s2} = 5.5$ | $C_{s2} = 20.5$ |
| 0.8 | 0.8 | 0.15 | 0.15 | 5.97149 | 8.01918 | 6.13365 | 8.18134 |
| | | | 0.45 | 6.56647 | 8.61416 | 6.72863 | 8.77632 |
| | | 0.45 | 0.15 | 6.43271 | 8.48040 | 6.59486 | 8.64255 |
| | | | 0.45 | 7.02769 | 9.07538 | 7.18984 | 9.23754 |
| | 2.0 | 0.15 | 0.60 | 6.73306 | 8.78075 | 6.89522 | 8.94291 |
| | | | 0.80 | 7.32804 | 9.37573 | 7.49020 | 9.53789 |
| | | 0.45 | 1.00 | 7.19428 | 9.24197 | 7.35644 | 8.64255 |
| | | | 2.00 | 7.78926 | 9.83695 | 7.95142 | 9.99911 |
| 2.0 | 0.8 | 0.15 | 0.15 | 6.35891 | 8.40660 | 6.52107 | 8.56876 |
| | | | 0.45 | 6.95389 | 9.00158 | 7.11605 | 9.16374 |
| | | 0.45 | 0.15 | 6.82013 | 8.86782 | 6.98229 | 9.02998 |
| | | | 0.45 | 7.41511 | 9.46280 | 7.57727 | 9.62496 |
| | 2.0 | 0.15 | 0.60 | 7.12049 | 9.16818 | 7.28264 | 9.33034 |
| | | | 0.80 | 7.71547 | 9.76316 | 7.87762 | 9.92532 |
| | | 0.45 | 1.00 | 7.58170 | 9.62940 | 7.74386 | 9.79155 |
| | | | 2.00 | 8.17668 | 10.22438 | 8.33884 | 10.38653 |

2. On the other hand, the expected total cost rate is maximum if the setup costs are equal and maximal.
3. If holding costs and perishability costs of commodity 2 increase, then the expected cost rate increases significantly.
4. The expected cost rate is highly sensitive when all costs increase.

## 7. Conclusion

We investigate the Markovian TC inventory system with a classical retrial facility. The system allows a customer who has purchased a free product to conduct Bernoulli trials at will. Also, it is assumed that the retrial process follows a classical retrial policy and an $(s, Q)$ ordering policy for replenishment. The system's stability is derived through the matrix-geometric approximation; indeed, the optimum total cost is computed along with the stationary probability vector. Further, through the numerical illustration, the merits of the proposed model are explained. This model helps the producers to cope with the market as well as the public to achieve better success in the sale of their products and earn profits. Evaluating the proposed model in an economy is one of the most crucial decision-making variables that a business must analyze to survive and grow in a competitive market.

## Acknowledgements

The authors would like to sincerely thank two anonymous reviewers for their valuable comments and constructive suggestions, which are very helpful in improving the presentation of this paper.

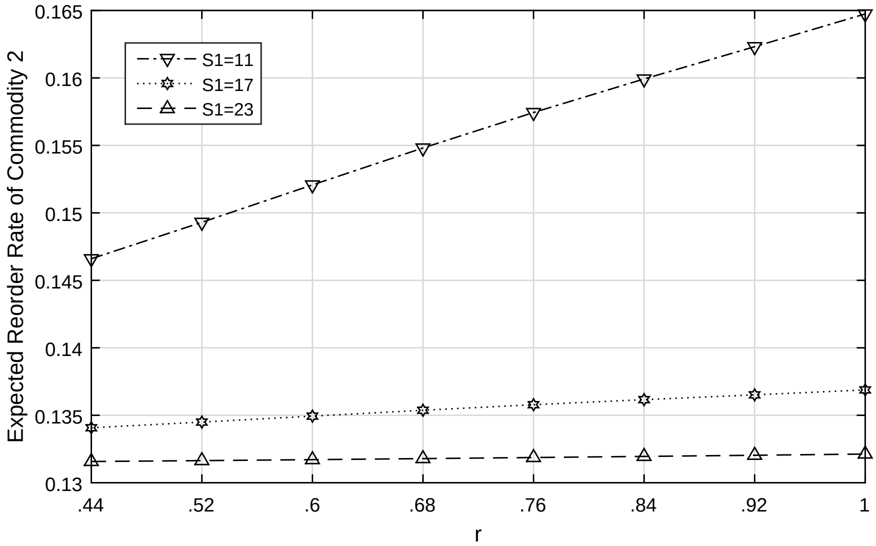



Figure 8. $ERC2$ vs $S_1$ and $r$

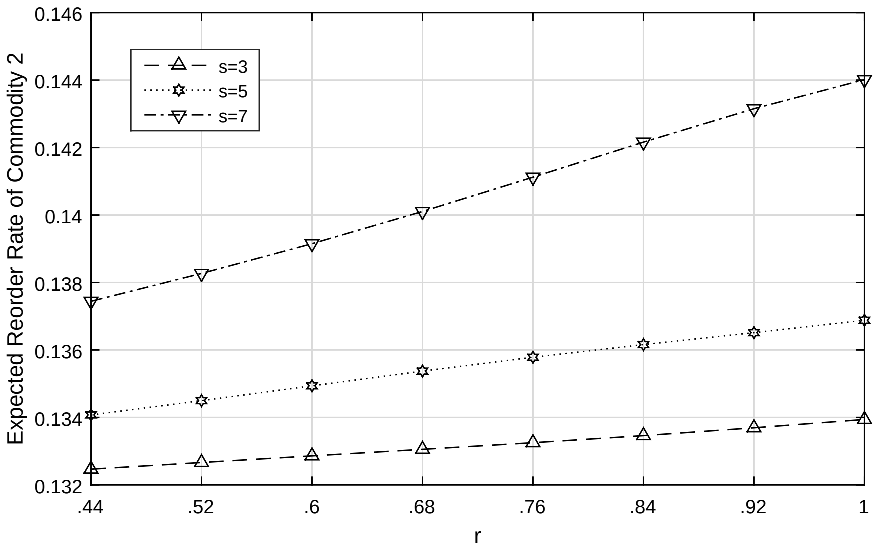



Figure 9. $ERC2$ vs $s$ and $r$

## REFERENCES


1. Abate J., Whitt W. Numerical inversion of Laplace transforms of probability distributions. *ORSA J. Comput.*, 1995. Vol. 7, No. 1. P. 36–43. DOI: 10.1287/ijoc.7.1.36
2. Anbazhagan N., Arivarignan G. Two commodity continuous review inventory system with coordinated reorder policy. *Int. J. Inf. Manag. Sci.*, 2000. Vol. 11, No. 3. P. 19–30.
3. Anbazhagan N., Arivarignan G. Two commodity inventory system with individual and joint ordering policies. *Int. J. Manag. Syst.*, 2003. Vol. 19, No. 2. P. 129–144.
4. Anbazhagan N., Arivarignan G. Analysis of two commodity Markovian inventory system with lead time. *Korean J. Comput. & Appl. Math.*, 2001. Vol. 8, No. 2. Art. no. 427. P. 427–438. DOI: 10.1007/BF02941976
5. Anbazhagan N., Jeganathan K. Two-commodity Markovian inventory system with compliment and retrial demand. *J. Adv. Math. Comput. Sci.*, 2013. Vol. 3, No. 2. P. 115–134. DOI: 10.9734/BJMCS/2013/1862

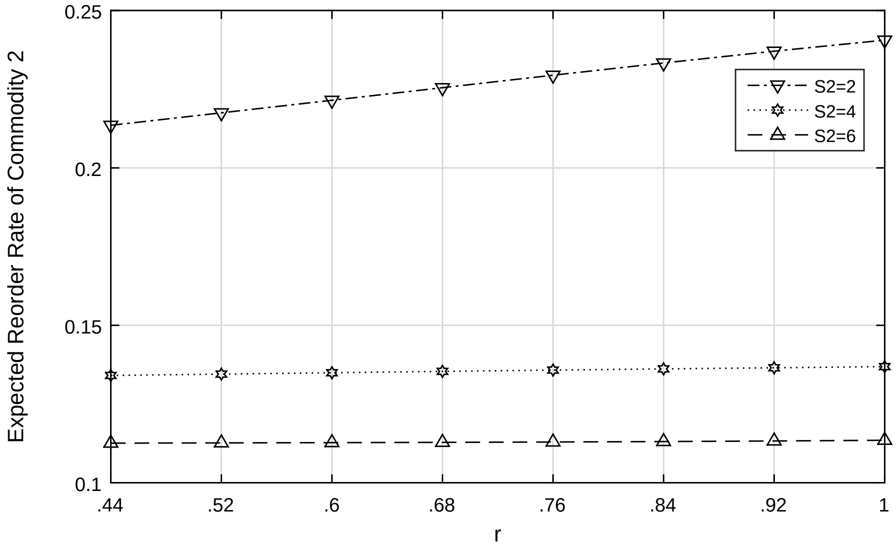



Figure 10. $ERC2$ vs $S_2$ and $r$

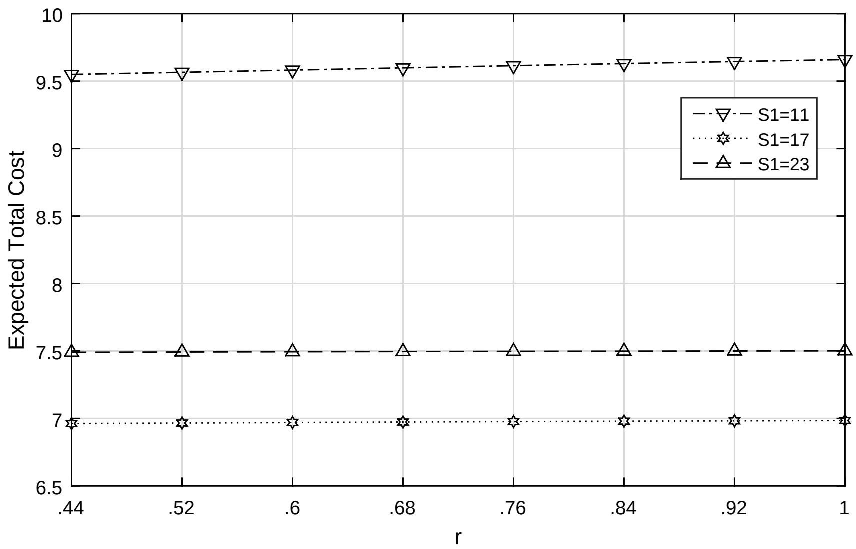



Figure 11. $ETC$ vs $S_1$ and $r$

6. Arivarignan G., Keerthana M., Sivakumar B. A production inventory system with renewal and retrial demands. In: *Applied Probability and Stochastic Processes. Joshua V., Varadhan S., Vishnevsky V. (eds)*. Infosys Sci. Found. Ser. Singapor: Springer, 2020. P. 129–142. DOI: 10.1007/978-981-15-5951-8_9

7. Artalejo J. R., Krishnamoorthy A., Lopez-Herrero M. J. Numerical analysis of $(s, S)$ inventory systems with repeated attempts. *Ann. Oper. Res.*, 2006. Vol. 141. P. 67–83. DOI: 10.1007/s10479-006-5294-8

8. Benny B., Chakravarthy S. R., Krishnamoorthy A. Queueing-inventory system with two commodities. *J. Indian Soc. Probab. Stat.*, 2018. Vol. 19, No. 1. P. 437–454. DOI: 10.1007/s41096-018-0052-1

9. Chakravarthy S. R., Krishnamoorthy A., Joshua V. C. Analysis of a multi-server retrial queue with search of customers from the orbit. *Perform. Eval.*, 2006. Vol. 63, No. 8. P. 776–798. DOI: 10.1016/j.peva.2005.09.002

10. Goyal S. K., Satir A. T. Joint replenishment inventory control: deterministic and stochastic models. *European J. Oper. Res.*, 1989. Vol. 38, No. 1. P. 2–13. DOI: 10.1016/0377-2217(89)90463-3

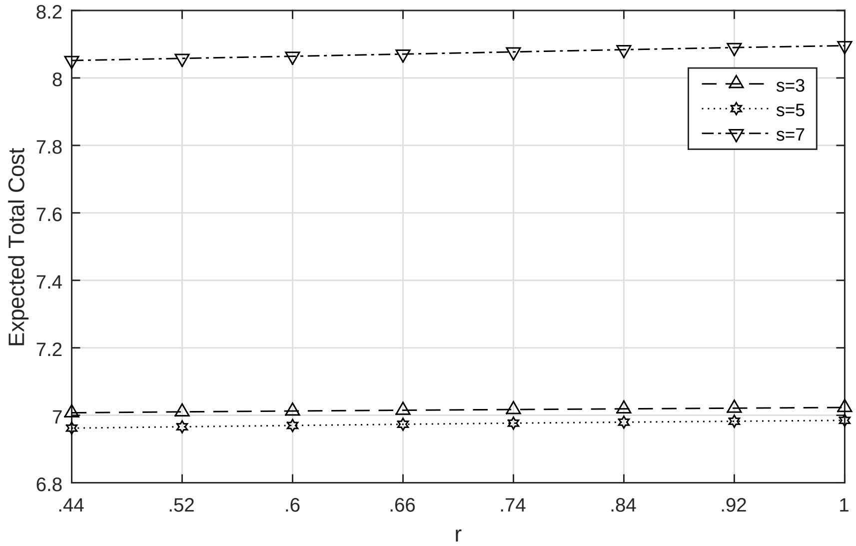



Figure 12. $ETC$ vs $s$ and $r$

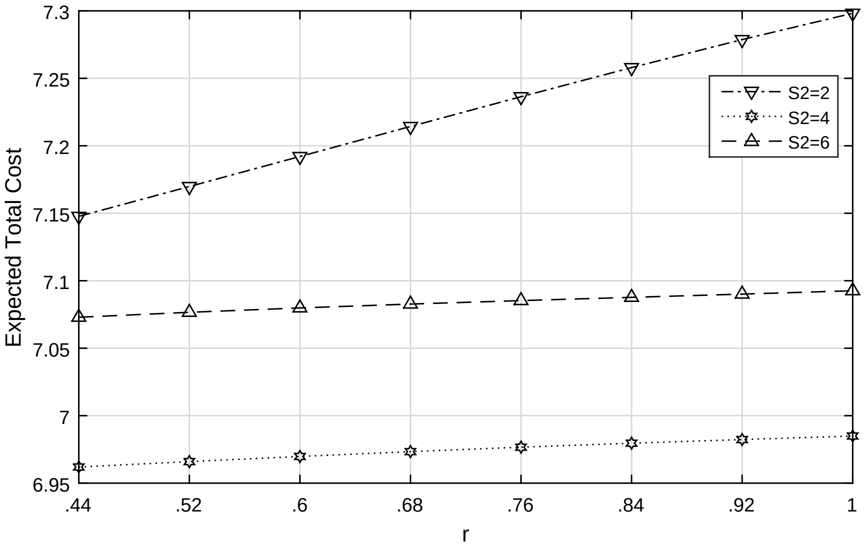



Figure 13. $ETC$ vs $S_2$ and $r$


11. Jeganathan K., Reiyas M. A., Padmasekaran S., Lakshmanan K. An $M/E_K/1/N$ queueing-inventory system with two service rates based on queue lengths. *Int. J. Appl. Comput. Math.*, 2017. Vol. 3, No. 1. P. 357–386. DOI: 10.1007/s40819-017-0360-2

12. Jeganathan K., Melikov A. Z., Padmasekaran S. et al. A stochastic inventory model with two queues and a flexible server. *Int. J. Appl. Comput. Math.*, 2019. Vol. 5, No. 1. Art. no. 21. P. 1–27. DOI: 10.1007/s40819-019-0605-3

13. Jeganathan K., Reiyas M. A. Two parallel heterogeneous servers Markovian inventory system with modified and delayed working vacations. *Math. Comput. Simulation*, 2020. Vol. 172. P. 273–304. DOI: 10.1016/j.matcom.2019.12.002

14. Jeganathan K., Harikrishnan T., Selvakumar S. et al. Analysis of interconnected arrivals on queueing-inventory system with two multi-server service channels and one retrial facility. *Electronics*, 2021. Vol. 10, No. 5. P. 1–35. DOI: 10.3390/electronics10050576

15. Jose K. P., Reshmi P. S. A production inventory model with deteriorating items and retrial demands. *OPSEARCH*, 2021. Vol. 58. P. 71–82. DOI: 10.1007/s12597-020-00471-8
16. Kalpakam S., Arivarignan G. A coordinated multicommodity $(s, S)$ inventory system. *Math. Comput. Model.*, 1993. Vol. 18, No. 11. P. 69–73. DOI: 10.1016/0895-7177(93)90206-E
17. Keerthana M., Saranya N., Sivakumar B. A stochastic queueing-inventory system with renewal demands and positive lead time. *Eur. J. Ind. Eng.*, 2020. Vol. 14, No. 4. P. 443–484. DOI: 10.1504/ejie.2020.10029435
18. Krishnamoorthy A., Iqbal Basha R., Lakshmy B. Analysis of two commodity inventory problem. *Int. J. Inf. Manag. Sci.*, 1994. Vol. 5, No. 1. P. 127–136.
19. Krishnamoorthy A., Merlymol J., Ravindranathan. Analysis of a bulk demand two commodity inventory problem. *Calcutta Stat. Assoc. Bull.*, 1999. Vol. 49, No. 1–2. P. 193–194. DOI: 10.1177/0008068319990108
20. Krishnamoorthy A., Manikandan R., Lakshmy B. A revisit to queueing-inventory system with positive service time. *Ann. Oper. Res.*, 2015. Vol. 233. P. 221–236. DOI: 10.1007/s10479-013-1437-x
21. Krishnamoorthy A., Shajin D., Lakshmy B. $GI/M/1$ type queueing-inventory systems with postponed work, reservation, cancellation and common life time. *Indian J. Pure Appl. Math.*, 2016. Vol. 47, No. 2. P. 357–388. DOI: 10.1007/s13226-016-0192-5
22. Krishnamoorthy A., Shajin D. $MAP/PH/1$ retrial queueing-inventory system with orbital search and reneging of customers. In: *Analytical and Computational Methods in Probability Theory (ACMPT 2017). Rykov V., Singpurwalla N., Zubkov A. (eds).* Lecture Notes in Comput. Sci., vol. 10684. Cham: Springer, 2017. P. 158–171. DOI: 10.1007/978-3-319-71504-9_15
23. Lakshmanan K., Padmasekaran S., Jeganathan K. Mathematical analysis of queueing-inventory model with compliment and multiple working vacations. *Int. J. Eng. Adv. Technol.*, 2019. Vol. 8, No. 6. P. 4239–4240. DOI: 10.35940/ijeat.F9003.088619
24. Neuts M. F. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach.* New York: Dover Publication Inc., 1994. 332 p.
25. Senthil Kumar P. A discrete time two commodity inventory system. *Int. J. Engrg. Res. Technol.*, 2016. Vol. 5, No. 1. P. 708–719. DOI: 10.17577/IJERTV5IS010602
26. Shajin D., Krishnamoorthy A., Dudin A. N. et al. On a queueing-inventory system with advanced reservation and cancellation for the next $K$ time frames ahead: the case of overbooking. *Queueing Syst.*, 2020. Vol. 94. P. 3–37. DOI: 10.1007/s11134-019-09631-0
27. Sivazlian B. D. Stationary analysis of a multicommodity inventory system with interacting set-up costs. *SIAM J. App. Math.*, 1971. Vol. 20. Vol. 2. P. 264–278. https://www.jstor.org/stable/2099925?seq=1
28. Sivakumar B., Anbazhagan N., Arivarignan G. A two-commodity perishable inventory system. *ORiON*, 2005. Vol. 21, No. 2. P. 157–172. DOI: 10.5784/21-2-27
29. Sivakumar B., Anbazhagan N., Arivarignan G. Two commodity continuous review perishable inventory system. *Int. J. Inf. Manag. Sci.*, 2006. Vol. 17, No. 3. P. 47–64.
30. Sivakumar B., Anbazhagan N., Arivarignan G. Two-commodity inventory system with individual and joint ordering policies and renewal demands. *Stoch. Anal. Appl.*, 2007. Vol. 25, No. 6. P. 1217–1241. DOI: 10.1080/07362990701567314
31. Sivakumar B. Two-commodity inventory system with retrial demand. *Eur. J. Oper. Res.*, 2008. Vol. 187, No. 1. P. 70–83. DOI: 10.1016/j.ejor.2007.02.036
32. Ushakumari P. V. On $(s, S)$ inventory system with random lead time and repeated demands. *J. Appl. Math. Stochastic Anal.*, 2006. Vol. 2006. Art. no. 81508. P. 1–22. DOI: 10.1155/JAMSA/2006/81508
33. Yadavalli V. S. S., Anbazhagan N., Arivarignan G. A two-commodity stochastic inventory system with lost sales. *Stoch. Anal. Appl.*, 2004. Vol. 22, No. 2. P. 479–497. DOI: 10.1081/SAP-120028606
34. Yadavalli V. S. S., Arivarignan G., Anbazhagan N. Two commodity coordinated inventory system with Markovian demand. *Asia-Pac. J. Oper. Res.*, 2006. Vol. 23, No. 4. P. 497–508. DOI: 10.1142/S0217595906001005
35. Yadavalli V. S. S., Sivakumar B., Arivarignan G., Olufemi Adetunji. A finite source multi-server inventory system with service facility. *Comput. Ind. Eng.*, 2012. Vol. 63. P. 739–753. DOI: 10.1016/j.cie.2012.04.014
36. Yadavalli V. S. S., Adetunji O., Sivakumar B., Arivarignan G. Two-commodity perishable inventory system with bulk demand for one commodity. *S. Afr. J. Ind. Eng.*, 2010. Vol. 21, No. 1. P. 137–155. DOI: 10.7166/21-1-73

**URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 117–127**

DOI: 10.15826/umj.2022.1.010

# MONOPOLISTIC COMPETITION MODEL WITH ENTRANCE FEE

**Olga A. Tilzo**

Novosibirsk State University,
2 Pirogova str., Novosibirsk, 630090, Russian Federation

kidanovaola@gmail.com

**Abstract:** We study the monopolistic competition model with producer-retailer-consumers two-level interaction. The industry is organized according to the Dixit–Stiglitz model. The retailer is the only monopolist. A quadratic utility function represents consumer preferences. We consider the case of the retailer's leadership; namely, we study two types of behavior: with and without the free entry condition. Earlier, we obtained the result: to increase social welfare and/or consumer surplus, the government needs to subsidize (not tax!) retailers. In the presented paper, we develop these results for the situation when the producer imposes an entrance fee for retailers.

**Keywords:** Monopolistic competition, Retailing, Equilibrium, Taxation, Entrance fee, Social welfare, Consumer surplus.

## 1. Introduction

The modeling the relations of market agents (producer, retailer, consumer, etc.) can be carried out in various ways. Let us distinguish the main clusters.

Firstly, the models may be characterized by different ways of interaction between participants in competitive relationships. In particular, the "leader-follower" model was firstly studied in detail in the classical works of Stackelberg. The model considers the case of leadership in terms of output under the conditions of an asymmetric duopoly, where one of the firms makes its choice before the other.

Secondary, the models of market spatial differentiation. In the spatial models theme, the foundational work is Hotelling's linear city model [9]. In this case, the horizontal differentiation of goods is characterized by the geographical location of the producer on a unit interval. Besides, the transport costs for the delivery of goods to the consumer are imposed. While consumers are distributed over the interval evenly, their preferences are asymmetric. Note that the number of firms in the market assumes constant. Another example of this class of models is the Salop model [12] (circular city model) with one producer and several retailers located along a circle (street) at an equal distance between each other; consumers are evenly distributed along the circle and have the same preferences.

Thirdly, the models are characterized by different types of utility functions. The work of Perry and Groff [11] considers the $CES$ utility function and evaluates the impact of integration on the change in the level of social welfare. In the work of Ottaviano, Tabuchi, and Thisse [10], a model of economic geography with a quadratic utility function is studied: here several production factors, as well as transport costs, are considered.

Fourthly, markets with different types of product differentiation, horizontal or vertical, are considered in the work of Gabzewicz and Thisse [8]. For the different locations of the stores (the firms), the authors study the conditions for the equilibrium existence.

In the presented paper, the industry is organized according to the Dixit–Stiglitz model [6, 7] with a quadratic utility function (cf. [10]). The quadratic utility function generates the linear demand function. The model is supplemented by a monopoly retailer. This way, we model a two-level interaction. The mass (the number) of producers is quite large (cf. [4, 5]). We study two types of retailer behavior: with and without the condition of free entry (cf. [1–3]). The result is that when the retailer imposes an entrance fee for each producer, it leads to an increase in both social welfare and consumer surplus.

Since the proofs of many statements are voluminous and rather technical, we give only proofs of some propositions. As to the other propositions, we provide only the schemes or the main ideas of proofs.

This paper continues the works [13, 14]. More precisely, in [14], we compared different types of interaction; as a result, we considered the situation from the point of view of the manufacturer, retailer, consumers, and society as a whole. In [13], we considered the case of retailer leadership; we studied two situations of retailer behavior: with free entry conditions and without free entry conditions; it turned out that social welfare increases when the retailer is stimulated by subsidies; a similar situation arises when considering consumer surplus.

In the present paper, we supplement the results of [13, 14] by introducing an entrance fee for the producer.

## 2. Model

We study the producer-retailer-consumers two-level interaction, monopolistic competition model. The model adopts several assumptions (see [13]) of monopolistic competition.

It is assumed that the number (mass) of firms is sufficiently large. Each firm produces only one type of product and sets its price. Firms produce goods of the same type ("variety") that are not absolutely substitutable. For firms, the free entry (zero profit) condition is assumed.

Goods on the market are represented by horizontally differentiated products. There are also other products on the market designated as "numéraire." In addition, it is assumed that there are several identical consumers, and each consumer supplies one unit of labor to the market.

Producers sell products through a monopoly retailer, which increases the retail price of goods by adding a mark-up.

We consider the case of linear demand corresponding to the quadratic utility function proposed by Ottaviano, Tabuchi, and Thisse [10]:

$$U(\mathbf{q}, N, A) = \alpha \int_0^N q(i)di - \frac{\beta - \gamma}{2} \int_0^N (q(i))^2 \, di - \frac{\gamma}{2} \left( \int_0^N q(i)di \right)^2 + A, \tag{2.1}$$

where $\alpha > 0$, $\beta > \gamma > 0$ are some parameters[1]; $N$ is the length of the product line, reflecting the range (interval) of varieties; $q(i) \geq 0$ is the consumption volume of variety $i$, $i \in [0, N]$; and $A \geq 0$ is the consumption of other aggregated products ("numéraire").

We introduce the notations: $\mathbf{q} = (q(i))_{i\in[0,N]}$ is an infinite-dimensional vector (profile) of the volume of goods; $\mathbf{p} = (p(i))_{i\in[0,N]}$ is the price profile; and $\mathbf{r} = (r(i))_{i\in[0,N]}$ is the trade mark-up profile.

Let us formulate the budget constraint

$$\int_0^N (p(i) + r(i))q(i)di + P_A A \leq wL + \int_0^N \pi_{\mathcal{M}}(i)di + \pi_{\mathcal{R}}, \tag{2.2}$$

[1] The economic meaning of the parameters can be found in the source [10, p. 413].

the right-hand side of (2.2) represents the gross domestic product (GDP) by income, and the left-hand side is expenditure. Here $p(i)$ is the wholesale price of variety $i$; $r(i)$ is the retailer's mark-up for product variety $i$; $p(i)+r(i)$ is the price of variety $i$ for the consumer, $w \equiv 1$ is the wage rate in the industry, normalized to one; $P_A$ is the price of other goods ("numéraire"), $\pi_{\mathcal{M}}(i)$ is the profit of the firm $i \in [0,N]$, while $\pi_{\mathcal{R}}$ is the retailer's profit.

Let us formulate the representative consumer problem:

$$\begin{cases} U(\mathbf{q},N,A) \to \max\limits_{\mathbf{q},A}, \\ \displaystyle\int_0^N p_{\mathcal{R}}(i)q(i)di + A \le L + \int_0^N \pi_{\mathcal{M}}(i)di + \pi_{\mathcal{R}}, \end{cases} \tag{2.3}$$

where $U(\mathbf{q},N,A)$ is defined in (2.1), and the price of other products $P_A$ and wage rate $w$ in (2.2) are normalized to one.

The problem (2.3) is solved using the Lagrange function. As a result, consumption characteristics can be determined for each $i \in [0,N]$:

$$q(i) = a - (b+cN)(p(i)+r(i)) + cP, \tag{2.4}$$

where the coefficients $a$, $b$, and $c$ are defined as

$$a = \frac{\alpha}{\beta+(N-1)\gamma}, \quad b = \frac{1}{\beta+(N-1)\gamma}, \quad c = \frac{\gamma}{(\beta-\gamma)(\beta+(N-1)\gamma)},$$

and $P$ is the price index

$$P = \int_0^N (p(j)+r(j))dj.$$

Let, as in [13, 14], $d$ be the producer's marginal costs and $F$ be the producer's fixed costs.

Then the problem of maximizing the firm's profit $i \in [0,N]$ can be written as

$$\pi_{\mathcal{M}}(i) = (p(i)-d)q(i,\mathbf{p}+\mathbf{r}) - F \ \to \ \max_{\mathbf{p}}, \tag{2.5}$$

where $q(i)$ is defined in (2.4).

Note that (2.5) is quadratic in $p(i)$.

Now let us formulate the retailer problem. Similarly to the producer problem (2.5) (see [13, 14]), let $d_{\mathcal{R}}$ be the retailer's marginal costs and $F_{\mathcal{R}}$ be the retailer's fixed costs. Let $p^*(i,r(i),N,P)$ be the optimal pricing policies, then the demand is $q(i,r(i),N,P)$ while the profile of mark-up is $\mathbf{r} = (r(i))_{i\in[0,N]}$. Then the problem of maximizing the retailer's profit is

$$\begin{cases} \pi_{\mathcal{R}} = \displaystyle\int_0^N (r(j)-d_{\mathcal{R}})\,q(j)dj - \int_0^N F_{\mathcal{R}}dj \ \to \ \max_{\mathbf{r},N}, \\ \pi_{\mathcal{M}}(p^*(i,r(i),N,P),r(i),N) \ge 0, \quad i \in [0,N]. \end{cases} \tag{2.6}$$

Due to the assumption that the firms are identical, two cases are possible when solving the problem (2.6), namely

- the free entry condition is not taken into account, i.e., $\pi_{\mathcal{M}}(i) > 0$;
- the free entry condition is taken into account, i.e., $\pi_{\mathcal{M}}(i) = 0$.

The Stackelberg equilibrium under the retailer's leadership is considered.

Let us denote the case of the retailer's leadership with the free entry condition as $RL$, and the case of the retailer's leadership without taking into account the free entry condition as $RL(I)$. These cases are described in detail in [13, 14].

**Case RL.** *The retailer simultaneously chooses a trade mark-up* $\mathbf{r} = (r(i))_{i\in[0,N]}$ *and a mass of firms* $N$*, correctly predicting the subsequent response of the producers.*

**Case RL(I).** *The retailer first uses the free entry condition to calculate* $N = N(\mathbf{r})$*, taking into account the subsequent response of producers, and then maximizes its profit through a trade mark-up* $\mathbf{r}$*.*

It turned out that which particular case (RL or RL(I)) arises is completely determined by the parameter

$$\mathcal{F} = \frac{F_{\mathcal{R}}}{2F}. \tag{2.7}$$

This allows us to formulate the following proposition.

**Proposition 1.** *1. The case RL is possible if and only if* $\mathcal{F} > 1$*.*

*2. The case RL(I) is possible if and only if* $\mathcal{F} \leq 1$*.*

The next proposition describes the Stackelberg equilibrium in the case of the retailer's leadership. Let

$$\Delta = \sqrt{\frac{F}{\beta-\gamma}} > 0, \quad \varepsilon = \frac{\beta-\gamma}{\gamma} > 0, \tag{2.8}$$

$$f = \sqrt{F\cdot(\beta-\gamma)} > 0, \quad D = \frac{\alpha - d - d_{\mathcal{R}}}{\sqrt{F\cdot(\beta-\gamma)}}. \tag{2.9}$$

**Proposition 2.** *In the cases RL and RL(I), the equilibrium demand* $q$*, wholesale price* $p$*, trade mark-up* $r$*, mass of firms* $N$*, and the retailer's profit* $\pi_{\mathcal{R}}$ *are presented in Tables 1 and 2, where* $\mathcal{F}, \Delta, \varepsilon, f,$ *and* $D$ *are defined in (2.7)–(2.9).*

Table 1. Equilibrium in different cases of the retailer's leadership

| | $q$ | $p$ | $r$ | $N$ |
|---|---|---|---|---|
| RL | $\Delta\sqrt{\mathcal{F}}$ | $d + f\sqrt{\mathcal{F}}$ | $d_{\mathcal{R}} + f\cdot\frac{D}{2}$ | $\frac{\varepsilon}{2}\cdot\left(\frac{D}{\sqrt{\mathcal{F}}} - 4\right)$ |
| RL(I) | $\Delta$ | $d + f$ | $d_{\mathcal{R}} + f\cdot\left(\frac{D}{2} + \mathcal{F} - 1\right)$ | $\frac{\varepsilon}{2}\cdot(D - 2\mathcal{F} - 2)$ |

Table 2. The retailer's profit in different cases of the retailer's leadership

| | $\pi_{\mathcal{R}}$ |
|---|---|
| RL | $\left(D - 4\sqrt{\mathcal{F}}\right)^2\cdot\frac{H}{2}$ |
| RL(I) | $(D - 2\mathcal{F} - 2)^2\cdot\frac{H}{2}$ |

### 2.1. Entrance fee

The relationship between producers and retailers is actually regulated. As a rule, the producer must pay the retailer. Let us denote the entrance fee by $F_{EF}$. Then the fixed costs of the producer and the retailer will change as follows:

$$\breve{F} = F + F_{EF},$$
$$\breve{F_{\mathcal{R}}} = F_{\mathcal{R}} - F_{EF}.$$

Taking into account $F_{EF}$, we write the profit of the $i$th producer as

$$\pi_{\mathcal{M}}(i) = (p(i) - d)q(i) - (F + F_{EF})$$

and the retailer's profit as

$$\pi_{\mathcal{R}} = \int_0^N (r(i) - d_{\mathcal{R}})\, q(i)di - \int_0^N (F_{\mathcal{R}} - F_{EF})di.$$

We get the following retailer's profit optimization problem:

$$\begin{cases} \pi_{\mathcal{R}} = \displaystyle\int_0^N (r(i) - d_{\mathcal{R}})\, q(i)di - \int_0^N (F_{\mathcal{R}} - F_{EF})di \;\to\; \max_{\mathbf{r},N,F_{EF}}, \\ \pi_{\mathcal{M}}(i) = (p(i) - d)q(i) - (F + F_{EF}) \geq 0. \end{cases} \tag{2.10}$$

In what follows, we will need the following lemmas.

**Lemma 1.** *The optimal trade mark-up $r$ is the same for all producers and is expressed in terms of $N$ as follows:*

$$r = r(N) = \frac{N(\alpha - d - d_{\mathcal{R}})}{2(N + \varepsilon)} + d_{\mathcal{R}}.$$

To prove Lemma 1, it is necessary to solve the optimization problem (2.10). Due to the fact that

$$\frac{\partial \pi_{\mathcal{R}}}{\partial F_{EF}} = N > 0,$$

$F_{EF}$ the optimum of the objective function is reached at the boundary, that is, $\pi_{\mathcal{M}}(i) = 0$. Substituting $F_{EF}$ into $\pi_{\mathcal{R}}$, we find the maximum of the function $\pi_{\mathcal{R}}$ over the variables $r$ and $N$. We solve the optimization problem by the method of needle variations; as a result, we determine the optimal mark-up of the retailer.

**Lemma 2.** *Under the symmetric equilibrium, the wholesale price $p$ and the demand $q$ are as follows:*

$$p = p(N) = \frac{\varepsilon(\alpha - d - d_{\mathcal{R}})}{2(N + \varepsilon)} + d,$$

$$q = q(N) = \frac{(\alpha - d - d_{\mathcal{R}})}{2\gamma(N + \varepsilon)} + d.$$

P r o o f. Under the symmetric equilibrium, solving the problem of the producer's profit maximization, we have

$$q(r, N) = \frac{(b + cN)(a - b(r + d))}{2b + cN}, \tag{2.11}$$

$$p(r,N)=\frac{q(r,N)}{b+cN}+d. \tag{2.12}$$

Note that

$$\frac{a}{b}=\alpha,\quad \frac{b}{c}=\frac{\beta-\gamma}{\gamma}=\varepsilon,\quad b+cN=\frac{1}{\beta-\gamma}.$$

Hence, due to Lemma 1, (2.11) is

$$\begin{aligned} q(N)=\frac{\alpha-r-d}{\gamma(N+2\varepsilon)}&=\frac{\alpha-\dfrac{N(\alpha-d-d_{\mathcal{R}})}{2(N+\varepsilon)}-d_{\mathcal{R}}-d}{\gamma(N+2\varepsilon)}=\\ &=\frac{(\alpha-d-d_{\mathcal{R}})(2N+2\varepsilon-N)}{2\gamma(N+\varepsilon)(N+2\varepsilon)}=\frac{\alpha-d-d_{\mathcal{R}}}{2\gamma(N+\varepsilon)}. \end{aligned} \tag{2.13}$$

Substituting (2.13) into (2.12), we get

$$p(N)=\frac{(\alpha-d-d_{\mathcal{R}})\varepsilon}{2(N+\varepsilon)}+d.$$

□

**Proposition 3.** *When an entrance fee is introduced, the equilibrium demand $q$, wholesale price $p$, trade mark-up $r$, entrance fee $F_{EF}$, mass of firms $N$, and retailer profit $\pi_{\mathcal{R}}$ are presented in Tables 3 and 4, where $\mathcal{F},\Delta,\varepsilon,f,$ and $D$ are defined in* (2.7)–(2.9)*.*

Table 3. Equilibrium in the case of introduction of an entrance fee

| | $q$ | $p$ | $r$ |
|---|---|---|---|
| EF | $\Delta\sqrt{2\mathcal{F}+1}$ | $f\sqrt{2\mathcal{F}+1}+d$ | $d_{\mathcal{R}}+f\cdot\dfrac{D}{2}-f\sqrt{2\mathcal{F}+1}$ |

Table 4. The retailer's fixed costs, mass of firms, and the retailer's profit with an entrance fee

| | $F_{EF}$ | $N$ | $\pi_{\mathcal{R}}$ |
|---|---|---|---|
| EF | $F_{\mathcal{R}}$ | $\dfrac{\varepsilon}{2}\cdot\left(\dfrac{D}{\sqrt{2\mathcal{F}+1}}-2\right)$ | $(D-2\sqrt{2\mathcal{F}+1})^2\cdot\dfrac{H}{2}$ |

P r o o f. From Lemmas 1 and 2, we get the following simplified retailer's profit optimization problem:

$$\begin{cases} \pi_{\mathcal{R}}=N\left((r(N)-d_{\mathcal{R}})q(N)-F_{\mathcal{R}}+F_{EF}\right)\ \to\ \max\limits_{N,F_{EF}}, \\ \pi_{\mathcal{M}}=(p(N)-d)\,q(N)-(F+F_{EF})\geq 0. \end{cases}$$

Note that

$$\frac{\partial\pi_{\mathcal{R}}}{\partial F_{EF}}=N>0.$$

Therefore, the optimum of the objective function is attained at the boundary, i.e., for $\pi_{\mathcal{M}} = 0$ (the entrance fee condition is fulfilled), whence we find

$$F_{EF} = F_{EF}(N) = -F + (p(N) - d)q(N).$$

Substituting $F_{EF} = F_{EF}(N)$ into $\pi_{\mathcal{R}}$, we get

$$\pi_{\mathcal{R}} = N\left((r(N) - d_{\mathcal{R}})q(N) - F_{\mathcal{R}} - F + (p(N) - d)q(N)\right) \ \to \ \max_{N}.$$

Since the derivative of the retailer's profit equals zero, we determine the mass of producers:

$$\frac{\partial \pi_{\mathcal{R}}}{\partial N} = 0 \Leftrightarrow N = \frac{\varepsilon}{2}\left(\frac{D}{\sqrt{2\mathcal{F}+1}} - 2\right).$$

Then we find

$$q = \Delta\sqrt{2\mathcal{F}+1}, \quad p = f\sqrt{2\mathcal{F}+1} + d,$$
$$r = d_{\mathcal{R}} + f\cdot\frac{D}{2} - f\sqrt{2\mathcal{F}+1}, \quad F_{EF} = F_{\mathcal{R}}, \quad \pi_{\mathcal{R}} = (D - 2\sqrt{2\mathcal{F}+1})^2\cdot\frac{H}{2}.$$

□

## 3. Social welfare and consumer surplus

In this section, we consider the functions of social welfare and consumer surplus and calculate the equilibrium social welfare and equilibrium consumer surplus in two cases: under the retailer's leadership and with an entrance fee.

### 3.1. Social welfare

Consider the social welfare function $W$ as a measure of the welfare of society. In the symmetric case, $W$ has the form

$$W = (\alpha - d - d_{\mathcal{R}})Nq - \frac{\beta-\gamma}{2}\cdot Nq^2 - \frac{\gamma}{2}\cdot N^2q^2 - (F + F_{\mathcal{R}})N. \tag{3.1}$$

In various equilibrium cases, we can formulate the following proposition for the social welfare function $W$.

**Proposition 4.** *The equilibrium social welfare under the retailer's leadership and with an entrance fee is presented in Table 5, where*

$$H = \frac{F\cdot(\beta-\gamma)}{2\gamma} > 0, \tag{3.2}$$

*while $\mathcal{F}, f$, and $D$ are defined in (2.7) and (2.9).*

We can prove Proposition 4 directly by substituting the equilibrium solutions from Proposition 2 and Proposition 3 into (3.1). After appropriate calculations, we get the results presented in Table 5.

Table 5. Social welfare in different equilibrium cases

| | $W$ |
|---|---|
| RL | $\left(D-4\sqrt{\mathcal{F}}\right)\cdot\left(\frac{3}{4}\cdot(D-2\sqrt{\mathcal{F}})-\frac{1}{\sqrt{\mathcal{F}}}\right)\cdot H$ |
| RL(I) | $(D-2\mathcal{F}-2)\cdot\left(\frac{3}{4}\cdot(D-2\mathcal{F})-1\right)\cdot H$ |
| EF | $\left(D-2\sqrt{2\mathcal{F}+1}\right)\cdot\left(3D-4\sqrt{2\mathcal{F}+1}\right)\cdot\frac{H}{4}$ |

### 3.2. Consumer surplus

The consumer surplus $CS$ is a measure of the well-being that consumers derive from the consumption of goods and services. In the case of symmetric equilibrium, it is represented in the form

$$CS=\alpha Nq-\frac{\beta-\gamma}{2}Nq^2-\frac{\gamma}{2}N^2q^2-(p+r)Nq. \tag{3.3}$$

For the consumer surplus function, for the various equilibrium cases, the following proposition can be formulated.

**Proposition 5.** *The equilibrium consumer surplus under the retailer's leadership and with an entrance fee is presented in Table 6, where $\mathcal{F}, f, D$, and $H$ are defined in (2.7), (2.9), and (3.2).*

Table 6. Consumer surplus in different equilibrium cases

| | $CS$ |
|---|---|
| RL | $\left(D-4\sqrt{\mathcal{F}}\right)\cdot\left(D-2\sqrt{\mathcal{F}}\right)\cdot\frac{H}{4}$ |
| RL(I) | $(D-2\mathcal{F}-2)\cdot(D-2\mathcal{F})\cdot\frac{H}{4}$ |
| EF | $D\cdot\left(D-2\sqrt{2\mathcal{F}+1}\right)\cdot\frac{H}{4}$ |

We can prove Proposition 5 directly by substituting the equilibrium solutions from Proposition 2 and Proposition 3 into (3.3). After appropriate calculations, we get the results presented in Table 6.

## 4. Comparison of RL and EF cases

In this section, we compare the obtained values $p, q, r, \pi_{\mathcal{R}}, W$, and $CS$ in the case of the retailer's leadership and in the case of an entrance fee (see Tables 1–6).

We get the following result, where the indices "RL" and "EF" mean that the corresponding values are calculated for the case of the retailer's leadership and for the case of an entrance fee, respectively.

**Proposition 6.** *For the equilibrium price $p$, mark-up $r$, retail price $p+r$, individual consumption $q$, total consumption $Q$, welfare $W$, consumer surplus $CS$, and the retailer's profit $\pi_{\mathcal{R}}$, we get*

- $p^{EF} > p^{RL}$, *i.e., the introduction of an entrance fee always increases the wholesale price, thereby offsetting the costs of the producer*;
- $r^{EF} < r^{RL}$, *i.e., the introduction of an entrance fee reduces the trade mark-up*;
- $p^{EF} + r^{EF} < p^{RL} + r^{RL}$, *i.e., the introduction of an entrance fee an entails a decrease in the retail price*;
- $q^{EF} > q^{RL}$, *i.e., the introduction of an entrance fee entails an increase in the individual consumption*;
- $Q^{EF} > Q^{RL}$, *where* $Q = qN$, *i.e., the introduction of an entrance fee increases the total consumption*;
- $W^{EF} > W^{RL}$, $CS^{EF} > CS^{RL}$, $\pi_{\mathcal{R}}^{EF} > \pi_{\mathcal{R}}^{RL}$, *i.e., the introduction of an entrance fee leads to an increase in the social welfare, consumer surplus, and the retailer's profit.*

P r o o f. Let us prove that $CS^{EF} \geq CS^{RL}$ (the rest can be proven in a similar way). In the case when $\mathcal{F} > 1$, we have

$$\begin{cases} N^{EF} & \geq & 0, \\ N^{RL} & \geq & 0, \end{cases} \quad \Leftrightarrow \quad \begin{cases} D & \geq & 4\sqrt{\mathcal{F}}, \\ D & \geq & 2\sqrt{2\mathcal{F}+1}. \end{cases}$$

Note that $2\sqrt{\mathcal{F}} \geq \sqrt{2\mathcal{F}+1}$. Hence $D \geq 4\sqrt{\mathcal{F}}$. Then

$$\begin{aligned} CS^{EF} - CS^{RL} &= \frac{H}{4} \cdot \Big( D(D - 2\sqrt{2\mathcal{F}+1}) - (D - 4\sqrt{\mathcal{F}})(D - 2\sqrt{\mathcal{F}}) \Big) = \\ &= \frac{H}{4} \cdot \Big( D^2 - 2D\sqrt{2\mathcal{F}+1} - D^2 + 6D\sqrt{\mathcal{F}} - 8\mathcal{F} \Big) = \\ &= \frac{H}{4} \cdot \left( D(\underbrace{6\sqrt{\mathcal{F}} - 2\sqrt{2\mathcal{F}+1}}_{\geq 0}) - 8\mathcal{F} \right) \geq \\ &\geq \frac{H}{4} \cdot \Big( 4\sqrt{\mathcal{F}}(6\sqrt{\mathcal{F}} - 2\sqrt{2\mathcal{F}+1}) - 8\mathcal{F} \Big) = \\ &= 2H \cdot \Big( \sqrt{\mathcal{F}}(3\sqrt{\mathcal{F}} - \sqrt{2\mathcal{F}+1}) - \mathcal{F} \Big) = \\ &= 2H \cdot \sqrt{\mathcal{F}} \Big( 2\sqrt{\mathcal{F}} - \sqrt{2\mathcal{F}+1} \Big) \geq 0, \end{aligned}$$

i.e., we get $CS^{EF} > CS^{RL}$ for $\mathcal{F} > 1$.

In the case when $0 < \mathcal{F} \leq 1$, we have

$$\begin{cases} N^{EF} & \geq & 0, \\ N^{RL} & \geq & 0, \end{cases} \quad \Leftrightarrow \quad \begin{cases} D & \geq & 2(\mathcal{F}+1), \\ D & \geq & 2\sqrt{2\mathcal{F}+1}. \end{cases}$$

Note that $\mathcal{F} + 1 \geq \sqrt{2\mathcal{F}+1}$. Hence $D \geq 2(\mathcal{F}+1)$. Then

$$\begin{aligned} CS^{EF} - CS^{RL} &= \frac{H}{4} \cdot \Big( D(D - 2\sqrt{2\mathcal{F}+1}) - (D - 2\mathcal{F} - 2)(D - 2\mathcal{F}) \Big) = \\ &= \frac{H}{4} \cdot \Big( D^2 - 2D\sqrt{2\mathcal{F}+1} - D^2 + 2D\mathcal{F} + (D - 2\mathcal{F})(2\mathcal{F}+2) \Big) = \end{aligned}$$

$$
\begin{aligned}
&= \frac{H}{4} \cdot \Big( D(\underbrace{4\mathcal{F} + 2 - 2\sqrt{2\mathcal{F}+1}}_{\geq 0}) - 2\mathcal{F}(2\mathcal{F}+2) \Big) \geq \\
&\geq \frac{H}{4} \cdot \Big( 2(\mathcal{F}+1)(4\mathcal{F} + 2 - 2\sqrt{2\mathcal{F}+1}) - 2\mathcal{F}(2\mathcal{F}+2) \Big) = \\
&= \frac{H}{4} \cdot 4 \Big( (\mathcal{F}+1)(2\mathcal{F} + 1 - \sqrt{2\mathcal{F}+1}) - \mathcal{F}(\mathcal{F}+1) \Big) = r \\
&= H \cdot (\mathcal{F}+1) \left( \mathcal{F} + 1 - \sqrt{2\mathcal{F}+1} \right) \geq 0,
\end{aligned}
$$

i.e., $CS^{EF} > CS^{RL}$ for $0 < \mathcal{F} \leq 1$. □

As for the values $N^{EF}$ and $N^{RL}$, there are two possibilities depending on the values $\mathcal{F}$ and $D$, see Fig. 1. In Fig. 1, only the areas $N^{EF} < N^{RL}$ and $N^{EF} > N^{RL}$ are of interest, since the number

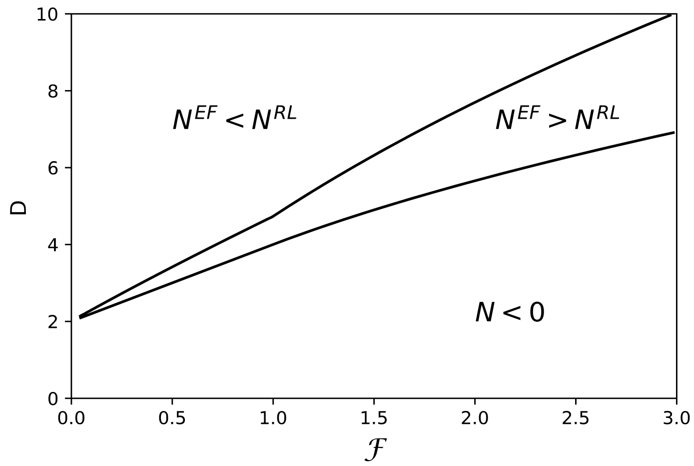



Figure 1. Comparison of values $N$ for $RL$ and $EF$ cases.

(mass) of producers in these areas is non-negative. An analytical representation of these areas is

$$
\begin{aligned}
&N^{EF} < N^{RL}, \quad D > \underline{D}, \\
&N^{EF} > N^{RL}, \quad \overline{D} > D > \underline{D},
\end{aligned}
$$

where

$$
\underline{D} = \begin{cases} 4\sqrt{\mathcal{F}} & \text{if } \mathcal{F} > 1; \\ 2(\mathcal{F}+1) & \text{if } \mathcal{F} \leq 1, \end{cases}
$$

$$
\overline{D} = \begin{cases} \left(1 + \sqrt{\dfrac{1}{\mathcal{F}} + 2}\right) \cdot \sqrt{1+2\mathcal{F}} \cdot \dfrac{2\mathcal{F}}{1+\mathcal{F}} & \text{if } \mathcal{F} > 1; \\ \left(1 + \sqrt{1+2\mathcal{F}}\right) \cdot \sqrt{1+2\mathcal{F}} & \text{if } \mathcal{F} \leq 1. \end{cases}
$$

## 5. Conclusion

The presented paper analyzes the monopolistic competition trade model with two-level interaction. The situation of the retailer's leadership is considered in detail. We show that, under the retailer's leadership, two ways are possible depending on $\mathcal{F}$: artificially restricted and unrestricted market. The parameter $\mathcal{F}$ is the ratio of the retailer's fixed costs to the twice fixed costs of each producer. In the case of an artificially limited market, the retailer independently restricts the entry of producers. Otherwise (i.e., in the case of an unrestricted market), the free entry condition is used, which means that producers enter the market until their profits become zero.

In addition, we study the possible effects when the retailer imposes the market entrance fee for producers. We show that the introduction of an entrance fee by the retailer is justifiable since it increases the social welfare and consumer surplus, as well as the retailer's profit.

## Acknowledgements

The author thanks scientific supervisor I.A. Bykadorov for helpful discussions. The study was partially carried out within the framework of the state assignments of the Sobolev Institute of Mathematics SB RAS (project FWNF-2022-0019).

## REFERENCES

1. Bykadorov I. Monopolistic competition with investments in productivity. *Optim. Lett.*, 2019. Vol. 13, No. 8. P. 1803–1817. DOI: 10.1007/s11590-018-1336-9
2. Bykadorov I., Ellero A., Funari S., Kokovin S., Pudova M. Chain Store Against Manufacturers: Regulation Can Mitigate Market Distortion. In: *Int. Conf. Discrete Optimization and Operations Research (DOOR-2016), Yu. Kochetov et all (eds.).* Lecture Notes in Comput. Sci., vol. 9869. Cham: Springer, 2016. P. 480–493. DOI: 10.1007/978-3-319-44914-2_38
3. Bykadorov I., Kokovin S. Can a larger market foster R&D under monopolistic competition with variable mark-ups? *Res. Econ.*, 2017. Vol. 71, No. 4. P. 663–674. DOI: 10.1016/j.rie.2017.10.006
4. Bykadorov I. A., Kokovin S. G. Effectiveness of retailer market power: monopolistic competition of producers. *Vestnik NGUEU* [Bulletin NSUEM], 2014. No. 1. P. 326–337. (in Russian)
5. Bykadorov I. A., Kokovin S. G., Zhelobodko E. V. Product diversity in a vertical distribution channel under monopolistic competition. *Autom. Remote Control*, 2014. Vol. 75, No. 8. P. 1503–1524. DOI: 10.1134/S0005117914080141
6. Dixit A. K. Vertical integration in a monopolistically competitive industry. *Int. J. Industrial Organization*, 1983. Vol. 1. P. 63–78. DOI: 10.1016/0167-7187(83)90023-1
7. Dixit A. K., Stiglitz J. E. *Monopolistic Competition and Optimum Product Diversity.* Working or Discussion Paper, no. 67. Warwick: University of Warwick, 1975. 54 p.
8. Gabszewicz J. J., Thisse J.-F. On the nature of competition with differentiated products. *J. Econom. Theory*, 1986. Vol. 96, No. 381. P. 160–172. DOI: 10.2307/2233431
9. Hotelling H. Stability in competition. *Econom. J.*, 1929. Vol. 39, No. 153. P. 41–57. DOI: 10.2307/2224214
10. Ottaviano G. I. P., Tabuchi T. and Thisse J.-F. Agglomeration and trade revisited. *Int. Econom. Review*, 2002. Vol. 43, No. 2. P. 409–436. DOI: 10.1111/1468-2354.t01-1-00021
11. Perry M. K., Groff R. H. Resale price maintenance and forward integration into a monopolistically competitive industry. *Quart. J. Econom.*, 1985. Vol. 100, No. 4. P. 1293–1311. DOI: 10.2307/1885684
12. Salop S. C. Monopolistic competition with outside goods. *Bell J. Econom.*, 1979. Vol. 10, No. 1. P. 141–156. DOI: 10.2307/3003323
13. Tilzo O., Bykadorov I. Monopolistic competition model with retailing. In: *Commun. Comput. Inf. Sci., vol. 1275: Mathematical Optimization Theory and Operations Research (MOTOR 2020), Kochetov Y., Bykadorov I., Gruzdeva T. (eds)*, 2020. P. 287–301. DOI: 10.1007/978-3-030-58657-7_24
14. Tilzo O., Bykadorov I. Retailing under monopolistic competition: a comparative analysis. In: *15th Int. Asian School-Seminar Optimization Problems of Complex Systems (OPCS), 26–30 August 2019, Novosibirsk, Russia.* IEEE Xplore, 2019. P. 156–161. DOI: 10.1109/OPCS.2019.8880233

**URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 128–135**

DOI: 10.15826/umj.2022.1.011

# HANKEL DETERMINANT OF CERTAIN ORDERS FOR SOME SUBCLASSES OF HOLOMORPHIC FUNCTIONS

**D. Vamshee Krishna**

Department of Mathematics, Gitam School of Science, GITAM University,
Visakhapatnam – 530 045, A.P., India
vamsheekrishna1972@gmail.com

**D. Shalini**

Department of Mathematics, Dr. B. R. Ambedkar University,
Srikakulam – 532 410, A.P., India
shaliniraj1005@gmail.com

**Abstract:** In this paper, we are introducing certain subfamilies of holomorphic functions and making an attempt to obtain an upper bound (UB) to the second and third order Hankel determinants by applying certain lemmas, Toeplitz determinants, for the normalized analytic functions belong to these classes, defined on the open unit disc in the complex plane. For one of the inequality, we have obtained sharp bound.

**Keywords:** Holomorphic function, Upper bound, Hankel determinant, Positive real function.

## 1. Introduction

Let $\mathcal{A}$ represent a family of mappings $f$ of the type

$$f(z) = z + \sum_{t=2}^{\infty} a_t z^t$$

in the open unit disc

$$\mathcal{U} = \{z \in \mathbb{C} : 1 > |z|\},$$

and $\mathcal{S}$ is the subfamily of $\mathcal{A}$, possessing univalent (schlicht) mappings. Pommerenke [17] characterized the $r^{th}$-Hankel determinant of order $n$, for $f$ with $r,\ n \in \mathbb{N}$, namely

$$H_{r,n}(f) = \begin{vmatrix} a_n & a_{n+1} & \cdots & a_{n+r-1} \\ a_{n+1} & a_{n+2} & \cdots & a_{n+r} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n+r-1} & a_{n+r} & \cdots & a_{n+2r-2} \end{vmatrix} \quad (a_1 = 1). \tag{1.1}$$

The Fekete–Szegö functional [7] is obtained for $r = 2$ and $n = 1$ in (1.1), denoted by $H_{2,1}(f)$. Further, sharp bounds to the functional $|H_{2,2}(f)|$, obtained for $r = 2$ and $n = 2$ in (1.1), are called as Hankel determinant of order two, given by

$$H_{2,2}(f) = \begin{vmatrix} a_2 & a_3 \\ a_3 & a_4 \end{vmatrix} = a_2 a_4 - a_3^2.$$

In recent years, the estimation of an upper bound (UB) to $|H_{2,2}(f)|$ was studied by many authors. The exact estimates of $|H_{2,2}(f)|$ for the functions namely, bounded turning, starlike and convex functions, each one is a subfamily of $\mathcal{S}$, symbolized as $\mathcal{R}$, $S^*$ and $\mathcal{K}$ respectively and fulfilling the conditions

$$\operatorname{Re} f'(z) > 0, \quad \operatorname{Re}\left\{\frac{zf'(z)}{f(z)}\right\} > 0, \quad \operatorname{Re}\left\{1 + \frac{zf''(z)}{f'(z)}\right\} > 0$$

in the unit disc $\mathcal{U}$, were proved by Janteng et al. [9, 10] and the derived bounds are 4/9, 1 and 1/8 respectively. Choosing $r = 2$ and $n = p+1$ in (1.1), we obtain Hankel determinant of second order for the $p$-valent function (see [20]), given by

$$H_{2,(p+1)}(f) = \begin{vmatrix} a_{p+1} & a_{p+2} \\ a_{p+2} & a_{p+3} \end{vmatrix} = a_{p+1}a_{p+3} - a_{p+2}^2.$$

The case $r = 3$ seems to be much tough than $r = 2$. Few papers were devoted for the study of third order Hankel determinant denoted as $H_{3,1}(f)$, with $r = 3$ and $n = 1$ in (1.1), namely

$$H_{3,1}(f) = \begin{vmatrix} a_1 = 1 & a_2 & a_3 \\ a_2 & a_3 & a_4 \\ a_3 & a_4 & a_5 \end{vmatrix}.$$

Calculating the determinant, we have

$$H_{3,1}(f) = a_1(a_3a_5 - a_4^2) + a_2(a_3a_4 - a_2a_5) + a_3(a_2a_4 - a_3^2). \tag{1.2}$$

The concept of estimation of an upper bound for $H_{3,1}(f)$ was firstly introduced and studied by Babalola [3], who tried to estimate this functional in the classes $\mathcal{R}$, $S^*$ and $\mathcal{K}$, his results are as follows

(i) $f \in S^* \Rightarrow |H_{3,1}(f)| \leq 16$;

(ii) $f \in \mathcal{K} \Rightarrow |H_{3,1}(f)| \leq 0.714$;

(iii) $f \in \mathcal{R} \Rightarrow |H_{3,1}(f)| \leq 0.742$.

As a result of the paper by Babalola [3], mach research associated with the Hankel determinant of order 3 and 4, for specific subfamilies of holomorphic functions have been done (see [1–5, 11, 12, 15, 18, 19]). Motivated by the results obtained by the indicated authors, here we make an attempt to derive an upper bound to $|H_{2,3}(f)| = a_3a_5 - a_4^2$, $|H_{3,1}(f)|$, when $f$ belongs to the following new subfamilies of holomorphic functions.

**Definition 1.** *A function $f(z) \in \mathcal{A}$ is said to be in the class $\mathcal{R}_b(\alpha)$, where $b \neq 0$ is a real number with $\alpha$ $(0 \leq \alpha < 1)$, if it satisfies the condition*

$$\operatorname{Re}\left(1 - \frac{2}{b} + \frac{2}{b}f'(z)\right) > \alpha, \quad z \in \mathcal{U}.$$

It is observed that for $b = 2$ and for the values $b = 2$, $\alpha = 0$, we have $\mathcal{R}(\alpha)$, the class consisting of functions whose derivative has positive real part of order $\alpha$ $(0 \leq \alpha < 1)$ and $\mathcal{R}$ respectively.

**Definition 2.** *A function $f(z) \in \mathcal{A}$ is said to be in the class $S_b^*(\alpha)$, where $b$ is a non-zero real number with $\alpha$ $(0 \leq \alpha < 1)$, if it satisfies the condition*

$$\operatorname{Re}\left(1 - \frac{2}{b} + \frac{2}{b}\left(\frac{zf'(z)}{f(z)}\right)\right) > \alpha, \quad z \in \mathcal{U}.$$

For the values $b = 2$ and $b = 2,\ \alpha = 0$, $S_b^*(\alpha)$ reduces to $S^*(\alpha)$, class consisting of starlike functions of order $\alpha\,(0 \leq \alpha < 1)$ and $S^*$ respectively.

**Definition 3.** *A function $f(z) \in \mathcal{A}$ is said to be in the class $\mathcal{K}_b(\alpha)$, where $b \neq 0$ is a real number with $\alpha\,(0 \leq \alpha < 1)$, if it satisfies the condition*

$$\operatorname{Re}\left(1 - \frac{2}{b} + \frac{2}{b}\left(1 + \frac{zf''(z)}{f'(z)}\right)\right) > \alpha, \quad z \in \mathcal{U}.$$

In particular for $b = 2$ and for the values $b = 2,\ \alpha = 0$, $\mathcal{K}_b(\alpha)$ reduces to $\mathcal{K}(\alpha)$, the class consisting of convex functions of order $\alpha\,(0 \leq \alpha < 1)$ and $\mathcal{K}$ respectively.

In proving our results, the following sharp estimates are needed, which are in the form of Lemmas hold good for functions possessing positive real part. Define the collection $\mathcal{P}$ of all functions $g$, each one called as Carathéodory function [6] of the form

$$g(z) = 1 + \sum_{t=1}^{\infty} c_t z^t,$$

which is holomorphic in $\mathcal{U}$ and $\operatorname{Re} g(z) > 0$ for $z \in \mathcal{U}$.

**Lemma 1** [8]. *If $g \in \mathcal{P}$, then the estimate $|c_i - \mu c_j c_{i-j}| \leq 2$ holds for $i, j \in \mathbb{N}$, with $i > j$ and $\mu \in [0, 1]$.*

**Lemma 2** [14]. *If $g \in \mathcal{P}$, then the estimate $|c_i - c_j c_{i-j}| \leq 2$ holds for $i, j \in \mathbb{N}$, with $i > j$.*

**Lemma 3** [16]. *If $g \in \mathcal{P}$, then $|c_t| \leq 2$, for $t \in \mathbb{N}$, equality occurs for the function*

$$h(z) = \frac{1+z}{1-z}, \quad z \in \mathcal{U}.$$

**Lemma 4** [21]. *If $g \in \mathcal{P}$, then $|c_2 c_4 - c_3^2| \leq 4 - 1/2 \cdot |c_2|^2 + 1/4 \cdot |c_2|^3$.*

In order to procure our results, we adopt the procedure framed through Libera and Zlotkiewicz [13].

## 2. Main results

**Theorem 1.** *If*

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{R}_b(\alpha),$$

*where $b$ is any real number with $0 < b \leq 1/(1-\alpha)$, for $0 \leq \alpha < 1$ then*

$$|H_{3,1}(f)| \leq \frac{41 b^2 (1-\alpha)^2}{240}.$$

P r o o f. For

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{R}_b(\alpha),$$

by virtue of Definition 1, we have

$$\frac{b(1-\alpha) + 2\{f'(z) - 1\}}{b(1-\alpha)} = g(z) \Leftrightarrow b(1-\alpha) + 2\left\{f'(z) - 1\right\} = b(1-\alpha)g(z). \tag{2.1}$$

Using the series representations for $f'(z)$ and $g(z)$ in (2.1), after simplifying, we get

$$a_n = \frac{tc_{n-1}}{2n}, \quad \text{where} \quad t = b(1-\alpha), \quad n \geq 2. \tag{2.2}$$

Putting the values of $a_i$, for $i \in \{2,3,4,5\}$ from (2.2), in $H_{3,1}(f)$, given in (1.2), we have

$$H_{3,1}(f) = t^2 \left[ \frac{c_2c_4}{60} - \frac{tc_2^3}{216} - \frac{c_3^2}{64} - \frac{tc_1^2c_4}{160} + \frac{tc_1c_2c_3}{96} \right]. \tag{2.3}$$

On grouping the terms in the expression (2.3), we obtain

$$\begin{aligned} H_{3,1}(f) = t^2 \Bigg[ & \frac{tc_4(c_2 - c_1^2)}{160} - \frac{c_3}{64}\left(c_3 - \frac{tc_1c_2}{2}\right) + \frac{tc_2(c_4 - c_2^2)}{216} \\ & - \frac{c_2}{192}\left(c_4 - \frac{tc_1c_3}{2}\right) + \frac{(189 - 94t)c_2c_4}{8640} \Bigg]. \end{aligned} \tag{2.4}$$

Applying the triangle inequality in (2.4), we get

$$\begin{aligned} \left|H_{3,1}(f)\right| \leq t^2 \Bigg[ & \frac{t|c_4||(c_2 - c_1^2)|}{160} + \frac{|c_3|}{64}\left|c_3 - \frac{tc_1c_2}{2}\right| + \frac{t|c_2||c_4 - c_2^2|}{216} \\ & + \frac{|c_2|}{192}\left|c_4 - \frac{tc_1c_3}{2}\right| + \frac{(189 - 94t)|c_2||c_4|}{8640} \Bigg]. \end{aligned} \tag{2.5}$$

Upon using the Lemmas 1–3 in the inequality (2.5), we obtain

$$|H_{3,1}(f)| \leq \frac{41t^2}{240} = \frac{41b^2(1-\alpha)^2}{240}. \tag{2.6}$$

□

*Remark 1.* Choosing $b = 2$ and $\alpha = 0$ in the inequality (2.6), it coincides with the result obtained by Zaprawa [22].

**Theorem 2.** *If*

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{R}_b(\alpha),$$

*where $b$ is any real number with $0 < b \leq 1/(1-\alpha)$, for $0 \leq \alpha < 1$ then $|H_{2,3}(f)| \leq b^2(1-\alpha)^2/15$.*

P r o o f. Substituting the values of $a_3$, $a_4$, and $a_5$ from (2.2) in $H_{2,3}(f)$, we have

$$\begin{aligned} H_{2,3}(f) = a_3a_5 - a_4^2 = t^2 \left[ \frac{c_2c_4}{60} - \frac{c_3^2}{64} \right] &= t^2 \left[ \frac{c_2c_4}{60} - \frac{c_2c_4}{64} + \frac{c_2c_4}{64} - \frac{c_3^2}{64} \right] \\ &= t^2 \left[ \frac{c_2c_4 - c_3^2}{64} + \frac{c_2c_4}{960} \right], \quad \text{where} \quad t = b(1-\alpha). \end{aligned} \tag{2.7}$$

Applying the triangle inequality in (2.7) and then using the Lemmas 3 and 4, after simplifying, we get

$$|H_{2,3}(f)| = |a_3a_5 - a_4^2| \leq \frac{b^2(1-\alpha)^2}{15}. \tag{2.8}$$

□

*Remark 2.* Choosing $b = 2$ and $\alpha = 0$ in the inequality (2.8), it coincides with the result obtained by Zaprawa [21]. At this stage, the inequality in (2.8) becomes sharp for the function

$$g(z) = \frac{1+z^2}{1-z^2}.$$

**Theorem 3.** *If*

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in S_b^*(\alpha),$$

*where $b$ is any real number with $0 < b \leq 1/(1-\alpha)$, for $0 \leq \alpha < 1$ then*

$$|H_{3,1}(f)| \leq \left[\frac{b(1-\alpha)}{12}\right]^2 [34 + b(1-\alpha)].$$

P r o o f. For

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in S_b^*(\alpha),$$

from the Definition 2, we have

$$\frac{\{b(1-\alpha)-2\} f(z) + 2zf'(z)}{b(1-\alpha)f(z)} = g(z) \Leftrightarrow \{b(1-\alpha)-2\} f(z) + 2zf'(z) = b(1-\alpha)f(z)g(z) \quad (2.9)$$

Replacing $f(z)$, $f'(z)$ and $g(z)$ with their equivalent series expressions in (2.9) and applying the same procedure as we carried in Theorem 1, we obtain

$$\begin{aligned} a_2 = \frac{tc_1}{2}, \quad a_3 = \frac{t}{8}\left(2c_2 + tc_1^2\right), \quad a_4 = \frac{t}{48}\left(8c_3 + 6tc_1c_2 + t^2c_1^3\right), \\ a_5 = \frac{t}{384}\left(48c_4 + 32tc_1c_3 + 12tc_2^2 + 12t^2c_1^2c_2 + t^3c_1^4\right), \quad \text{where} \quad t = b(1-\alpha). \end{aligned} \quad (2.10)$$

Substituting the values of $a_2$, $a_3$, $a_4$, and $a_5$ from (2.10) in the functional given in (1.2), we get

$$\begin{aligned} H_{3,1}(f) = \left(\frac{t}{94}\right)^2 \Big[ - t^4c_1^6 + 6t^3c_1^4c_2 + 32t^2c_1^3c_3 - 36t^2c_1^2c_2^2 - 144tc_1^2c_4 \\ + 192tc_1c_2c_3 - 72tc_2^3 + 288c_2c_4 - 256c_3^2\Big]. \end{aligned} \quad (2.11)$$

On grouping the terms in (2.11), we have

$$\begin{aligned} H_{3,1}(f) = \left(\frac{t}{94}\right)^2 \Big[160\Big(c_2 - \frac{tc_1^2}{2}\Big)\Big(c_4 - \frac{tc_2^2}{2}\Big) + 8t\Big(c_2 - \frac{tc_1^2}{2}\Big)^3 + \\ 128\Big(c_2 - \frac{tc_1^2}{2}\Big)\Big(c_4 - \frac{tc_1c_3}{2}\Big) - 256\Big(c_3 - \frac{8tc_1c_2}{16}\Big)^2\Big]. \end{aligned} \quad (2.12)$$

On applying the triangle inequality in (2.12), we obtain

$$\begin{aligned} |H_{3,1}(f)| \leq \left(\frac{t}{94}\right)^2 \Big[160\Big|c_2 - \frac{tc_1^2}{2}\Big|\Big|c_4 - \frac{tc_2^2}{2}\Big| + 8t\Big|c_2 - \frac{tc_1^2}{2}\Big|^3 + \\ 128\Big|c_2 - \frac{tc_1^2}{2}\Big|\Big|c_4 - \frac{tc_1c_3}{2}\Big| + 256\Big|c_3 - \frac{8tc_1c_2}{16}\Big|^2\Big]. \end{aligned}$$

Further, the above inequality simplifies to

$$|H_{3,1}(f)| \leq \left(\frac{t}{12}\right)^2 [34+t] = \left[\frac{b(1-\alpha)}{12}\right]^2 [34+b(1-\alpha)]. \tag{2.13}$$

□

*Remark 3.* Choosing $b=2$ and $\alpha=0$ in the inequality (2.13), we see that it coincides with that of Zaprawa [22].

**Theorem 4.** *If*

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{K}_b(\alpha),$$

*where $b$ is any real number with $0 < b \leq 1/(1-\alpha)$, $0 \leq \alpha < 1$ then*

$$|H_{3,1}(f)| \leq \left[\frac{b(1-\alpha)}{12\sqrt{15}}\right]^2 [33+8b(1-\alpha)].$$

P r o o f. For

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{K}_b(\alpha),$$

from Definition 3, we have

$$\frac{\{b(1-\alpha)-2\} f(z) + 2zf'(z)}{b(1-\alpha)f(z)} = g(z) \Leftrightarrow \{b(1-\alpha)-2\} f(z) + 2zf'(z) \; = b(1-\alpha)f(z)g(z).$$

Applying the same procedure as we did in Theorem 1, we obtain

$$a_2 = \frac{tc_1}{4}, \quad a_3 = \frac{t}{24}\left(2c_2 + tc_1^2\right), \quad a_4 = \frac{t}{192}\left(8c_3 + 6tc_1c_2 + t^2c_1^3\right),$$
$$a_5 = \frac{t}{1920}\left(48c_4 + 32tc_1c_3 + 12tc_2^2 + 12t^2c_1^2c_2 + t^3c_1^4\right), \quad \text{where} \quad t = b(1-\alpha).$$

Further, we have

$$H_{3,1}(f) = \frac{t^2}{552960}\Big[ -t^4c_1^6 + 12t^3c_1^4c_2 + 48t^2c_1^3c_3 - 84t^2c_1^2c_2^2 - 288tc_1^2c_4$$
$$+288tc_1c_2c_3 - 32tc_2^3 + 1152c_2c_4 - 960c_3^2\Big].$$

On grouping the suitable terms in the above expression, we have

$$\begin{aligned} H_{3,1}(f) = \frac{t^2}{552960}\Bigg[&64t\left(c_2 - \frac{tc_1^2}{4}\right)^3 + 384c_4\left(c_2 - \frac{tc_1^2}{2}\right) + 576c_2\left(c_4 - \frac{tc_2^2}{2}\right) \\ &+192\left(c_2 - \frac{tc_1^2}{2}\right)\left(c_4 - \frac{tc_1c_3}{2}\right) - 960c_3\left(c_3 - \frac{2tc_1c_2}{5}\right) + 192tc_2^2\left(c_2 - \frac{3tc_1^2}{16}\right)\Bigg]. \end{aligned} \tag{2.14}$$

Applying the triangle inequality and then the Lemmas 1–3 in (2.14), we get

$$|H_{3,1}(f)| \leq \left[\frac{t}{12\sqrt{15}}\right]^2 [33+8t] = \left[\frac{b(1-\alpha)}{12\sqrt{15}}\right]^2 [33+8b(1-\alpha)]. \tag{2.15}$$

*Remark 4.* Choosing $b=2$ and $\alpha=0$ in the inequality (2.15), we see that it coincides with the result obtained by Zaprawa [22].

## 3. Conclusion

The upper bounds to the fourth order Hankel determinants for all the above defined subclasses of analytic functions were derived.

## Acknowledgements

The authors are highly grateful to the esteemed Referee(s) for a comprehensive reading of the manuscript and making valuable suggestions, leading to a better paper presentation.

## REFERENCES

1. Arif M., Rani L., Raza M., Zaprawa P. Fourth Hankel determinant for the family of functions with bounded turning. *Bull. Korean Math. Soc.*, 2018. Vol. 55, No. 6. P. 1703–1711. DOI: 10.4134/BKMS.b170994
2. Kowalczyk B., Lecko A., Sim Y. J. The sharp bound for the Hankel determinant of the third kind for convex functions. *Bull. Aust. Math. Soc.*, 2018. Vol. 97, No. 3. P. 435–445. DOI: 10.1017/S0004972717001125
3. Babalola K. O. *On $H_3(1)$ Hankel Determinant for Some Classes of Univalent Functions.* 2009. 7 p. arXiv:0910.3779v13 [math.CV]
4. Bansal D., Maharana S., Prajapat J. K. Third order Hankel determinant for certain univalent functions. *J. Korean Math. Soc.*, 2015. Vol. 52, No. 6. P. 1139–1148. DOI: 10.4134/JKMS.2015.52.6.1139
5. Cho N. E., Kowalczyk B., Kwon O. S. et al. The bounds of some determinants for starlike functions of order alpha. *Bull. Malays. Math. Sci. Soc.*, 2018. Vol. 41, No. 1. P. 523–535. DOI: 10.1007/s40840-017-0476-x
6. Duren P. L. *Univalent Functions.* Grundlehren Math. Wiss., vol. 259. New York: Springer, 1983. 384 p.
7. Fekete M., Szegö G. Eine bemerkung über ungerade schlichte funktionen. *J. Lond. Math. Soc.*, 1933. Vol. s1-8, No. 2. P. 85–89. DOI: 10.1112/jlms/s1-8.2.85
8. Hayami T., Owa S. Generalized Hankel determinant for certain classes. *Int. J. Math. Anal.*, 2010. Vol. 4, No. 52, P. 2573–2585.
9. Janteng A., Halim S. A., Darus M. Hankel determinant for starlike and convex functions. *Int. J. Math. Anal.,* 2007. Vol. 1, No. 13. P. 619–625.
10. Janteng A., Halim S. A., Darus M. Coefficient inequality for a function whose derivative has a positive real part. *J. Inequal. Pure Appl. Math.*, 2006. Vol. 7, No. 2. P. 1–5.
11. Lecko A., Sim Y. J., Smiarowska B. The sharp bound of the Hankel determinant of the third kind for starlike functions of order 1/2. *Complex Anal. Oper. Theory*, 2019. Vol. 13, No. 5. P. 2231–2238. DOI: 10.1007/s11785-018-0819-0
12. Lee S. K., Ravichandran V., Supramaniam S. Bounds for the second Hankel determinant of certain univalent functions. *J. Inequal. Appl.*, 2013. Art. no. 281. DOI: 10.1186/1029-242X-2013-281
13. Libera R. J., Złotkiewicz E. J. Coefficient bounds for the inverse of a function with derivative in $\mathcal{P}$. *Proc. Amer. Math. Soc.*, 1983. Vol. 87, No. 2. P. 251–257. DOI: 10.1090/S0002-9939-1983-0681830-8
14. Livingston A. E. The coefficients of multivalent close-to-convex functions. *Proc. Amer. Math. Soc.*, 1969. Vol. 21, No. 3. P. 545–552. DOI: 10.1090/S0002-9939-1969-0243054-0
15. Orhan H., Zaprawa P. Third Hankel determinants for starlike and convex functions of order alpha. *Bull. Korean Math. Soc.,* 2018. Vol. 55, No. 1. P. 165–173. DOI: 10.4134/BKMS.b160916
16. Pommerenke Ch. *Univalent Functions. With a Chapter on Quadratic Differentials by Gerd Jensen.* Studia Mathematic Band XXV. GmbH: Vandenhoeck and Ruprecht, 1975. 376 p.
17. Pommerenke Ch. On the coefficients and Hankel determinants of univalent functions. *J. Lond. Math. Soc.*, 1966. Vol. s1-41, No. 1. P. 111–122. DOI: 10.1112/jlms/s1-41.1.111
18. Raza M., Malik S. N. Upper bound of third Hankel determinant for a class of analytic functions related with lemniscate of Bernoulli. *J. Inequal. Appl.*, 2013. Art. no. 412. DOI: 10.1186/1029-242X-2013-412
19. Sudharsan T. V., Vijayalakshmi S. P., Stephen B. A. Third Hankel determinant for a subclass of analytic functions. *Malaya J. Mat.*, 2014. Vol. 2, No. 4. P. 438–444.

20. Vamshee Krishna D., RamReddy T. Coefficient inequality for certain $p$-valent analytic functions. *Rocky Mountain J. Math.*, 2014. Vol. 44, No. 6. P. 1941–1959. DOI: 10.1216/RMJ-2014-44-6-1941
21. Zaprawa P. On Hankel determinant $H_2(3)$ for univalent functions. *Results Math.*, 2018. Vol. 73, No. 3. P. 1–12. DOI: 10.1007/s00025-018-0854-1
22. Zaprawa P. Third Hankel determinants for subclasses of univalent functions. *Mediterr. J. Math.*, 2017. Vol. 14, No. 1. Art. no. 19. P. 1–10. DOI: 10.1007/s00009-016-0829-y

**URAL MATHEMATICAL JOURNAL, Vol. 8, No. 1, 2022, pp. 136–144**

DOI: 10.15826/umj.2022.1.012

# EVOLUTION OF A MULTISCALE SINGULARITY OF THE SOLUTION OF THE BURGERS EQUATION IN THE 4-DIMENSIONAL SPACE-TIME

**Sergey V. Zakharov**

Krasovskii Institute of Mathematics and Mechanics,
Ural Branch of the Russian Academy of Sciences,
16 S. Kovalevskaya str., Ekaterinburg, 620108, Russian Federation

svz@imm.uran.ru

**Abstract:** The solution of the Cauchy problem for the vector Burgers equation with a small parameter of dissipation $\varepsilon$ in the 4-dimensional space-time is studied:

$$\mathbf{u}_t + (\mathbf{u}\nabla)\mathbf{u} = \varepsilon\triangle\mathbf{u}, \quad u_\nu(\mathbf{x}, -1, \varepsilon) = -x_\nu + 4^{-\nu}(\nu+1)x_\nu^{2\nu+1},$$

With the help of the Cole–Hopf transform $\mathbf{u} = -2\varepsilon\nabla \ln H$, the exact solution and its leading asymptotic approximation, depending on six space-time scales, near a singular point are found. A formula for the growth of partial derivatives of the components of the vector field $\mathbf{u}$ on the time interval from the initial moment to the singular point, called the formula of the gradient catastrophe, is established:

$$\frac{\partial u_\nu(0,t,\varepsilon)}{\partial x_\nu} = \frac{1}{t}\left[1 + O\left(\varepsilon|t|^{-1-1/\nu}\right)\right], \quad \frac{t}{\varepsilon^{\nu/(\nu+1)}} \to -\infty, \quad t \to -0.$$

The asymptotics of the solution far from the singular point, involving a multistep reconstruction of the space-time scales, is also obtained:

$$u_\nu(\mathbf{x}, t, \varepsilon) \approx -2\left(\frac{t}{\nu+1}\right)^{1/2\nu} \tanh\left[\frac{x_\nu}{\varepsilon}\left(\frac{t}{\nu+1}\right)^{1/2\nu}\right], \quad \frac{t}{\varepsilon^{\nu/(\nu+1)}} \to +\infty.$$

**Keywords:** Vector Burgers equation, Cauchy problem, Cole–Hopf transform, Singular point, Laplace's method, Multiscale asymptotics.

## 1. Statement of the problem

In the present work, we study the solution of the following Cauchy problem for the vector Burgers equation in the $(3+1)$-dimensional space-time:

$$\frac{\partial\mathbf{u}}{\partial t} + (\mathbf{u}\nabla)\mathbf{u} = \varepsilon\triangle\mathbf{u}, \quad t \geqslant -1, \tag{1.1}$$

$$u_\nu(\mathbf{x}, -1, \varepsilon) = -x_\nu + \frac{\nu+1}{4^\nu}x_\nu^{2\nu+1}, \tag{1.2}$$

where $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$, $\mathbf{u} = (u_1, u_2, u_3)$ is a *potential* vector field, $\varepsilon$ is a small positive parameter of dissipation frequently called viscosity, and the index $\nu$ changes from 1 to 3.

The evolutionary equation (1.1) is widely used in the mechanics of continuous media [5], in particular, for modeling the formation and the propagation of shock waves (in limit of vanishing viscosity $\varepsilon$), in addition, it successfully serves as a basic instrument of the theoretical investigation of the large-scale structure of the Universe [6]; it is worth noting that the Burgers equation, in the case of small values of parameter $\varepsilon$, simulates good enough the observed mosaic distribution of the matter in the space at the distances measured by billions of light-years.

The aim of the present work is to study the arising microlocal singularity, i.e., the solution of problem (1.1), (1.2) as $\varepsilon \to +0$ near the singular point, which coincides with the origin because of the special choice of the initial data for $t = -1$. Our investigation has to find the scales of the localization of the singularity and explicit asymptotic formulas for the solution $\mathbf{u}(\mathbf{x}, t, \varepsilon)$.

In the context of the present paper, the terms "singularity" and "singular point" are understood in the sense of a large space gradient of the solution $\mathbf{u}(\mathbf{x}, t, \varepsilon)$ in some small neighborhood of the origin, while the solution itself is *smooth*; in this connexion, see [8, Sect. 2] or detailed explanations in the introductory part of survey [9].

Here, it is appropriate to mention that Arnold's scientific school performed a detailed topological classification of singular points and reconstructions[1] of singular sets of the solution of equation (1.1) in limit of vanishing viscosity $\varepsilon$ [1, Ch. 2, § 2.5], including the theorems forbidding some metamorphoses of singularities of solutions, for example, see [3, § 3; 4]; while, in the present investigation, we are mainly interested in analytical results of studying the asymptotic behavior of the solution for small, however, *not equal to zero*, values of parameter $\varepsilon$.

## 2. Exact solution and its asymptotics

By the standard Cole–Hopf transform

$$\mathbf{u} = -2\varepsilon\nabla \ln H \tag{2.1}$$

equation (1.1) is reduced to the linear heat equation $\partial H/\partial t = \varepsilon \triangle H$, whose solution with the initial condition (1.2) is easily obtained in the explicit form:

$$H(\mathbf{x}, t, \varepsilon) = \frac{1}{8\pi^{3/2}(1+t)^{3/2}} \int\limits_{\mathbb{R}^3} \exp\left\{\frac{1}{\varepsilon}\left[-\frac{|\mathbf{x}-\mathbf{s}|^2}{4(1+t)} + \frac{|\mathbf{s}|^2}{4} - \sum_{\mu=1}^{3}\left(\frac{s_\mu}{2}\right)^{2\mu+2}\right]\right\} \prod_{\mu=1}^{3} ds_\mu. \tag{2.2}$$

With the help of the scaling change of variables of integration

$$s_\mu = 2\varepsilon^{1/(2\mu+2)}\sigma_\mu,$$

from expression (2.2), we find the following formula:

$$\begin{gathered} H(\mathbf{x}, t, \varepsilon) = \frac{\varepsilon^{13/24}}{\pi^{3/2}(1+t)^{3/2}} \exp\left\{-\frac{|\mathbf{x}|^2}{4\varepsilon(1+t)}\right\} \\ \times \int\limits_{\mathbb{R}^3} \exp\sum_{\mu=1}^{3}\left[-\sigma_\mu^{2\mu+2} + \frac{t\sigma_\mu^2}{\varepsilon^{\mu/(\mu+1)}(1+t)} + \frac{x_\mu\sigma_\mu}{\varepsilon^{(2\mu+1)/(2\mu+2)}(1+t)}\right] \prod_{\mu=1}^{3} d\sigma_\mu. \end{gathered} \tag{2.3}$$

Whence, by elementary differentiation, we obtain

$$\begin{gathered} \frac{\partial H(\mathbf{x}, t, \varepsilon)}{\partial x_\nu} = -\frac{1}{2\varepsilon^{11/24}\pi^{3/2}(1+t)^{5/2}} \exp\left\{-\frac{|\mathbf{x}|^2}{4\varepsilon(1+t)}\right\} \\ \times \int\limits_{\mathbb{R}^3} \left(x_\nu - 2\varepsilon^{1/(2\nu+2)}\sigma_\nu\right) \exp\sum_{\mu=1}^{3}\left[-\sigma_\mu^{2\mu+2} + \frac{t\sigma_\mu^2}{\varepsilon^{\mu/(\mu+1)}(1+t)} + \frac{x_\mu\sigma_\mu}{\varepsilon^{(2\mu+1)/(2\mu+2)}(1+t)}\right] \prod_{\mu=1}^{3} d\sigma_\mu. \end{gathered} \tag{2.4}$$

[1]The equivalent terms used in the relevant literature are as follows: bifurcations, metamorphoses, perestroikas, transitions.

Using formulas (2.3) and (2.4), from transform (2.1) we immediately get the exact solution of the Cauchy problem (1.1), (1.2) in the component-wise form:

$$u_\nu(\mathbf{x},t,\varepsilon) = \frac{\displaystyle\int_{-\infty}^{+\infty} \left(x_\nu - 2\varepsilon^{1/(2\nu+2)}\sigma_\nu\right) \exp\left[-\sigma_\nu^{2\nu+2} + \frac{\Theta_\nu\sigma_\nu^2 + \Lambda_\nu\sigma_\nu}{(1+t)}\right] d\sigma_\nu}{\displaystyle(1+t)\int_{-\infty}^{+\infty} \exp\left[-\sigma_\nu^{2\nu+2} + \frac{\Theta_\nu\sigma_\nu^2 + \Lambda_\nu\sigma_\nu}{(1+t)}\right] d\sigma_\nu}, \tag{2.5}$$

where, for convenience, the inner variables

$$\Theta_\nu = \frac{t}{\varepsilon^{\nu/(\nu+1)}}, \quad \Lambda_\nu = \frac{x_\nu}{\varepsilon^{(2\nu+1)/(2\nu+2)}} \tag{2.6}$$

are introduced.

First of all, we must find the leading approximation of the exact solution obtained above, since the explicit expression (2.5) itself tells us few about the asymptotic structure of the solution.

**Statement 1.** *As* $|\mathbf{x}|+|t| \to 0$ *and* $\varepsilon \to +0$, *for the solution of problem* (1.1), (1.2), *there holds the asymptotic formula*

$$u_\nu(\mathbf{x},t,\varepsilon) = -2\varepsilon^{1/(2\nu+2)}\frac{\partial}{\partial\Lambda_\nu}\ln\int_{-\infty}^{+\infty}\exp\left(-\sigma_\nu^{2\nu+2}+\Theta_\nu\sigma_\nu^2+\Lambda_\nu\sigma_\nu\right)d\sigma_\nu + O(|\mathbf{x}|+|t|). \tag{2.7}$$

P r o o f. Near the origin, by the elementary passage to the limit $|\mathbf{x}|+|t| \to 0$ in formula (2.5), we obtain the expression for the leading approximation:

$$U_\nu(\mathbf{x},t,\varepsilon) = -2\varepsilon^{1/(2\nu+2)}\frac{\displaystyle\int_{-\infty}^{+\infty}\sigma_\nu\exp\left(-\sigma_\nu^{2\nu+2}+\Theta_\nu\sigma_\nu^2+\Lambda_\nu\sigma_\nu\right)d\sigma_\nu}{\displaystyle\int_{-\infty}^{+\infty}\exp\left(-\sigma_\nu^{2\nu+2}+\Theta_\nu\sigma_\nu^2+\Lambda_\nu\sigma_\nu\right)d\sigma_\nu}. \tag{2.8}$$

In the argument of the integrand exponent of this expression, we recognize the truncated generating family (in other words, the truncated versal deformation of the germ) of the Lagrange singularities $A_{2\nu+1}$; the first of them $A_3$ is usually called the Whitney fold; see [1, Ch. 2; 2, Ch. II, § 11, § 17].

With the help of some obvious transforms of formula (2.8), by formula (2.5) for small values of the independent variables $(\mathbf{x},t)$, we arrive at the desired result. □

Now, we are ready to move to the very center of the singularity of the solution $u_\nu(\mathbf{x},t,\varepsilon)$.

**Statement 2.** *As* $\Theta_\nu = \varepsilon^{-\nu/(\nu+1)}t \to -\infty$, *there holds the asymptotic formula*

$$\frac{\partial U_\nu(0,t,\varepsilon)}{\partial x_\nu} = \frac{1}{t}\left[1 + O\left(\varepsilon|t|^{-1-1/\nu}\right)\right]. \tag{2.9}$$

P r o o f. Using formula (2.8), let us show that the point $(\mathbf{x},t)=(0,0)$ is singular by computing the asymptotics of the derivative

$$\frac{\partial U_\nu(0,t,\varepsilon)}{\partial x_\nu} = -2\varepsilon^{-\nu/(\nu+1)}\frac{\displaystyle\int_{-\infty}^{+\infty}\sigma^2\exp\left(-\sigma^{2\nu+2}-|\Theta_\nu|\sigma^2\right)d\sigma}{\displaystyle\int_{-\infty}^{+\infty}\exp\left(-\sigma^{2\nu+2}-|\Theta_\nu|\sigma^2\right)d\sigma} \tag{2.10}$$

as $\Theta_\nu \to -\infty$. After the change of the variable of integration $\sigma = |\Theta_\nu|^{1/2\nu}\eta$, we have:

$$\frac{\partial U_\nu(0,t,\varepsilon)}{\partial x_\nu} = -2\varepsilon^{-\nu/(\nu+1)}|\Theta_\nu|^{1/\nu}\frac{\displaystyle\int_{-\infty}^{+\infty}\eta^2\exp\left(-|\Theta_\nu|^{1+1/\nu}S(\eta)\right)d\eta}{\displaystyle\int_{-\infty}^{+\infty}\exp\left(-|\Theta_\nu|^{1+1/\nu}S(\eta)\right)d\eta},$$

where the phase function $S(\eta) = \eta^{2\nu+2} + \eta^2$ has clearly only a unique point of minimum: $\eta = 0$.

In this special case, it is convenient to make use of the asymptotic formula for integrals of the Laplacian type:

$$\int\limits_{-\infty}^{+\infty} A(\eta)\exp\left(-\omega S(\eta)\right)d\eta = \exp\left(-\omega S(0)\right)\sqrt{\frac{2\pi}{\omega S''(0)}}$$
$$\times\left\{A(0) + \frac{1}{2\omega}\left[\frac{A''(0)}{S''(0)} + \frac{A'(0)S'''(0)}{(S''(0))^2} + A(0)\left(\frac{5(S'''(0))^2}{12(S''(0))^3} + \frac{S''''(0)}{4(S''(0))^2}\right)\right] + O\left(\frac{1}{\omega^2}\right)\right\},$$

where $\omega \to +\infty$. Taking into account that $S''(0) = 2$, for our phase function, after elementary calculations, we find a very simple approximation:

$$\frac{\partial U_\nu(0,t,\varepsilon)}{\partial x_\nu} = -\varepsilon^{-\nu/(\nu+1)}|\Theta_\nu|^{-1}\left[1 + O\left(|\Theta_\nu|^{-1-1/\nu}\right)\right], \quad \Theta_\nu \to -\infty;$$

whence we obtain the necessary result. □

*Remark 1.* Relation (2.9) as $t \to -0$ can be called the *formula of the gradient catastrophe*, because the variable $t$ enters the denominator. Strictly speaking, exactly in the sense of this statement, we use the term "singular point" with reference to the point $(\mathbf{x}, t) = (0, 0)$. Let us emphasize that for $t = 0$ the gradient $\partial U_\nu/\partial x_\nu$ still does not become infinite, although it has, according to formula (2.10), the order of the value $\varepsilon^{-\nu/(\nu+1)} \to +\infty$ as $\varepsilon \to +0$.

Now, let us look into the future: in other words, let us calculate the asymptotics of the function of the leading approximation $U_\nu(\mathbf{x}, t, \varepsilon)$ as $\Theta_\nu \to +\infty$.

**Statement 3.** *As $\Theta_\nu \to +\infty$ there holds the asymptotic formula*

$$U_\nu(\mathbf{x},t,\varepsilon) = -2\varepsilon^{1/(2\nu+2)}\left(\frac{\Theta_\nu}{\nu+1}\right)^{1/2\nu}\tanh\left[\Lambda_\nu\left(\frac{\Theta_\nu}{\nu+1}\right)^{1/2\nu}\right] + O\left(\Theta_\nu^{-(1+1/4\nu)}\right).$$

P r o o f. Using the change of the variable $\sigma_\nu = \Theta_\nu^{1/2\nu} z_\nu$, for the integral in the denominator of expression (2.8), we obtain

$$\int\limits_{-\infty}^{+\infty}\exp\left(-\sigma_\nu^{2\nu+2} + \Theta_\nu\sigma_\nu^2 + \Lambda_\nu\sigma_\nu\right)d\sigma_\nu = \Theta_\nu^{1/2\nu}\int\limits_{-\infty}^{+\infty}\exp\left[\Theta_\nu^{1+1/\nu}(z_\nu^2 - z_\nu^{2\nu+2}) + \Lambda_\nu\Theta_\nu^{1/2\nu}z_\nu\right]dz_\nu.$$

Following now Laplace's method, for the phase function

$$F(z) = z^2 - z^{2\nu+2},$$

we find two stationary points

$$z^\pm = \pm\frac{1}{(\nu+1)^{1/2\nu}}, \quad F'(z^\pm) = 0, \quad F''(z^\pm) = -4\nu,$$

and the necessary formula of the leading approximation for the integral

$$\int_{-\infty}^{+\infty} \exp\left(-\sigma_\nu^{2\nu+2} + \Theta_\nu\sigma_\nu^2 + \Lambda_\nu\sigma_\nu\right) d\sigma_\nu$$

$$= \Theta_\nu^{1/4\nu}\left(\frac{2\pi}{\nu}\right)^{1/2} \exp\left[\nu\left(\frac{\Theta_\nu}{\nu+1}\right)^{1+1/\nu}\right] \cosh\left[\Lambda_\nu\left(\frac{\Theta_\nu}{\nu+1}\right)^{1/2\nu}\right] + O\left(\Theta_\nu^{-(1+3/4\nu)}\right);$$

in addition, by the same method, the asymptotic formula

$$\int_{-\infty}^{+\infty} \sigma_\nu \exp\left(-\sigma_\nu^{2\nu+2} + \Theta_\nu\sigma_\nu^2 + \Lambda_\nu\sigma_\nu\right) d\sigma_\nu$$

$$= \Theta_\nu^{3/4\nu}\left(\frac{2\pi}{\nu(\nu+1)^{1/\nu}}\right)^{1/2} \exp\left[\nu\left(\frac{\Theta_\nu}{\nu+1}\right)^{1+1/\nu}\right] \sinh\left[\Lambda_\nu\left(\frac{\Theta_\nu}{\nu+1}\right)^{1/2\nu}\right] + O\left(\Theta_\nu^{-(1+1/4\nu)}\right)$$

is established. Substituting these formulas into expression (2.8), we easily arrive at the desired result. □

*Remark 2.* Using the change (2.6), from Statement 3 in the leading approximation, we obtain the relation

$$U_\nu(\mathbf{x}, t, \varepsilon) \approx -2\left(\frac{t}{\nu+1}\right)^{1/2\nu} \tanh\left[\frac{x_\nu}{\varepsilon}\left(\frac{t}{\nu+1}\right)^{1/2\nu}\right], \quad \Theta_\nu \to +\infty,$$

which gives a mathematical formulation of happening reconstructions of the scales of time and space in the solution of the problem under consideration:

$$\Theta_\nu = \frac{t}{\varepsilon^{\nu/(\nu+1)}} \longmapsto t, \quad \Lambda_\nu = \frac{x_\nu}{\varepsilon^{(2\nu+1)/(2\nu+2)}} \longmapsto \frac{x_\nu}{\varepsilon}.$$

## 3. Survey of the asymptotic structure

Using the form of the inner variables (2.6) and Statements 1–3, we can establish the boundaries of domains, where the obtained asymptotic approximations of the solution of problem (1.1), (1.2) remain valid.

Since in the formula for the solution (2.5) we have specific space-time scales, which are determined by changes (2.6), it is natural to define correspondingly the following sets of the independent variables:

$$\Omega_\nu = \{(\mathbf{x}, t) : |x_\nu| < \varepsilon^{(2\nu+1)/(2\nu+2)}, \ |t| < \varepsilon^{\nu/(\nu+1)}\}.$$

In the smallest domain $\Omega_3$ (in Figures 1–3, it is conventionally shown with lilac color), bounded in time by the value of order $\varepsilon^{3/4}$, for the solution $\mathbf{u}(\mathbf{x}, t, \varepsilon)$, a fortiori, there holds the asymptotic formula (2.7) of the leading approximation.

As $\Theta_3 \to +\infty$, according to Statement 3 for $\nu = 3$, the natural scale of the space localization in $x_3$ is constricted to the value of order $\varepsilon$ and for the component $u_3$ one should use the following approximate relation:

$$u_3(\mathbf{x}, t, \varepsilon) \approx -2\left(\frac{t}{4}\right)^{1/6} \tanh\left[\frac{x_3}{\varepsilon}\left(\frac{t}{4}\right)^{1/6}\right], \quad \Theta_3 \to +\infty, \tag{3.1}$$

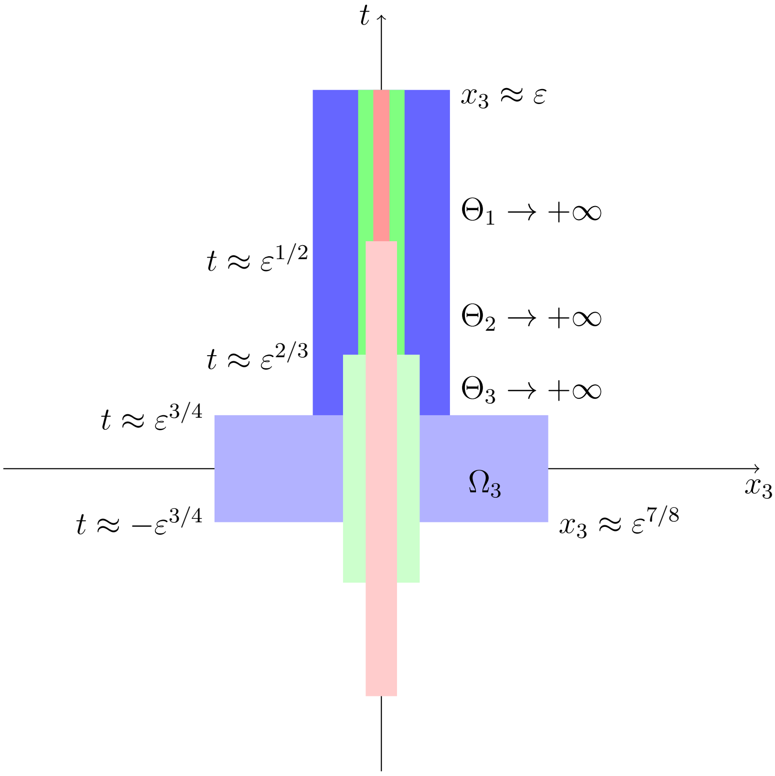



Figure 1. A schematic projection of the domain of the localization of the singularity in the plane $(x_3, t)$.

where $t = \varepsilon^{3/4}\Theta_3$.

In the intermediate domain $\Omega_2$, (in our figures, it is shown with light-green color) bounded in time by the value of order $\varepsilon^{2/3}$, the "remote future" from the point of view of domain $\Omega_3$, i.e., the times such that $\Theta_3 \to +\infty$, turns out to be a relatively short interval, because we have the relation $\Theta_2 = \varepsilon^{1/12}\Theta_3$.

*Remark 3.* Elegantly confirming Newton's principle *relativus de relativo in relativum*[2] from his famous "*Philosophiæ Naturalis Principia Mathematica*", a similar phenomenon is also observed in another Cauchy problem for equation (1.1) with an additional small parameter in the initial condition [8].

As $\Theta_2 \to +\infty$, according to Statement 3 for $\nu = 2$, the natural scale of the space localization in $x_2$ is constricted to the value of order $\varepsilon$ and for the component $u_2$ one should use the following approximate relation:

$$u_2(\mathbf{x}, t, \varepsilon) \approx -2\left(\frac{t}{3}\right)^{1/4} \tanh\left[\frac{x_2}{\varepsilon}\left(\frac{t}{3}\right)^{1/4}\right], \quad \Theta_2 \to +\infty,$$

where $t = \varepsilon^{2/3}\Theta_2$.

At last, in the largest domain $\Omega_1$ (in figures, it is shown with pink color), bounded in time by the value of order $\varepsilon^{1/2}$, the "remote future" already from the point of view of domain $\Omega_2$, i.e. the times such that $\Theta_2 \to +\infty$, again turns out to be only a short interval, because we have the relation $\Theta_1 = \varepsilon^{1/6}\Theta_2$.

[2]In author's free translation from Latin: anything relative [passes] from relative to relative.

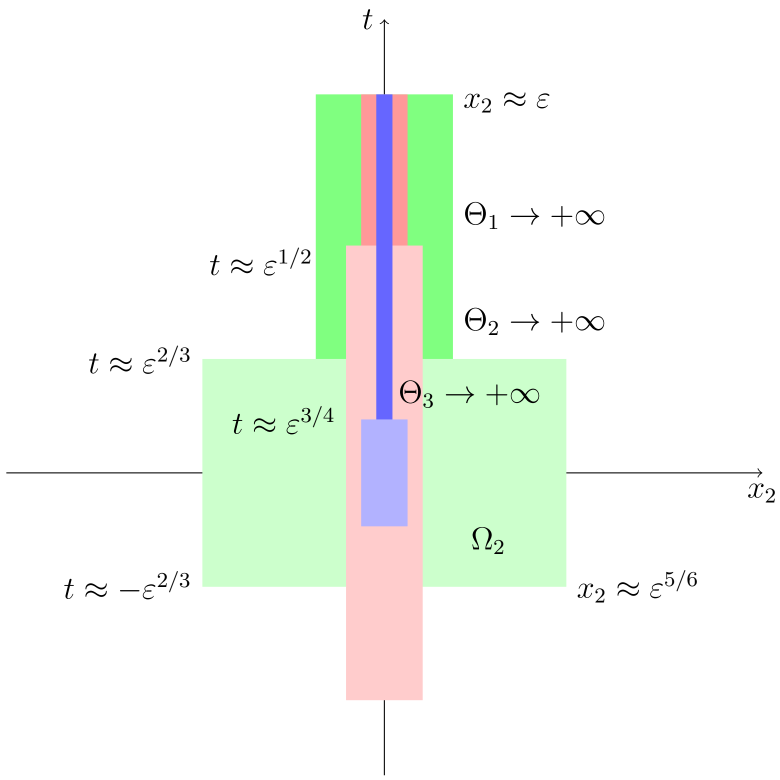



Figure 2. A schematic projection of the domain of the localization of the singularity in the plane $(x_2, t)$.

As $\Theta_1 \to +\infty$, according to Statement 3 for $\nu = 2$, the natural scale of the space localization in $x_1$ is constricted to the value of order $\varepsilon$ and for the component $u_1$ one should use the following approximate relation:

$$u_1(\mathbf{x}, t, \varepsilon) \approx -\sqrt{2t}\,\tanh\left[\frac{x_1}{\varepsilon}\sqrt{\frac{t}{2}}\right], \quad \Theta_1 \to +\infty, \tag{3.2}$$

where $t = \varepsilon^{1/2}\Theta_1$.

*Remark 4.* For correct understanding of the whole picture presented above, one important explanation should be given, although it is a quite trivial moment when one uses the standard matching method [7]. The indicated boundaries of the fragments of the asymptotic structure of the singularity of the solution $\mathbf{u}(\mathbf{x}, t, \varepsilon)$, that is the domains of the space-time scales of its localization, are not perfectly defined, since they can be displaced to some inessential distances, for example, with the help of multiplication by the value $\varepsilon^{\delta}$, where $0 < \delta \ll 1$, even without prejudice to the strictness of mathematical statements if only the overlapping of the transition regions, i.e. the domains of the reconstructions of the scales, is not upset.[3]

[3]The dialectic image of these metamorphoses and the place of particular ones in the whole structure of the singularity may be excellently reproduced by the sharp-witted phrase from the first book of Hegel's "*Wissenschaft der Logik*": "So ist das Endliche in dem Vergehen nicht vergangen; es ist zunächst nur ein *anderes* Endliches geworden... ." (In author's translation: "Thus the finite had not passed in the passage; first of all, it became only some *other* finite.")

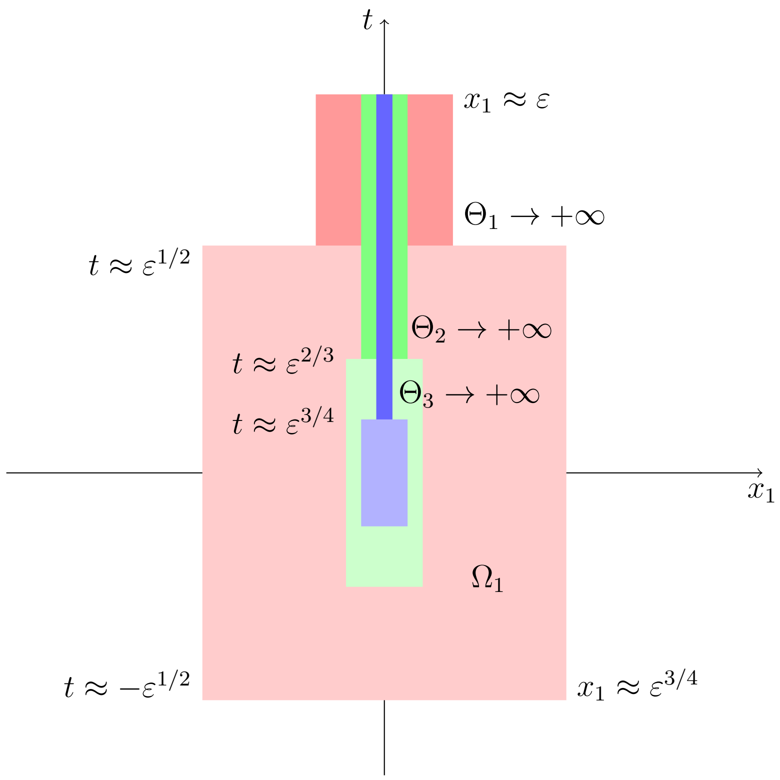



Figure 3. A schematic projection of the domain of the localization of the singularity in the plane $(x_1, t)$.

## 4. Summary

1. The explicit formula (2.5) for the exact solution of the investigated Cauchy problem and expression (2.8) for its leading asymptotic approximation clearly show that the specific form the initial condition (1.2) in a finite time generates a peculiar multiscale microlocal singularity, whose evolution is determined by the joint effect of the Lagrange singularities $A_3$, $A_5$, and $A_7$; as we have seen, their truncated versal deformations appear in the arguments of the corresponding integrand exponents.

2. As shown by Statement 3 and further detailed explanations in Section 3, in particular, see relations (3.1)–(3.2), moving away from the singular point $(\mathbf{x}, t) = (0, 0)$ is accompanied by the multistep reconstruction of the natural space-time scales of the asymptotics of the solution, in other words, by a successive "switching" of the orders of their values with respect to the small parameter of dissipation.

   In view of this interesting property, the case of the origin and the evolution of the multiscale singularity of the solution under consideration is conceptually close to the nontrivial hierarchy of the space-time reconstructions corresponding to the multiscale evolution of the initial singularity obtained in [8]. Taking into account the picture of asymptotic relations clarified above, we may say that the case considered in the present paper has the advantage of the statement of the problem itself, since the vector field (1.2) at the initial moment of time is smooth and does not depend on additional small parameters.

3. The summarizing thesis of the present paper, that confirms Hilbert's thought about the

importance of studying specific problems[4], can be expressed as follows: the results of our study give an obvious example when a rather simple experiment in mathematical physics — the exactly solvable Cauchy problem for an evolutionary differential equation with only one small parameter — was able to generate the *multiscale* structure of metamorphoses of the "life" of the solution in the 4-dimensional space-time.

## REFERENCES


1. Arnold V. I. *Singularities of Caustics and Wave Fronts.* Math. Its Appl. Ser., vol. 62. Dordrecht: Kluwer Acad. Publ., 1990. 258 p. DOI: 10.1007/978-94-011-3330-2
2. Arnold V. I., Gusein-Zade S. M., Varchenko A. N. *Singularities of Differentiable Maps. Vol. I. Classification of Critical Points, Caustics and Wave Fronts.* Monogr. Math., vol. 82. Boston, MA: Birkhäuser, 1985. 392 p.
3. Bogaevsky I. A. Reconstructions of singularities of minimum functions, and bifurcations of shock waves of the Burgers equation with vanishing viscosity. *St. Petersburg Math. J.*, 1990. Vol. 1, No. 4. P. 807–823.
4. Bogaevsky I. A. Perestroikas of shocks and singularities of minimum functions. *Phys. D: Nonlinear Phenomena*, 2002. Vol. 173, No. 1–2. P. 1–28. DOI: 10.1016/S0167-2789(02)00652-8
5. Gurbatov S. N., Rudenko O. V., Saichev A. I. *Waves and Structures in Nonlinear Nondispersive Media: General Theory and Applications to Nonlinear Acoustics.* Berlin: Springer, 2011. 472 p. DOI: 10.1007/978-3-642-23617-4
6. Gurbatov S. N., Saichev A. I., Shandarin S. F. Large-scale structure of the Universe. The Zeldovich approximation and the adhesion model. *Phys. Usp.*, 2012. Vol. 55, No. 3. P. 223–249. DOI: 10.3367/UFNr.0182.201203a.0233
7. Il'in A. M. *Matching of Asymptotic Expansions of Solutions of Boundary Value Problems.* Transl. Math. Monogr., vol. 102. Providence, R.I.: Am. Math. Soc., 1992. 281 p.
8. Zakharov S. V. Asymptotic solution of the multidimensional Burgers equation near a singularity. *Theor. Math. Phys.*, 2018. Vol. 196, No. 1. P. 976–982. DOI: 10.1134/S0040577918070048
9. Zakharov S. V. Singular points and asymptotics in the singular Cauchy problem for the parabolic equation with a small parameter. *Comp. Math. Math. Phys.*, 2020. Vol. 60, No. 5. P. 821–832. DOI: 10.1134/S0965542520050164


[4]In his well-known talk, in 1900, David Hilbert literally said: "Eine noch wichtigere Rolle als das Verallgemeinern spielt — wie ich glaube — bei der Beschäftigung mit mathematischen Problemen das Specialisiren." (In author's translation: "An even more important role is played, as I believe, by studying rather special mathematical problems than general ones".)

Editor: Tatiana F. Filippova

Managing Editor: Oxana G. Matviychuk

Design: Alexander R. Matviychuk

Contact Information

16 S. Kovalevskaya str., Ekaterinburg, Russia, 620990

Phone: +7 (343) 375-34-73

Fax: +7 (343) 374-25-81

Email: secretary@umjuran.ru

Web-site: https://umjuran.ru

N.N. Krasovskii Institute of Mathematics and Mechanics
of the Ural Branch of Russian Academy of Sciences

Ural Federal University named after the first President of Russia B.N. Yeltsin

Distributed for free