ISSN 1997-1397 (Print)
ISSN 2313-6022 (Online)

Журнал Сибирского федерального университета

Математика и физика

Journal of Siberian Federal University

Mathematics & Physics

2025 18 (1)

ISSN 1997-1397 (Print)

ISSN 2313-6022 (Online)

ЖУРНАЛ СИБИРСКОГО ФЕДЕРАЛЬНОГО УНИВЕРСИТЕТА Математика и Физика

2025 18 (1)

JOURNAL OF SIBERIAN FEDERAL UNIVERSITY Mathematics & Physics

Издание индексируется Scopus (Elsevier), Emerging Sources Citation Index (WoS, Clarivate Analytics), Российским индексом научного цитирования (НЭБ), представлено в международных и российских информационных базах: Ulrich's periodicals directiory, ProQuest, EBSCO (США), Google Scholar, MathNet.ru, КиберЛенинке.

Включено в список Высшей аттестационной комиссии «Рецензируемые научные издания, входящие в международные реферативные базы данных и системы цитирования».

Все статьи представлены в открытом доступе http://journal.sfu-kras.ru/en/series/mathematics_physics.

Журнал Сибирского федерального университета. Математика и физика.
Journal of Siberian Federal University. Mathematics & Physics.

Учредитель: Федеральное государственное автономное образовательное учреждение высшего образования "Сибирский федеральный университет"(СФУ)
Главный редактор: А.М. Кытманов. Редакторы: В.Е. Openers Зализняк, А.В. Щуплев.
Компьютерная верстка: Г.В. Хрусталева
№ 1. 26.02.2025. Индекс: 42327. Тираж: 1000 экз. Свободная цена
Адрес редакции и издателя: 660041 г. Красноярск, пр. Свободный, 79, оф. 32-03.
Отпечатано в типографии Издательства БИК СФУ
660041 г. Красноярск, пр. Свободный, 82а.
Свидетельство о регистрации СМИ ПИ № ФС 77-28724 от 29.06.2007 г., выданное Федеральной службой по надзору в сфере массовых коммуникаций, связи и охраны культурного наследия
http://journal.sfu-kras.ru
Подписано в печать 15.02.25. Формат 84×108/16. Усл.печ. л. 12,0.
Уч.-изд. л. 11,8. Бумага тип. Печать офсетная.
Тираж 1000 экз. Заказ 23348
Возрастная маркировка в соответствии с Федеральным законом № 436-ФЗ:16+

CONTENTS

СОДЕРЖАНИЕ

Journal of Siberian Federal University. Mathematics & Physics 2025, 18(1), 5–13

EDN: ACYSPC
УДК 517.5

Certain Integral Formulas Involving Products of Two Incomplete Beta Functions

Ahmed Ali Atash*
Department of Mathematics
Faculty of Education – Shabwah
Shabwah University
Yemen

Received 10.08.2024, received in revised form 11.09.2024, accepted 24.10.2024

Abstract. The aim of this paper is to obtain some integral formulas involving products of two incomplete beta functions in terms of general triple hypergeometric series and Kampé de Fériet function. Some new particular integral formulas involving the incomplete beta function are also calculated as an application of our main results with the help of Whipple, Dixon and extension of Dixon summation theorems.

Keywords: incomplete beta function, Integral formulas, Kampé de Fériet function, General triple hypergeometric series.

Citation: A.A.Atash, Certain Integral Formulas Involving Products of Two Incomplete Beta Functions, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 5–13. EDN: ACYSPC.

1. Introduction

The generalized hypergeometric function ${}_pF_q$ with p numerator parameters and q denominator parameters (p and q are positive integers or zero and z is complex variable) is defined by (see [10, 11])

$$ {}_pF_q\left[\begin{matrix} a_1,\ldots,a_p\,; \\ b_1,\ldots,b_q\,; \end{matrix} z\right]=\sum_{n=0}^{\infty}\frac{(a_1)_n\ldots(a_p)_n}{(b_1)_n\ldots(b_q)_n}\frac{z^n}{n!}, \tag{1}$$

where $(\lambda)_n$ denotes the Pochhammer's symbol defined by

$$\begin{aligned}(\lambda)_n &= \begin{cases} 1 & , \quad (n=0) \\ \lambda(\lambda+1)(\lambda+2)\ldots(\lambda+n-1) & , \quad (n\in\mathbb{N}) \end{cases} \\ &= \frac{\Gamma(\lambda+n)}{\Gamma(\lambda)}, \qquad (\lambda \in \mathbb{C}\backslash\mathbb{Z}_0^-) \end{aligned} \tag{2}$$

and $\Gamma(\lambda)$ is the gamma function defined by

$$\Gamma(\lambda)=\int_0^{\infty} t^{\lambda-1}e^{-t}dt, \qquad \Re(\lambda)>0. \tag{3}$$

The classical beta function $B(a,b)$ is defined by (see [11])

$$B(a,b)=\begin{cases} \displaystyle\int_0^1 t^{a-1}(1-t)^{b-1}dt & , \quad \Re(a)>0,\ \Re(b)>0, \\ \\ \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} & , \quad a,b\neq 0,-1,-2,\ldots\ . \end{cases} \tag{4}$$

*ah-a-atash@hotmail.com

The incomplete beta function is defined as follows [8]:

$$B_z(a,b)=\int_0^z t^{a-1}(1-t)^{b-1}dt, \quad 0\leqslant z\leqslant 1,\ a,b>0. \tag{5}$$

Further, the substitution $t=\sin^2(\theta)$ gives

$$B_z(a,b)=2\int_0^{\arcsin\sqrt{z}} \sin^{2a-1}(\theta)\cos^{2b-1}(\theta)d\theta. \tag{6}$$

The hypergeometric representation of incomplete beta function is given by [8]

$$B_z(a,b)=a^{-1}z^a{}_2F_1[a,1-b;a+1;z]. \tag{7}$$

Also, we recalling the following formulas for the incomplete beta function [7]:

$$B_z(a,b)=B(a,b)-B_{1-z}(b,a), \tag{8}$$

$$B_z(1,1)=z, \quad B_z(a,1)=\frac{z^a}{a}. \tag{9}$$

The Kampé de Fáriet function of two variables $F^{p:q;k}_{l:m;n}[x,y]$ is defined and represented as follows [10, 11]:

$$F\begin{matrix} p:\ q\ ;k \\ \\ l:m\,;n \end{matrix}\left[\begin{matrix} (a_p):\ (b_q)\ ;(c_k)\,; \\ \\ (\alpha_l):(\beta_m)\,;(\gamma_n)\,; \end{matrix}\ x\ ,\ y\right]$$

$$=\sum_{r,s=0}^{\infty}\frac{\prod\limits_{j=1}^{p}(a_j)_{r+s}\ \prod\limits_{j=1}^{q}(b_j)_r\ \prod\limits_{j=1}^{k}(c_j)_s}{\prod\limits_{j=1}^{l}(\alpha_j)_{r+s}\ \prod\limits_{j=1}^{m}(\beta_j)_r\ \prod\limits_{j=1}^{n}(\gamma_j)_s}\ \frac{x^r}{r!}\ \frac{y^s}{s!}. \tag{10}$$

Furthermore, we recall that the general triple hypergeometric series $F^{(3)}[x,y,z]$ is defined by [10, 11]:

$$F^{(3)}[x,y,z]=F^{(3)}\left[\begin{matrix} (a)::(b)\,;(b')\,;(b''):(c)\,;(c')\,;(c'')\,; \\ \\ (e)::(g)\,;(g')\,;(g''):(h)\,;(h')\,;(h'')\,; \end{matrix}\ x,y,z\right]$$

$$=\sum_{m,n,p=0}^{\infty}\Lambda(m,n,p)\frac{x^m}{m!}\ \frac{y^n}{n!}\ \frac{z^p}{p!}, \tag{11}$$

where, for convenience,

$$\Lambda(m,n,p)=\frac{\prod_{j=1}^{A}(a_j)_{m+n+p}\prod_{j=1}^{B}(b_j)_{m+n}\prod_{j=1}^{B'}(b'_j)_{n+p}\prod_{j=1}^{B''}(b''_j)_{p+m}}{\prod_{j=1}^{E}(e_j)_{m+n+p}\prod_{j=1}^{G}(g_j)_{m+n}\prod_{j=1}^{G'}(g'_j)_{n+p}\prod_{j=1}^{G''}(g''_j)_{p+m}}\times$$

$$\times\frac{\prod_{j=1}^{C}(c_j)_m\prod_{j=1}^{C'}(c'_j)_n\prod_{j=1}^{C''}(c''_j)_p}{\prod_{j=1}^{H}(h_j)_m\prod_{j=1}^{H'}(h'_j)_n\prod_{j=1}^{H''}(h''_j)_p} \tag{12}$$

and (a) abbreviates the array of A parameters $a_1,a_2,\ldots,a_A$, with similar interpretations for $(b),(b'),(b'')$ and so on.

Recently some works for the incomplete beta function with applications have been considered by several authors, see [1,3,4]. In this paper, we obtain some integral formulas involving products of two incomplete beta functions. Further, we apply these results with the help of Whipple, Dixon and extension of Dixon summation theorems to compute some new particular integral formulas involving incomplete beta function.

2. Integral formulas for the incomplete beta function

In this section, we establish four integral formulas involving products of two incomplete beta functions asserted by the following theorems:

Theorem 2.1. *The following integral formula holds true:*

$$\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)B_z(c,d)dz=$$
$$=\frac{x^{a+c+k}}{ac(a+c+k)}F^{(3)}\left[\begin{matrix} a+c+k :: -;-;-: a,1-b;c,1-d;1-p; \\ a+c+k+1 :: -;-;-: a+1 ; c+1 ; - ; \end{matrix} x,x,x\right]. \tag{13}$$

Proof. Denoting the left hand side of (13) by L, replacing the two incomplete beta functions by their hypergeometric representations given in (7), expanding the two ${}_2F_1$ in a power series, changing the order of summation and integration and using (5), we get

$$L=\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)B_z(c,d)dz=$$
$$=\frac{1}{ac}\sum_{m,n=0}^{\infty}\frac{(a)_m(1-b)_m(c)_n(1-d)_n}{(a+1)_m(c+1)_n m!n!}B_x(a+c+k+m+n,p). \tag{14}$$

Again, replacing the incomplete beta function in the right hand side of (14) by its hypergeometric representation given in (7) and expanding ${}_2F_1$ in a power series we have

$$L=\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)B_z(c,d)dz=$$
$$=\frac{1}{ac}\sum_{m,n=0}^{\infty}\frac{(a)_m(1-b)_m(c)_n(1-d)_n x^m x^n}{(a+1)_m(c+1)_n m!n!}\times$$
$$\times\frac{x^{a+c+k}}{a+c+k+m+n}\sum_{s=0}^{\infty}\frac{(a+c+k+m+n)_s(1-p)_s x^s}{(a+c+k+1+m+n)_s s!}. \tag{15}$$

Finally, by using the following identities:

$$\frac{a}{a+m}=\frac{(a)_m}{(a+1)_m}, \tag{16}$$

$$(a)_{m+n}=(a)_m(a+m)_n, \tag{17}$$

we get the right hand side of (13).This completes the proof of Theorem 2.1. □

Corollary 2.1. *For $c=d=1$ in Theorem 2.1 yields the following result:*

$$\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)dz=$$
$$=\frac{x^{a+k}}{a(a+k)}F^{1:2;1}_{1:1;0}\left[\begin{matrix} a+k : a,1-b;1-p; \\ a+k+1 : a+1 ; - ; \end{matrix} x , x\right]. \tag{18}$$

Theorem 2.2. *The following integral formula holds true:*

$$\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)B_{1-z}(c,d)dz=$$
$$=\frac{B(c,d)x^{a+k}}{a(a+k)}F^{1:2;1}_{1:1;0}\left[\begin{array}{l} a+k \;:a,1-b;1-p; \\ \qquad\qquad\qquad\qquad x\,,\,x \\ a+k+1:\; a+1\;;\;-\;; \end{array}\right]-$$
$$-\frac{x^{a+d+k}}{ad(a+d+k)}F^{(3)}\left[\begin{array}{l} a+d+k \;::-;-;-:a,1-b;d,1-c;1-p; \\ \qquad\qquad\qquad\qquad\qquad\qquad x,x,x \\ a+d+k+1::-;-;-:\; a+1\;;\;d+1\;;\;-\;; \end{array}\right]. \tag{19}$$

Proof. Denoting the left hand side of (19) by L and then applying the result (8), we have

$$L=\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)B_{1-z}(c,d)dz=$$
$$=\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)(B(c,d)-B_z(d,c))dz=$$
$$=B(c,d)\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)dz-\int_0^x z^{k-1}(1-z)^{p-1}B_z(a,b)B_z(d,c)dz$$

Now, using (13) and (18), we obtain the desired result. $\square$

If we use the same technique as in the proof of the integral (13) asserted in the Theorem 2.1, we have the following theorem:

Theorem 2.3. *The following integral formula holds true:*

$$\int_0^1 z^{k-1}(1-z)^{p-1}B_z(a,b)B_{1-z}(c,d)dz=$$
$$=\frac{B(a+k,c+p)}{ac}F^{0:3;3}_{1:1;1}\left[\begin{array}{l} \qquad - \qquad :a,1-b,a+k;c,1-d,c+p; \\ \qquad\qquad\qquad\qquad\qquad\qquad 1\,,\,1 \\ a+c+k+p:\quad a+1\quad;\quad c+1\quad; \end{array}\right]. \tag{20}$$

Corollary 2.2. *For $c=d=1$ in Theorem 2.3 yields the following result:*

$$\int_0^1 z^{k-1}(1-z)^{p-1}B_z(a,b)dz=$$
$$=\frac{B(a+k,p)}{a}{}_3F_2\left[\begin{array}{l} a,1-b,a+k\;; \\ \qquad\qquad\qquad 1 \\ a+1,a+k+p; \end{array}\right]. \tag{21}$$

Corollary 2.3. *For $a=b=1$ in Theorem 2.3 yields the following result:*

$$\int_0^1 z^{k-1}(1-z)^{p-1}B_{1-z}(c,b)dz=$$
$$=\frac{B(k,c+p)}{c}{}_3F_2\left[\begin{array}{l} c,1-d,c+p\;; \\ \qquad\qquad\qquad 1 \\ c+1,c+k+p; \end{array}\right]. \tag{22}$$

Theorem 2.4. *The following integral formula holds true:*

$$\int_0^1 z^{k-1}(1-z)^{p-1}B_z(a,b)B_z(c,d)dz=$$

$$=\frac{B(c,d)B(a+k,p)}{a}{}_3F_2\left[\begin{matrix}a,1-b,a+k\ ;\\ \\ a+1,a+k+p\ ;\end{matrix}\ 1\right]-$$

$$-\frac{B(a+k,d+p)}{ad}F^{0:3\,;\,3}_{1:1\,;\,1}\left[\begin{matrix}- & :a,1-b,a+k\ ; & d,1-c,d+p\ ; \\ & & & 1\ ,\ 1\\ a+d+k+p: & a+1\ \ ; & d+1\ \ ;\end{matrix}\right]. \tag{23}$$

Proof. Denoting the left hand side of (23) by L and then applying the result (8), we have

$$L=\int_0^1 z^{k-1}(1-z)^{p-1}B_z(a,b)B_z(c,d)dz=$$

$$=\int_0^1 z^{k-1}(1-z)^{p-1}B_z(a,b)(B(c,d)-B_{1-z}(d,c))dz=$$

$$=B(c,d)\int_0^1 z^{k-1}(1-z)^{p-1}B_z(a,b)dz-\int_0^1 z^{k-1}(1-z)^{p-1}B_z(a,b)B_{1-z}(d,c)dz$$

Now, using (20) and (21), we obtain the desired result. □

Corollary 2.4. *For* $k=p=1$ *in Theorem 2.4 yields the following result:*

$$\int_0^1 B_z(a,b)B_z(c,d)dz=$$

$$=B(c,d)B(a,b+1)-\frac{B(d,a+c+1)}{a(a+1)}{}_3F_2\left[\begin{matrix}a,1-b,a+c+1\ ;\\ \\ a+2,a+c+d+1\ ;\end{matrix}\ 1\right]. \tag{24}$$

Remark 2.1. *Note that*

$$\int_0^1 B_z(a,b)B_{1-z}(c,d)dz=$$

$$=\frac{B(c,a+d+1)}{a(a+1)}{}_3F_2\left[\begin{matrix}a,1-b,a+d+1\ ;\\ \\ a+2,a+c+d+1\ ;\end{matrix}\ 1\right]. \tag{25}$$

Corollary 2.5. *For* $k=2$, $p=1$ *in Theorem 2.4 yields the following result:*

$$\int_0^1 zB_z(a,b)B_z(c,d)dz=\frac{1}{2}B(c,d)(B(a,b)-B(a+2,b))-$$

$$-\frac{B(d,a+c+2)}{a(a+2)}{}_4F_3\left[\begin{matrix}a,1-b,a+2,a+c+2\ ;\\ \\ a+1,a+3,a+c+d+2\ ;\end{matrix}\ 1\right]. \tag{26}$$

Remark 2.2. *Note that*

$$\int_0^1 zB_z(a,b)B_{1-z}(c,d)dz=$$

$$=\frac{B(c,a+d+2)}{a(a+2)}{}_4F_3\left[\begin{matrix} a,1-b,a+2,a+d+2\ ; \\ \\ a+1,a+3,a+c+d+2\ ; \end{matrix} 1\right]. \tag{27}$$

3. Some particular integrals with examples

In this section, we compute some particular integrals involving the incomplete beta function as an applications of our main results given in Section 2.
I. Taking $p=1$ in (18) and using the following result [9]:

$${}_3F_2\left[\begin{matrix} a,b,c\ ; \\ \\ b+1,c+1\ ; \end{matrix} x\right]=\frac{1}{c-b}\left[c\ {}_2F_1\left[\begin{matrix} a,b\ ; \\ \\ b+1\ ; \end{matrix} x\right]-b\ {}_2F_1\left[\begin{matrix} a,c\ ; \\ \\ c+1\ ; \end{matrix} x\right]\right], \tag{28}$$

thus, after considering the result (7) we obtain the following integral formula:

$$\int_0^x z^{k-1}B_z(a,b)dz=\frac{1}{k}[x^kB_x(a,b)-B_x(a+k,b)]. \tag{29}$$

Example 3.1. *For* $x=\frac{1}{2}$, $a=b=\frac{3}{2}$, $k=2$ *in (29), we get*

$$\int_0^{\frac{1}{2}} zB_z\left(\frac{3}{2},\frac{3}{2}\right)dz=\frac{1}{48}-\frac{\pi}{512}. \tag{30}$$

Example 3.2. *For* $x=\frac{1}{4}$, $a=b=\frac{3}{2}$, $k=3$ *in (29), we get*

$$\int_0^{\frac{1}{4}} z^2B_z\left(\frac{3}{2},\frac{3}{2}\right)dz=\frac{27\sqrt{3}}{5120}-\frac{13\pi}{4608}. \tag{31}$$

Remark 3.1. *For* $k=1$ *in (29), we get the well-known result [7]*

$$\int_0^x B_z(a,b)dz=xB_x(a,b)-B_x(a+1,b). \tag{32}$$

Remark 3.2. *For* $x=1$ *in (29), we get*

$$\int_0^1 z^{k-1}B_z(a,b)dz=\frac{1}{k}[B(a,b)-B(a+k,b)]. \tag{33}$$

Further, using (8) in (33), we get

$$\int_0^1 z^{k-1}B_{1-z}(a,b)dz=\frac{1}{k}[B(a,b+k)]. \tag{34}$$

II. Taking $a=b=\frac{1}{2}$ in (29) and using the result (6), we get

$$\int_0^{\sqrt{x}} t^{2k-1}\arcsin\ t\ dt=\frac{1}{2k}\left[x^k\arcsin\ \sqrt{x}-\int_0^{\arcsin\ \sqrt{x}}\sin^{2k}\ t\ dt\right]. \tag{35}$$

Example 3.3. *For* $x=\frac{1}{4}$, $k=2$ *in (35), we get*

$$\int_0^{\frac{1}{2}} t^3 \arcsin\ t\ dt = \frac{7\sqrt{3}}{256} - \frac{5\pi}{384}. \tag{36}$$

Example 3.4. *For* $x=\frac{1}{4}$, $k=3$ *in (35), we get*

$$\int_0^{\frac{1}{2}} t^5 \arcsin\ t\ dt = \frac{3\sqrt{3}}{192} - \frac{19\pi}{2304}. \tag{37}$$

Remark 3.3. *For* $x=1$ *in (35), we get the well-known result [5]*

$$\int_0^1 t^{2k-1} \arcsin\ t\ dt = \frac{\pi}{4k}\left[1 - \frac{(2k-1)!!}{2^k k!}\right]. \tag{38}$$

III Taking $a=b$, $c=d$ in (24) and using classical Whipple theorem for ${}_3F_2(1)$ [2], we get

$$\int_0^1 B_z(a,a)B_z(c,c)dz = B(c,c)B(a,a+1) - \frac{(a+c+\frac{1}{2})B(a+c+1,\frac{1}{2})}{2^{2(a+c)}ac}. \tag{39}$$

Example 3.5. *For* $a=\frac{3}{2}$, $c=\frac{1}{2}$ *in (39), we get*

$$\int_0^1 B_z\left(\frac{3}{2},\frac{3}{2}\right)B_z\left(\frac{1}{2},\frac{1}{2}\right)dz = \frac{\pi^2}{16} - \frac{2}{9}. \tag{40}$$

Example 3.6. *For* $a=\frac{5}{2}$, $c=\frac{1}{2}$ *in (39), we get*

$$\int_0^1 B_z\left(\frac{5}{2},\frac{5}{2}\right)B_z\left(\frac{1}{2},\frac{1}{2}\right)dz = \frac{3\pi^2}{256} - \frac{1}{25}. \tag{41}$$

Remark 3.4. *Note that*

$$\int_0^1 B_z(a,a)B_{z-1}(c,c)dz = \frac{(a+c+\frac{1}{2})B(a+c+1,\frac{1}{2})}{2^{2(a+c)}ac}. \tag{42}$$

Example 3.7. *For* $a=\frac{3}{2}$, $c=\frac{1}{2}$ *in (42), we get*

$$\int_0^1 B_z\left(\frac{3}{2},\frac{3}{2}\right)B_{1-z}\left(\frac{1}{2},\frac{1}{2}\right)dz = \frac{2}{9}. \tag{43}$$

Example 3.8. *For* $a=\frac{5}{2}$, $c=\frac{1}{2}$ *in (42), we get*

$$\int_0^1 B_z\left(\frac{5}{2},\frac{5}{2}\right)B_{1-z}\left(\frac{1}{2},\frac{1}{2}\right)dz = \frac{1}{25}. \tag{44}$$

IV Taking $a=c$, $b=d$ in (24) and using classical Dixon theorem for ${}_3F_2(1)$ [2], we get

$$\int_0^1 [B_z(a,b)]^2 dz = \frac{B(a,b)}{(a+b)}\left(b\ B(a,b) - \frac{B(a+\frac{1}{2},b+\frac{1}{2})}{B(a+b+\frac{1}{2},\frac{1}{2})}\right). \tag{45}$$

Example 3.9. *For* $a=\frac{3}{2}$, $b=\frac{1}{2}$ *in (45), we get*

$$\int_0^1 \left[B_z\left(\frac{3}{2},\frac{1}{2}\right)\right]^2 dz=\frac{\pi^2}{16}-\frac{1}{3}. \tag{46}$$

Example 3.10. *For* $a=\frac{5}{2}$, $b=\frac{1}{2}$ *in (45), we get*

$$\int_0^1 \left[B_z\left(\frac{5}{2},\frac{1}{2}\right)\right]^2 dz=\frac{3\pi^2}{128}-\frac{2}{15}. \tag{47}$$

V Taking $a=c$, $b=d$ in (26) and using the extension of Dixon theorem for ${}_4F_3(1)$ [6], we get

$$\int_0^1 z[B_z(a,b)]^2 dz=\frac{B(a,b)}{2(a+b)(a+b+1)}\left((b^2+2ab+b)B(a,b)-\frac{(2a+1)B(a+\frac{1}{2},b+\frac{1}{2})}{B(a+b+\frac{1}{2},\frac{1}{2})}\right). \tag{48}$$

Example 3.11. *For* $a=\frac{3}{2}$, $b=\frac{1}{2}$ *in (48), we get*

$$\int_0^1 z\left[B_z\left(\frac{3}{2},\frac{1}{2}\right)\right]^2 dz=\frac{3\pi^2}{64}-\frac{2}{9}. \tag{49}$$

Example 3.12. *For* $a=\frac{5}{2}$, $b=\frac{1}{2}$ *in (48), we get*

$$\int_0^1 z\left[B_z\left(\frac{5}{2},\frac{1}{2}\right)\right]^2 dz=\frac{39\pi^2}{2048}-\frac{1}{10}. \tag{50}$$

References

[1] R.Alahmadi, H.Almefleh, Antiderivatives and integrals involving incomplete beta functions with applications, *Aust. J. Math. Anal. Appl.*, **17**(2020), 7 pages.

[2] W.N.Bailey, Generalized Hypergeometric Series, *Cambridge Tracts in Math. And Math. Phys.*, no. 32, Cambridge Univ. Press, London, 1935.

[3] J.González-Santander, A note on some reduction formulas for the incomplete beta function and the Lerch transcendent, *Mathematics*, **9**(2021), 1486.

[4] J.L.González-Santander, F.Sánchez Lasheras, Sums involving the Digamma function connected to the incomplete beta function and the Bessel functions, *Mathematics*, **11**(2023), 1937.

[5] I.S.Gradshteyn, I.M.Ryzhik, Tables of Integrals, Series, and Products. 7th ed.; Academic Press, Boston, MA, 2007.

[6] Y.S.Kim, M.A.Rakha, A.K.Rathie, Extensions of certain classica summation theorems for the series ${}_2F_1$, ${}_3F_2$, and ${}_4F_3$ with applications in Ramanujan's summations, *Intern. J. Mathematics and Math. Sci.*, **2010**(2010). DOI: 10.1155/2010/309503

[7] K.B.Oldham, J.C.Myland, J.Spanier, *An Atlas of Functions: with equator, the atlas function calculator.* Springer Science & Business Media, New York, 2009.

[8] F.W.Olver, D.W.Lozier, R.F.Boisvert, C.W.Clark (editors), NIST Handbook of Mathematical Functions, Cambridge University Press, 2010.

[9] A.P.Prudnikov, Yu.A.Brychkov, O.I.Marichev, Integrals and Series, Vol. 3, More Special Functions, Nauka, Moscow, 1986.

[10] H.M.Srivastava, P.W.Karlsson, Multiple Gaussian Hypergeometric Series, Halsted Press (Ellis Horwood Limited, Chichester), John Wiley & Sons, New York, Chichester, Brisbane, Toronto, 1985.

[11] H.M.Srivastava, H.L.Manocha, A Treatise on Generating Functions, Halsted Press, New York, 1984.

Некоторые интегральные формулы, включающие произведения двух неполных бета-функций

Ахмед Али Аташ
Кафедра математики
Факультет образования – Шабва
Университет Шабва
Шабва, Йемен

Аннотация. Целью данной статьи является получение некоторых интегральных формул, включающих произведения двух неполных бета-функций в терминах общих тройных гипергеометрических рядов и функции Кампаé де Фéриета. Некоторые новые частные интегральные формулы, включающие неполную бета-функцию, также вычисляются как приложение наших основных результатов с помощью теорем Уиппла, Диксона и расширения теоремы Диксона о суммировании.

Ключевые слова: неполная бета-функция, интегральные формулы, функция Кампа де Фье, общий тройной гипергеометрический ряд.

EDN: GHRLIM
УДК 517.958

Global Solvability of a Kernel Determination Problem in 2D Heat Equation with Memory

Durdumurod K. Durdiev*
Bukhara Branch of the Institute of mathematics
Academy of Sciences of the Republic of Uzbekistan
Bukhara State University
Bukhara, Uzbekistan

Zhavlon Z. Nuriddinov†
Bukhara State University
Bukhara, Uzbekistan

Received 10.08.2024, received in revised form 15.09.2024, accepted 24.10.2024

Abstract. The inverse problem of determining two dimensional kernel in the integro-differential heat equation is considered in this paper. The kernel depends on the time variable t and space variable x. Assuming that kernel function is given, the direct initial-boundary value problem with Neumann conditions on the boundary of a rectangular domain is studied for this equation. Using the Green's function, the direct problem is reduced to integral equation of the Volterra-type of the second kind. Then, using the method of successive approximation, the existence of a unique solution of this equation is proved. The direct problem solution on the plane $y = 0$ is used as an overdetermination condition for inverse problem. This problem is replaced by an equivalent auxiliary problem which is more suitable for further study. Then the last problem is reduced to the system of integral equations of the second order with respect to unknown functions. Applying the fixed point theorem to this system in the class of continuous in time functions with values in the Hölder spaces with exponential weight norms, the main result of the paper is proved. It consists of the global existence and uniqueness theorem for inverse problem solution.

Keywords: integro-differential heat equation, inverse problem, Banach theorem, existence, uniqueness.

Citation: D.K. Durdiev, Z.Z. Nuriddinov, Global Solvability of a Kernel Determination Problem in 2D Heat Equation with Memory, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 14–24. EDN: GHRLIM.

1. Introduction and preliminaries

The integro-differential equations with an integral term of convolution type are used in the mathematical modeling of biological phenomena and engineering sciences when it is necessary to take into consideration the history of the processes. In these integro-differential equations the convolution kernel accounts for memory influences. The key point when considering memory effects is that the kernel cannot be considered a known function because there is no way to measure it directly. Kernel can be reconstructed by additional measurements of physical field taken on a suitable subset of the body. Thus, an inverse problem has to be solved. The constitutive

*d.durdiev@mathinst.uz https://orcid.org/0000-0002-6054-2827
†j.zafarovich@mail.ru https://orcid.org/0000-0001-7379-3433

relations for a linear non-homogeneous heat propagation and diffusion processes in medium with memory contain a time- and space-dependent kernel functions in the integral of time convolution type [1, 2]. The memory effect phenomenon is governed by hyperbolic and parabolic integro–differential equations with time dependent memory kernel when the medium is homogeneous and time-space dependent memory kernel when the medium is heterogeneous. The kernel determination problems in one-dimensional heat conduction equations are widely encountered where memory kernel depends only on time variable. For example, in [3–13] (see also references therein) these problems were studied on the basis of the fixed point argument, and the local/global in time existence and uniqueness of inverse problems were derived. The numerical solutions for this problems were considered and efficient computational algorithms were constructed [14–17].

In this paper, the inverse problems of determining kernels of an integral convolution-type term in the integro-differential heat equation are studied with the use of the solution of the initial-boundary value problem in a rectangular domain given on the boundary $y = 0$. Unlike existing works, here the unknown kernel depends on both time and spatial coordinates. Consider the problem of determining functions $u(x, y, t)$ and $k(x, t)$ from the following equations:

$$u_t - \Delta u = \int_0^t k(x, t-\tau)u(x, y, \tau)d\tau + f(x, y, t), \quad (x, y, t) \in D_T, \tag{1}$$

$$u\mid_{t=0}= \varphi(x, y),\ (x, y) \in \overline{D}, \tag{2}$$

$$u_x\mid_{x=0}= u_x\mid_{x=1}= 0,\ u_y\mid_{y=0}= u_y\mid_{y=1}= 0,\ (x, y) \in \partial D \times [0, T], \tag{3}$$

$$u\mid_{y=0}= h(x, t),\ (x, t) \in [0, 1] \times [0, T], \tag{4}$$

where $\Delta = \dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}$ is the Laplace operator, $D_T = D \times (0, T]$, $D = \{(x, y) : x \in (0, 1),\ y \in (0, 1)\}$, $T > 0$ is an arbitrary fixed number. In the theory of inverse problems for differential equations, initial-boundary value problem (1)–(3) of determining function $u(x, y, t)$ with Neumann boundary conditions is called the *direct problem.* Function $u(x, y, t) \in C^{2,1}_{x,t}(D_T) \cap C^{1,0}_{x,t}\left(\overline{D}_T\right)$ is regular solution of the direct problem if it satisfies equalities (1)–(3).

Regular solution of (1)–(4) presupposes the fulfilment of the following conditions

$$\varphi_x(0, y) = \varphi_x(1, y) = 0, \quad \varphi_y(x, 0) = \varphi_y(x, 1) = 0, \quad \varphi(x, 0) = h(x, 0).$$

Let us introduce the class $H^l(D)$ of Hölder continuous functions on D with $l \in (0, 1)$. The space $H^{m+l}(D)$ (m is a nonnegative integer) and norms $|\cdot|^l$, $|\cdot|^{m+l}$ are defined in [18, pp. 16–27].The class of j times continuous differentiable with respect to t variable on the segment $[0,\ T]$ with values in $H^l(D)$ functions is denoted by $C^j\left(H^l(D),\ [0,\ T]\right)$. For a fixed t, the norm of function $g(x, y, t)$ in the $H^l(D)$ is denoted by $|g|^l(t)$. The norm of function $g(x, y, t)$ in $C^j\left(H^l(D),\ [0,\ T]\right)$ is defined by the equality

$$\|g\|^l := \sum_{i=0}^{j} \max_{t\in[0,\ T]} \left|\frac{\partial^i g}{\partial t^i}\right|^l (t).$$

2. Study of direct problem

The solution of problem (1)–(3) is equivalent to the following Volterra type integral equation

$$u(x,y,t)=\int_0^1\int_0^1 G(x,y,\xi,\eta,t)\varphi(\xi,\eta)d\xi d\eta+\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)f(\xi,\eta,\tau)d\xi d\eta d\tau+$$

$$+\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\int_0^\tau k(\xi,\tau-\alpha)u(\xi,\eta,\alpha)d\alpha d\xi d\eta d\tau, \tag{5}$$

where $G(x,y,\xi,\eta,t)$ is the Green function and it is defined as

$$G(x,y,\xi,\eta,t)=1+4\sum_{m,n=1}^{\infty} e^{-\lambda_{mn}t}\cos\pi nx\cos\pi my\cos\pi n\eta\cos\pi m\xi,\ \ \lambda_{mn}=\pi\sqrt{m^2+n^2}.$$

Lemma 2.1. *Suppose that* $\varphi(x,y)\in H^l(D)$, $f(x,y,t)\in C\Big(H^l(D),[0,T]\Big)$ *and* $k(x,t)\in C\Big(H^l([0,1]),[0,T]\Big)$. *Then there is a unique solution of integral equation (5) such that* $u(x,y,t)\in C^1\Big(H^{l+2}(D),[0,T]\Big)$.

Proof. To prove this Lemma, the method of successive approximations is used. At the first step, the following sequences of functions is constructed

$$\begin{aligned}u_0(x,y,t)&=\int_0^1\int_0^1 G(x,y,\xi,\eta,t)\varphi(\xi,\eta)d\xi d\eta+\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)f(\xi,\eta,\tau)d\xi d\eta d\tau,\\ u_i(x,y,t)&=\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\int_0^\tau k(\xi,\tau-\alpha)u_{i-1}(\xi,\eta,\alpha)d\alpha d\xi d\eta d\tau,\ i=1,2,\dots .\end{aligned} \tag{6}$$

For brevity, introduce the following notations

$$\varphi_{00}:=|\varphi|^l,\ \ f_0:=\|f\|^l,\ \ k_0:=\|k\|^l.$$

Let us estimate modules of functions $u_i(x,y,t)$. Using the Green's function property $\int\limits_0^1\int\limits_0^1 G(x,y,\xi,\eta,t)d\xi d\eta=1$,one can obtain from (6) for $(x,y,t)\in\overline{D}_T$ that

$$\Big|u_0(x,y,t)\Big|^l\leqslant\Big|\int_0^1\int_0^1 G(x,y,\xi,\eta,t)\varphi(\xi,\eta)d\xi d\eta\Big|^l+$$

$$+\Big|\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)f(\xi,\eta,\tau)d\xi d\eta d\tau\Big|^l\leqslant\varphi_{00}+f_0t,$$

$$\Big|u_i(x,y,t)\Big|^l\leqslant\Big|\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\int_0^\tau k(\xi,\tau-\alpha)u_{i-1}(\xi,\eta,\alpha)d\alpha d\xi d\eta d\tau\Big|^l\leqslant$$

$$\leqslant k_0^i\Big(\varphi_{00}\frac{t^{2i}}{2i!}+f_0\frac{t^{2i+1}}{(2i+1)!}\Big),\ i=1,2,\dots .$$

Let us define functional series $\sum\limits_{i=0}^{\infty}u_i(x,y,t)$. Using values obtained above, this series can be estimated as follows

$$\sum_{i=0}^{\infty}\Big|u_i(x,y,t)\Big|\leqslant\sum_{i=0}^{\infty}k_0^i\Big(\varphi_{00}\frac{T^{2i}}{2i!}+f_0\frac{T^{2i+1}}{(2i+1)!}\Big),\quad (x,y,t)\in\overline{D}_T.$$

Since the last number series converges, series $\sum\limits_{i=0}^{\infty} u_i(x,y,t)$ converges uniformly and absolutely. Obviously, under conditions of the Lemma the inclusion $u_0(x,y,t) \in C^{2,1}_{x,t}(D_T)$ is satisfied. Consequently, all $u_i(x,y,t)$ have this property, i.e., $u_i(x,y,t) \in C^{2,1}_{x,t}(D_T)$, $i = 1,2,\ldots$. Then, according to the general theory of linear integral equations of Volterra type, $\sum\limits_{i=0}^{\infty} u_i(x,y,t)$ is a regular solution of direct problem (1)–(3).

Let us show that equation (5) has a unique solution. For this, let us assume the opposite, that is, integral equation (5) has two different solutions $u^1(x,y,t)$ and $u^2(x,y,t)$ with the same data:

$$u^i(x,y,t) = \int_0^1\int_0^1 G(x,y,\xi,\eta,t)\varphi(\xi,\eta)d\xi d\eta + \int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)f(\xi,\eta,\tau)d\xi d\eta d\tau +$$
$$+\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\int_0^\tau k(\xi,\tau-\alpha)u^i(\xi,\eta,\alpha)d\alpha d\xi d\eta d\tau,\ i=1,2.$$

The difference of these functions is defined by $Z(x,y,t) = u^1(x,y,t) - u^2(x,y,t)$:

$$Z(x,y,t) = \int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\int_0^\tau k(\xi,\tau-\alpha)Z(\xi,\eta,\alpha)d\alpha d\xi d\eta d\tau. \tag{7}$$

Let us denote the modular supremum of function $Z(x,y,t)$ on $(x,y) \in D$ for each $t \in [0,T]$ as

$$\widetilde{Z}(t) = \sup_{(x,y)\in D}\Big|Z(x,y,t)\Big|,\quad t\in[0,T].$$

It follows from integral equation (7) that

$$\widetilde{Z}(t) \leqslant k_0 T\int_0^t \widetilde{Z}(\tau)d\tau.$$

According to the Gronuolla–Bellman inequality, the last integral inequality has only $\widetilde{Z}(t)\equiv 0$ solution. It means that $Z(x,y,t)=0$ or $u^1(x,y,t)=u^2(x,y,t)$ in domain D_T. The lemma is proved. □

3. Auxiliary problem

Suppose that functions in problem (1)–(4) are sufficiently smooth. The degree of smoothness for each function will be determined later.

The following assertion is true.

Lemma 3.1. *Problem (1)–(4) is equivalent to the following auxiliary problem for functions $\omega(x,y,t)$, $k(x,t)$:*

$$\omega_t - \Delta\omega = k(x,t)\varphi_{yy}(x,y) + f_{tyy}(x,y,t) + \int_0^t k(x,t-\tau)\omega(x,y,\tau)d\tau,\quad (x,y,t)\in D_T, \tag{8}$$

$$\omega\mid_{t=0} = \Delta\varphi_{yy}(x,y) + f_{yy}(x,y,0),\ (x,y)\in D, \tag{9}$$

$$\omega_x\mid_{x=0} = \omega_x\mid_{x=1} = 0,\ \omega_y\mid_{y=0} = \omega_y\mid_{y=1} = 0,\ \partial D\times[0,T], \tag{10}$$

$$\begin{aligned}\omega\mid_{y=0} = h_{tt}(x,t) - h_{xxt}(x,t) - k(x,t)\varphi(x,0) - f_t(x,0,t) -\\ -\int_0^t k(x,t-\tau)h_t(x,0,\tau)d\tau,\ (x,t)\in[0,1]\times[0,T],\end{aligned} \tag{11}$$

where $\omega(x,y,t) := u_{tyy}(x,y,t)$.

Proof. Upon differentiating equations (1)–(4) with respect to t and setting $\vartheta(x,y,t) := u_t(x,y,t)$, one can obtain the following equivalent problem for functions ϑ, k

$$\vartheta_t - \Delta\vartheta = k(x,t)\varphi(x,y) + f_t(x,y,t) + \int_0^t k(x,t-\tau)\vartheta(x,y,\tau)d\tau, \quad (x,y,t) \in D_T, \tag{12}$$

$$\vartheta\mid_{t=0} = \Delta\varphi(x,y) + f(x,y,0), \quad (x,y) \in D, \tag{13}$$

$$\vartheta_x\mid_{x=0} = \vartheta_x\mid_{x=1} = 0, \quad \vartheta_y\mid_{y=0} = \vartheta_y\mid_{y=1} = 0, \ (x,y) \in \partial D \times [0,T], \tag{14}$$

$$\vartheta\mid_{y=0} = h_t(x,t), \ (x,t) \in [0,1] \times [0,T]. \tag{15}$$

Here, it is assumed that

$$\Delta\varphi_x(0,y) + f_x(0,y,0) = \Delta\varphi_x(1,y) + f_x(1,y,0),$$

$$\Delta\varphi_y(x,0) + f_y(x,0,0) = \Delta\varphi_y(x,1,0) + f_y(x,1,0), \ \Delta\varphi(x,0) = h_t(x,0).$$

Hence it follows that if (u,k) is a solution of problem (1)–(4) then (12)–(15) has a solution (ϑ,k) with the same k. Let us prove the converse. Let (ϑ,k) satisfy relations (12)–(15) then

$$u(x,y,t) = \int_0^t \vartheta(x,y,\tau)d\tau + \varphi(x,y).$$

Let us show that relation (1) holds. It follows from (12)—(15) that

$$u_t - \Delta u - \int_0^t k(x,\tau)u(x,y,t-\tau)d\tau - f(x,y,t) =$$

$$= \vartheta(x,y,t) - \int_0^t \Delta\vartheta(x,y,\tau)d\tau - \Delta\varphi(x,y) - \int_0^t k(x,\tau)\int_0^{t-\tau}\vartheta(x,y,\alpha)d\alpha d\tau - \int_0^t k(x,\tau)\varphi(x,y)d\tau -$$

$$-f(x,y,t) = \int_0^t \vartheta_\tau(x,y,\tau)d\tau + \Delta\varphi(x,y) + f(x,y,0) - \int_0^t \Delta\vartheta(x,y,\tau)d\tau - \Delta\varphi(x,y) -$$

$$-\int_0^t k(x,\tau)\int_0^\tau \vartheta(x,y,\tau-\alpha)d\alpha d\tau - \int_0^t k(x,\tau)\varphi(x,y)d\tau - \int_0^t f_\tau(x,y,\tau)d\tau - f(x,y,0) =$$

$$= \int_0^t \Big[\vartheta_\tau - \Delta\vartheta - \int_0^\tau k(x,\alpha)\vartheta(x,y,\tau-\alpha)d\alpha - k(x,\tau)\varphi(x,y) - f_\tau(x,y,\tau)\Big]d\tau = 0.$$

This completes the proof of equivalence of problems (1)–(4) and (12)–(15).

Now consider the second auxiliary problem. It can be obtained from problem (12)–(15) for function $p(x,y,t) := \vartheta_y(x,y,t)$:

$$p_t - \Delta p = k(x,t)\varphi_y(x,y) + f_{ty}(x,y,t) + \int_0^t k(x,t-\tau)p(x,y,\tau)d\tau, \quad (x,y,t) \in D_T, \tag{16}$$

$$p\mid_{t=0} = \Delta\varphi_y(x,y) + f_y(x,y,0), \quad (x,y) \in D, \tag{17}$$

$$p_x\mid_{x=0} = p_x\mid_{x=1} = 0, \quad p_y\mid_{y=0} = p_y\mid_{y=1} = 0, \quad \partial D \times [0,T], \tag{18}$$

$$\begin{aligned} p_y\mid_{y=0} = {} & h_{tt}(x,t) - h_{xxt}(x,t) - k(x,t)\varphi(x,0) - f_t(x,0,t) - \\ & - \int_0^t k(x,t-\tau)h_t(x,0,\tau)d\tau, \quad (x,t) \in [0,1] \times [0,T]. \end{aligned} \tag{19}$$

It is assumed that

$$\Delta\varphi_{xy}(0,y)+f_{xy}(0,y,0)=\Delta\varphi_{xy}(1,y)+f_{xy}(1,y,0),$$

$$\Delta\varphi_{yy}(x,0)+f_{yy}(x,0,0)=\Delta\varphi_{yy}(x,1)+f_{yy}(x,1,0),$$

$$\Delta\varphi_{yy}(x,0)+f_{yy}(x,0,0)=h_{tt}(x,0)-h_{xxt}(x,0)-k(x,0)\varphi(x,0)-f_t(x,0,0).$$

This follows from (12—(15), and it can be proved by complete analogy with the previous case.

Therefore, if problem (12)–(15) has solution (ϑ,k), then problem (16)–(19) has solution (p,k) with the same k. Moreover, $p(x,y,t)=\vartheta_y(x,y,t)$. Conversely, let (p,k) satisfy relations (16)–(19).

Hence it follows that

$$\vartheta(x,y,t)=\int_0^y p(x,z,t)dz+h_t(x,t),$$

$$\begin{aligned}\vartheta_t-\Delta\vartheta-k(x,t)\varphi(x,y)-f_t(x,y,t)-\int_0^t k(x,t-\tau)\vartheta(x,y,\tau)d\tau=\\
=\int_0^y p_t(x,z,t)dz+h_{tt}(x,t)-\int_0^y\Delta p(x,z,t)dz-p_y(x,0,t)-h_{xxt}(x,t)-\\
-\int_0^y k(x,t)\varphi_z(x,z)dz-k(x,t)\varphi(x,0)-\int_0^y f_{tz}(x,z,t)dz-f_t(x,0,t)-\\
-\int_0^y\int_0^t k(x,t-\tau)\vartheta_z(x,z,\tau)d\tau dz-\int_0^t k(x,t-\tau)h_t(x,0,\tau)d\tau=\\
=\int_0^y\Big[p_t-\Delta p-k(x,t)\varphi_z(x,z)-f_{tz}(x,z,t)-\int_0^t k(x,t-\tau)\vartheta_z(x,z,\tau)d\tau\Big]dz-\\
-p_y(x,0,t)+h_{tt}(x,t)-h_{xxt}(x,t)-f_t(x,0,t)-k(x,t)\varphi(x,0)-\int_0^t k(x,t-\tau)h_t(x,0,\tau)d\tau=0\end{aligned}$$

Then the equivalence of problems (12)–(15) and (16)—(19) is proved. In similar way, one can show that problem (16)–(19) is equivalent to problem (10)–(13) for function $\omega:=p_y(x,y,t)$. This implies the equivalence of problems (1)–(4) and (8)–(11). The lemma is proved. □

4. Study of inverse problem (8)–(11)

The solution of problem (8)–(10) is equivalent to the following Volterra type integral equation

$$\begin{aligned}\omega(x,y,t)=\int_0^1\int_0^1 G(x,y,\xi,\eta,t)\Big(\Delta\varphi_{yy}(\xi,\eta)+f_{yy}(\xi,\eta)\Big)d\xi d\eta+\\
+\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)f_{tyy}(\xi,\eta,\tau)d\xi d\eta d\tau+\\
+\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)k(\xi,t-\tau)\varphi_{yy}(\xi,\eta)d\xi d\eta d\tau+\\
+\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\int_0^\tau k(\xi,\tau-\alpha)\omega(\xi,\eta,\alpha)d\alpha d\xi d\eta d\tau.\end{aligned}\tag{20}$$

Let $\varphi(x,0) \neq 0$, for all $x \in [0,1]$. Using equation (20) and additional conditions (11), one can obtain the following integral equation with respect to function $k(x,t)$

$$
\begin{aligned}
k(x,t) =& \frac{1}{\varphi(x,0)} \Big[h_{tt}(x,t) - h_{xxt}(x,t) - f_t(x,0,t) \Big] - \frac{1}{\varphi(x,0)} \int_0^t k(x,t-\tau) h_t(x,0,\tau) d\tau - \\
& - \frac{1}{\varphi(x,0)} \int_0^1 \int_0^1 G(x,0,\xi,\eta,t) \Big(\Delta\varphi_{yy}(\xi,\eta) - f_{yy}(\xi,\eta) \Big) d\xi d\eta - \\
& - \frac{1}{\varphi(x,0)} \int_0^t \int_0^1 \int_0^1 G(x,0,\xi,\eta,t-\tau) f_{tyy}(\xi,\eta,\tau) d\xi d\eta d\tau - \\
& - \frac{1}{\varphi(x,0)} \int_0^t \int_0^1 \int_0^1 G(x,0,\xi,\eta,t-\tau) k(\xi,t-\tau) \varphi_{yy}(\xi,\eta) d\xi d\eta d\tau - \\
& - \frac{1}{\varphi(x,0)} \int_0^t \int_0^1 \int_0^1 G(x,0,\xi,\eta,t-\tau) \int_0^\tau k(\xi,\tau-\alpha) \omega(\xi,\eta,\alpha) d\alpha d\xi d\eta d\tau.
\end{aligned}
\tag{21}
$$

The main result of this work is the following assertion.

Theorem 4.1. *Assume that* $\varphi(x,y) \in H^{l+4}(D)$, $|\varphi(x,0)| \geqslant \varphi_0 = const > 0$, $f(x,y,t) \in C^1\Big(H^{l+2}(D); [0,T]\Big)$, $h(x,t) \in C^2\Big(H^{l+2}([0,1]); [0,T]\Big)$. *In addition, all the above matching conditions with respect to the specified functions are fulfilled. Then for any fixed* $T > 0$, *there exists a unique solution of integral equations (20), (21) and* $\omega(x,y,t) \in C\Big(H^{l+2}(D); [0,T]\Big)$, $k(x,t) \in C\Big(H^l([0,1]); [0,T]\Big)$.

Proof. The system of equation (20), (21) is closed system of integral equations with respect to functions $\omega(x,y,t)$ and $k(x,t)$. Let us write this system in the form of a non-linear operator equation

$$\psi = A\psi, \tag{22}$$

where $\psi = (\psi_1, \psi_2,) = (\omega(x,y,t), k(x,t))^*$, $*$ is the symbol of transposition. The operator in (22) has the form $A\psi = \Big[(A\psi)_1, (A\psi)_2\Big]$;

$$
\begin{aligned}
(A\psi)_1 =& \psi_{01}(x,y,t) + \int_0^t \int_0^1 \int_0^1 G(x,y,\xi,\eta,t-\tau) \psi_2(\xi,t-\tau) \varphi_{yy}(\xi,\eta) d\xi d\eta d\tau + \\
& + \int_0^t \int_0^1 \int_0^1 G(x,y,\xi,\eta,t-\tau) \int_0^\tau \psi_2(\xi,\tau-\alpha) \psi_1(\xi,\eta,\alpha) d\alpha d\xi d\eta d\tau.
\end{aligned}
\tag{23}
$$

$$
\begin{aligned}
(A\psi)_2 =& \psi_{02}(x,t) - \frac{1}{\varphi(x,0)} \int_0^t \psi_2(x,t-\tau) h_t(x,0,\tau) d\tau - \\
& - \frac{1}{\varphi(x,0)} \int_0^t \int_0^1 \int_0^1 G(x,0,\xi,\eta,t-\tau) \psi_2(\xi,t-\tau) \varphi_{yy}(\xi,\eta) d\xi d\eta d\tau - \\
& - \frac{1}{\varphi(x,0)} \int_0^t \int_0^1 \int_0^1 G(x,0,\xi,\eta,t-\tau) \int_0^\tau \psi_2(\xi,\tau-\alpha) \psi_1(\xi,\eta,\alpha) d\alpha d\xi d\eta d\tau.
\end{aligned}
\tag{24}
$$

The following designations are used in equations (23), (24)

$$
\begin{aligned}
\psi_{01}(x,y,t) =& \int_0^1 \int_0^1 G(x,y,\xi,\eta,t) \Big(\Delta\varphi_{yy}(\xi,\eta) + f_{yy}(\xi,\eta) \Big) d\xi d\eta + \\
& + \int_0^t \int_0^1 \int_0^1 G(x,y,\xi,\eta,t-\tau) f_{tyy}(\xi,\eta,\tau) d\xi d\eta d\tau,
\end{aligned}
$$

$$\psi_{02}(x,t) = \frac{1}{\varphi(x,0)}\Big[h_{tt}(x,t) - h_{xxt}(x,t) - f_t(x,0,t)\Big] -$$

$$-\frac{1}{\varphi(x,0)}\int_0^1\int_0^1 G(x,0,\xi,\eta,t)\Big(\Delta\varphi_{yy}(\xi,\eta) - f_{yy}(\xi,\eta)\Big)d\xi d\eta -$$

$$-\frac{1}{\varphi(x,0)}\int_0^t\int_0^1\int_0^1 G(x,0,\xi,\eta,t-\tau)f_{tyy}(\xi,\eta,\tau)d\xi d\eta d\tau.$$

Let $C_\sigma\left(H^l(D),\ [0,T]\right)$ be the Banach space of continuous with respect to t variable on the segment $[0,\,T]$ with values in $H^l(D)$ functions with the family of weighted norms $\|\cdot\|^l_\sigma$, $\sigma \geqslant 0$

$$\|\psi\|^l_\sigma = \max_{t\in[0,T]} e^{-\sigma t}|\psi_i|^l,\ i = 1,2. \tag{25}$$

Obviously, C_σ with $\sigma = 0$ is the usual space of continuous in t on $[0,\,T]$ with values in $H^l(D)$ functions with the ordinary norm (see Introduction). In what follows it is denoted by $\|\cdot\|^l$. Because

$$e^{-\sigma t}\|\psi\|^l \leqslant \|\psi\|^l_\sigma \leqslant \|\psi\|^l \tag{26}$$

norms $\|\cdot\|^l_\sigma$ and $\|\cdot\|^l$ are equivalent for any $t \in [0,T]$. Parameter σ will be defined later.

Consider space C_σ with $\sigma \geqslant 0$. Let us introduce the ball $S_\sigma(\psi,\|\psi_0\|^l) := \{\psi : \|\psi - \psi_0\|^l_\sigma \leqslant \|\psi_0\|^l\}$ of radius $|\psi_0\|^l$ centred at the point ψ_0, where vector function ψ_0 has components ψ_{0i}, $i = 1,2$ and $\|\psi_0\|^l = \max\limits_{i=1,2}|\psi_{0i}|^l$. Obviously, the estimate $\|\psi\|^l_\sigma \leqslant \|\psi\|^l_\sigma + \|\psi_0\|^l \leqslant 2\|\psi_0\|^l$ holds for a function $\psi \in S_\sigma(\psi_0,\|\psi_0\|^l)$. Let $\psi \in S_\sigma(\psi_0,\|\psi_0\|^l)$. Let us prove that operator A is contracting operator on set $\psi \in S_\sigma(\psi_0,\|\psi_0\|^l)$ for an appropriately chosen $\sigma > 0$. First let us show that if $\sigma > 0$ is chosen appropriately then operator A maps ball $S_\sigma(\psi_0,\|\psi_0\|^l)$ into the same ball, i.e., $A\psi \in S_\sigma(\psi_0,\|\psi_0\|^l)$.

Indeed, using relations (20), (21) for the norm of differences and denoting $\varphi_1 := |\varphi|^{l+4}$, $h_0 := |h|^{l+2}$ for $(x,t) \in [0,1]\times[0,T]$, one can obtain

$$\|(A\psi)_1 - \psi_{01}\|^l_\sigma = \max_{t\in[0,t]}\Big|((A\psi)_1 - \psi_{01})\Big|^l e^{-\sigma t} \leqslant \max_{t\in[0,t]}\Big|\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\times$$
$$\times e^{-\sigma t}\psi_2(\xi,t-\tau)e^{-\sigma(t-\tau)}\varphi_{yy}(\xi,\eta)d\xi d\eta d\tau +$$

$$+\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\int_0^\tau e^{-\sigma(\tau-\alpha)}\psi_2(\xi,\tau-\alpha)e^{-\sigma(\alpha)}\psi_1(\xi,\eta,\alpha)e^{-\sigma(t-\tau)}d\alpha d\xi d\eta d\tau\Big|^l \leqslant$$
$$\leqslant \|\psi_2\|^l_\sigma\varphi_1\frac{1}{\sigma} + 2\|\psi_2\|^l_\sigma\|\psi_1\|^l_\sigma\frac{T}{\sigma} \leqslant 2\|\psi_0\|^l(\varphi_1 + 4\|\psi_0\|^l T)\frac{1}{\sigma},$$

$$\|(A\psi)_2 - \psi_{02}\|^l_\sigma = \max_{t\in[0,T]}\Big|((A\psi)_2 - \psi_{02})\Big|^l e^{-\sigma t} \leqslant \varphi_0^{-1}\Big[\|\psi_2\|^l_\sigma h_0\frac{1}{\sigma} + \|\psi_2\|^l_\sigma\varphi_1\frac{1}{\sigma} +$$

$$+2\|\psi_2\|^l_\sigma\|\psi_1\|^l_\sigma\frac{T}{\sigma}\Big] \leqslant 2\|\psi_0\|^l\varphi_0^{-1}\Big[h_0 + \varphi_1 + 4\|\psi_0\|^l T\Big]\frac{1}{\sigma}.$$

Let $\sigma \geqslant \sigma_0$, where

$$\sigma_0 = 2\max\{\varphi_1 + 4\|\psi_0\|^l T, h_0 + \varphi_1 + 4\|\psi_0\|^l T\}.$$

Then operator A maps $S_\sigma(\psi_0,\|\psi_0\|^l)$ into itself, i.e., $A\psi \in S_\sigma(\psi_0,\|\psi_0\|^l)$.

Let us show the fulfilment of the second property of construction map for operator A. First, one should note that inequalities for $\psi^{(1)} = \left(\psi_1^{(1)},\ \psi_2^{(1)}\right) \in S_\sigma\left(\psi_0, \|{}^l\psi_0\|^l\right)$, $\psi^{(2)} = = \left(\psi_1^{(2)}, \psi_2^{(2)}\right) \in S_\sigma\left(\psi_0, \|\psi_0\|^l\right)$.

$$\left|\psi_2^{(1)}\psi_1^{(1)} - \psi_2^{(2)}\psi_1^{(2)}\right|^l = \left|\left(\psi_2^{(1)} - \psi_2^{(2)}\right)\psi_1^{(1)} + \psi_2^{(2)}\left(\psi_1^{(1)} - \psi_1^{(2)}\right)\right|^l \leqslant$$

$$\leqslant 2\left|\psi^{(1)} - \psi^{(2)}\right|^l \max\left(\left|\psi_1^{(1)}\right|^l,\ \left|\psi_2^{(2)}\right|^l\right) \leqslant 4\|\psi_0\|^l\left|\psi^{(1)} - \psi^{(2)}\right|^l.$$

holds. Then one can obtain

$$\left\|((A\psi)^{(1)} - (A\psi)^{(2)})_1\right\|_\sigma^l = \max_{t\in[0,T]}\left|((A\psi)^{(1)} - (A\psi)^{(2)})_1\right|^l e^{-\sigma t} \leqslant$$

$$\leqslant \max_{t\in[0,T]}\Big|\int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)e^{-\sigma t}\Big(\psi_2^{(1)}(\xi,t-\tau) - \psi_2^{(2)}(\xi,t-\tau)\Big)\times$$

$$\times e^{-\sigma(t-\tau)}\varphi_{yy}(\xi,\eta)d\xi d\eta d\tau + \int_0^t\int_0^1\int_0^1 G(x,y,\xi,\eta,t-\tau)\int_0^\tau\Big(e^{-\sigma(\tau-\alpha)}\psi_2^{(1)}(\xi,\tau-\alpha)\times$$

$$\times e^{-\sigma(\alpha)}\psi_1^{(1)}(\xi,\eta,\alpha) - \psi_2^{(2)}(\xi,\tau-\alpha)\psi_1^{(2)}(\xi,\eta,\alpha)e^{-\sigma(t-\tau)}\Big)d\alpha d\xi d\eta d\tau\Big|^l \leqslant$$

$$\leqslant |\psi^{(1)} - \psi^{(2)}|^l\varphi_1\frac{1}{\sigma} + 8\|\psi_0\|^l|\psi^{(1)} - \psi^{(2)}|^l\frac{T}{\sigma} \leqslant |\psi^{(1)} - \psi^{(2)}|^l\Big(\varphi_1 + 8\|\psi_0\|^l T\Big)\frac{1}{\sigma}.$$

$$\left\|((A\psi)^{(1)} - (A\psi)^{(2)})_2\right\|_\sigma^l = \max_{t\in[0,T]}\left|((A\psi)^{(1)} - (A\psi)^{(2)})_2\right|^l e^{-\sigma t} \leqslant$$

$$\leqslant \varphi_0^{-1}\Big[h_0|\psi^{(1)} - \psi^{(2)}|^l\frac{1}{\sigma} + |\psi^{(1)} - \psi^{(2)}|^l\varphi_1\frac{1}{\sigma} + 8\|\psi_0\|^l|\psi^{(1)} - \psi^{(2)}|^l\frac{T}{\sigma}\Big] \leqslant$$

$$\leqslant |\psi^{(1)} - \psi^{(2)}|^l\varphi_0^{-1}\Big[h_0 + \varphi_1 + 8\|\psi_0\|^l T\Big]\frac{1}{\sigma}.$$

Let $\sigma \geqslant \sigma^*$, where

$$\sigma^* = \max\left\{\varphi_1 + 8\|\psi_0\|^l T, \varphi_0^{-1}\Big[h_0 + \varphi_1 + 8\|\psi_0\|^l T\Big]\right\}.$$

Then operator A is contracting operator on $S_\sigma(\psi_0, \|\psi_0\|^l)$. It follows from the Banach fixed-point theorem that (22) is solvable and has a unique solution in $S_\sigma(\psi_0, \|\psi_0\|^l)$ for any fixed $T > 0$.

Since $\omega =: \psi_1$ then

$$u_{yyt}(x,y,t) = \psi_1(x,y,t). \tag{27}$$

Function $u(x,y,t)$ is determined from equation (27) as follows

$$u(x,y,t) = h(x,t) + \varphi(x,y) - \varphi(x,0) + \int_0^y\int_0^t (y-\eta)\omega(x,\eta,\tau)d\tau d\eta.$$

Thus, the solution of inverse problem (1)–(4) is found. □

References

[1] L.Pandolfi, Systems with Persistent Memory, Springer Nature Switzerland, 2021.

[2] D.K.Durdiev, Z.D.Totieva, Kernel Determination Problems in Hyperbolic Integro-Differential Equations, Singapore, Springer, 2023.

[3] S.Avdonin, S.Ivanov, J.Wang, Inverse problems for the heat equation with memory, *Inverse problems and imaging*, **13**(2019), no. 1, 31–38.

[4] D.Guidetti, Reconstruction of a convolution kernel in a parabolic problem with a memory term in the boundary conditions, *Bruno Pini Mathematical Analysis Seminar*, **4**(2013), no. 1, 47–55. DOI: 10.6092/issn.2240-2829/4154

[5] C.Cavaterra, D.Guidetti, Identification of a convolution kernel in a control problem for the heat equation with a boundary memory term, *Annali di Matematica*, **193**(2014), 779–816. DOI: 10.1007/s10231-012-0301-y.

[6] D.K.Durdiev, J.J.Jumaev, Memory kernel reconstruction problems in the integro-differential equation of rigid heat conductor, *Mathematical Methods in the Applied Sciences*, **45**(2022), 8374–8388. DOI: 10.1002/mma.7133

[7] D.K.Durdiev, Zh.Zh.Zhumaev, One-dimensional inverse problems of finding the kernel of the integro-differential heat equation in a bounded domain, *Ukrains'kyi Matematychnyi Zhurnal*, **73**, (2021), no 11, 1492–1506.

[8] D.K.Durdiev, Zh.Zh.Zhumaev On determination of the coefficient and kernel in an integro-differential equation of parabolic type, *Eurasian Journal of Mathematical and Computer Applications*, **11** 2023, no 1, pp. 49–65

[9] D.K.Durdiev, Z.Z.Nuriddinov, Determination of a multidimensional kernel in some parabolic integro-differential equation, *Journal of Siberian Federal University. Mathematics and Physics*, **14**(2021), no 1, стр. 117–127.

[10] D.K.Durdiev, Z.Z.Nuriddinov, Uniqueness of the Kernel Determination Problem in a Integro-Differential Parabolic Equation with Variable Coefficients, *Russian Mathematics*, **67**(2023), 11, 1–11.

[11] S.Avdonin, S.Ivanov, J.Wang, Inverse problems for the heat equation with memory, *Inverse problems and imaging*, **13**(2019), no. 1, 31–38.

[12] D.Guidetti, Reconstruction of a convolution kernel in a parabolic problem with a memory term in the boundary conditions, *Bruno Pini Mathematical Analysis Seminar*, **4**(2013), no. 1, 47–55. DOI: 10.6092/issn.2240-2829/4154

[13] D. Guidetti, Some inverse problems of identification for integrodifferential parabolic systems with a boundary memory term, *Discrete & Continuous Dynamical Systems – S*, **8**(2015), no. 4, 749–756. DOI: 10.3934/dcdss.2015.8.749

[14] A.L.Karchevsky, A.G. Fatianov, Numerical solution of the inverse problem for a system of elasticity with the aftereffect for a vertically inhomogeneous medium, *Sib. Zh. Vychisl. Mat.*, **4**(2001), no. 3, 259–268.

[15] U.D.Durdiev, Numerical method for determining the dependence of the dielectric permittivity on the frequency in the equation of electrodynamics with memory, *Sib. Elektron. Mat. Izv.*, **17**(2020), 179–189.

[16] Z.R.Bozorov, Numerical determining a memory function of a horizontally-stratified elastic medium with aftereffect, *Eurasian journal of mathematical and computer applications*, **8**(2020), no. 2, 4–16.

[17] S.I.Kabanikhin, A.L.Karchevsky, A.Lorenzi, Lavrent'ev Regularization of Solutions to Linear Integro-differential Inverse Problems, *J. Inverse Ill-Posed Probl.*, **2**(1993), no. 1, 115–140.

[18] A.Ladyzhenskaya, V.A.Solonnikov, N.N.Ural'tseva, Linear and quasilinear equations of parabolic type, Moscow, Nauka, 1967.

Глобальная разрешимость задачи определения ядра в двумерном уравнении теплопроводности с памятью

Дурдимурод К. Дурдиев
Бухарское отделение Института математики АН РУз
Государственный университет Бухары
Бухара, Узбекистан
Жавлон З. Нуриддинов
Государственный университет Бухары
Бухара, Узбекистан

Аннотация. В статье исследуется обратная задача определения двумерного ядра интегрального члена, зависящего от временной переменной t и первой компоненты пространственной переменной (x, y) в интегро-дифференциальном уравнении теплопроводности. Для этого уравнения при заданном ядра изучается прямая начально-краевая задача с условиями Неймана на границе прямоугольной области. С помощью функции Грина эта задача сводится к интегральному уравнению вольтерровского типа второго рода, а затем методом последовательных приближений доказывается существование единственного решения. В обратной задаче в качестве условия переопределения используется решение прямой задачи на плоскости $y = 0$. Обратная задача заменяется эквивалентной вспомогательной задачей, более удобной для дальнейшего исследования. Далее эта задача сводится к системе интегральных уравнений второго рода относительно неизвестных функций. Применяя к этой системе теорему о неподвижной точке в классе непрерывных по времени со значениями в пространствах Гёльдера функций с экспоненциальной весовой нормой, доказывается основной результат статьи, состоящий в глобальной теореме существования и единственности решения обратной задачи.

Ключевые слова: интегро-дифференциальное уравнение, обратная задача, Теорема Банаха, существование, единственность.

EDN: GMRCJR
УДК 517.9

Characteristic Determinant of a Perturbed Regular Third-order Differential Operator on an Interval

Nurlan S. Imanbaev*
Zhanibekov South Kazakhstan Pedagogical University
Shymkent, Kazakhstan
Institute of Mathematics and Mathematical Modeling
Almaty, Kazakhstan

Received 10.08.2024, received in revised form 13.09.2024, accepted 01.11.2024

Abstract. In the paper, we consider the spectral problem for a third-order operator with an integral perturbation of one of the boundary value conditions that are regular and at the same time strongly regular; a feature of the problem is that the conjugate operator will be a loaded third-order differential operator with regular (strongly regular) boundary value conditions. Moreover, a characteristic determinant of the spectral problem is constructed, on the basis of which conclusions about eigenvalues of the perturbed operator are assumed.

Keywords: differential operator, integral perturbation, conjugate operator, loaded, eigenvalues, eigenfunctions, characteristic determinant, entire function.

Citation: N.S. Imanbaev, Characteristic Determinant of a Perturbed Regular Third-order Differential Operator on an Interval, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 25–31. EDN: GMRCJR.

Introduction and problem statement

In the functional space $W_2^3(0,1)$, we consider a differential operator L_1, given by the following expression:

$$l(u) = -u'''(x) = \lambda u(x), \quad 0 < x < 1 \tag{1}$$

with the "perturbed" boundary value conditions:

$$U_1(u) \equiv u(0) = \int_0^1 \overline{P(x)} u(x)\,dx, \quad U_2(u) \equiv u'(0) = 0, \quad U_3(u) \equiv u(1) = 0, \tag{2}$$

where λ is a spectral parameter, $U_1(u), U_2(u), U_3(u)$ are linear homogeneous independent forms, regular by G. D. Birkhoff [1, 2], $P(x) \in L_2(0,1)$. In the monography by M. A. Naimark ([3], p. 67) it was noted, that all differential operators of odd order with strongly regular boundary value conditions are with regular boundary value conditions. A third-order linear differential operator with nonlocal boundary value conditions under integral perturbation was studied in [4], and eigenvalue problems with periodic boundary value conditions were studied in [5–7], which are regular boundary value conditions. The questions of regularity and strong regularity of boundary value conditions for the Sturm–Liouville operator are related to the questions of basis property of the system of root vectors. In this case, when the boundary value conditions are strongly regular, the results of V. P. Mikhailov [8] and G. M. Keselman [9] imply the Riesz basis property of the systems of eigen- and associated functions of the Sturm–Liouville operator in $L_2(0,1)$. For

*imanbaevnur@mail.ru https://orcid.org/0000-0002-5220-9899

regular boundary value conditions, A. A. Shkalikov [10] proved unconditional basis property with brackets, and A. M. Minkin [11] proved the opposite statement, namely, that the unconditional basis property of a system of root vectors implies regularity of the operator. In [12–17], the issues of stability and instability of the basis property of systems of eigen- and associated functions of the multiple differentiation operator with regular, but not strongly regular, boundary value conditions were studied, that is, under integral perturbation of one of the boundary value conditions.

Calculation of eigenvalues and eigenfunctions of third-order equations of composite type is described in [18]. The spectrum of the boundary value problem with shift for the wave equation was studied in [19].

We will construct the characteristic determinant of the "perturbed" spectral problem (1)–(2) for the operator L_1. Based on the obtained formula, conclusions about the eigenvalues of the operator L_1 are established.

Conjugate problem

We define the operator L_1^*. By using Lagrange formula $(L_1, u, \nu) - (u, L_1^*, \nu) = \int\limits_0^1 l\left(u\right)\overline{v\left(x\right)}dx - \int\limits_0^1 u\left(x\right)\overline{l^*\left(v\right)}dx$, for all functions $u \in D\left(L_1\right)$ and $v \in D\left(L_1^*\right)$, and the boundary value conditions (2), we find:

$$-\int_0^1 u'''\left(x\right)\overline{v\left(x\right)}dx = \overline{v\left(0\right)}u''\left(0\right) - \overline{v\left(1\right)}u''\left(1\right) + u'\left(1\right)\overline{v'\left(1\right)} -$$

$$-u'\left(0\right)\overline{v'\left(0\right)} - u\left(1\right)\overline{v''\left(1\right)} + \int_0^1 u\left(x\right)dx\left[P\left(x\right)\cdot\overline{v''\left(0\right)} + \overline{v'''\left(x\right)}\right].$$

Due to linear independence of the forms $U_1\left(u\right), U_2\left(u\right), U_3\left(u\right)$ and $V_1\left(v\right), V_2\left(v\right), V_3\left(v\right)$, we get, that the operator L_1^* is given by the loaded differential expression:

$$L_1^* v \equiv l^*\left(v\right) = v'''\left(x\right) + P\left(x\right)v''\left(0\right) = \overline{\lambda}v\left(x\right), \quad 0 < x < 1,\ P\left(x\right) \in L_2\left(0,1\right) \tag{1*}$$

and the boundary value conditions:

$$V_1\left(v\right) \equiv v\left(0\right) = 0, \quad V_2\left(v\right) = v\left(1\right) = 0, \quad V_3\left(v\right) = v'\left(1\right) = 0, \tag{2*}$$

which is one of the features of the considered spectral problem (1)–(2).

Construction of the characteristic determinant of the spectral problem (1)-(2)

The general solution of the equation (1) has the form:

$$u\left(x\right) = C_1 e^{2\rho x} + \left(C_2\cdot\cos\sqrt{3}\rho x + C_3\cdot\sin\sqrt{3}\rho x +\right)e^{-\rho x}, \tag{3}$$

where C_1, C_2, C_3 are arbitrary constants.

$$\rho = \frac{\sqrt[3]{-\lambda}}{2}. \tag{4}$$

Putting (3) into the boundary value condition (2), we will have the linear system with respect to coefficients C_1, C_2, C_3:

$$\begin{cases} C_1\left(1-\int\limits_0^1 \overline{P(x)}e^{2\rho x}dx\right)+C_2\left(1-\int\limits_0^1 \overline{P(x)}e^{-\rho x}\cos\sqrt{3}\rho x dx\right)- \\ -C_3\int\limits_0^1 \overline{P(x)}e^{-\rho x}\sin\sqrt{3}\rho x dx=0, \\ 2C_1-C_2+C_3\sqrt{3}=0, \\ C_1\cdot e^{3\rho}+C_2\cos\sqrt{3}\rho+C_3\sin\sqrt{3}\rho=0. \end{cases}$$

Determinant of this system will be the characteristic determinant of the spectral problem (1)–(2):

$$\Delta_1(\rho)=\begin{vmatrix} 1-\int\limits_0^1 \overline{P(x)}e^{2\rho x}dx & 1-\int\limits_0^1 \overline{P(x)}e^{-\rho x}\cos\sqrt{3}\rho x dx & -\int\limits_0^1 \overline{P(x)}e^{-\rho x}\sin\sqrt{3}\rho x dx \\ 2 & -1 & \sqrt{3} \\ e^{3\rho} & \cos\sqrt{3}\rho & \sin\sqrt{3}\rho \end{vmatrix}=$$

$$=\left(-\sin\sqrt{3}\rho-\sqrt{3}\cos\sqrt{3}\rho\right)\left(1-\int_0^1 \overline{P(x)}e^{2\rho x}dx\right)-$$

$$-\left(2\sin\sqrt{3}\rho-\sqrt{3}e^{3\rho}\right)\left(1-\int_0^1 \overline{P(x)}e^{-\rho x}\cos\sqrt{3}\rho x dx\right)=$$

$$=\left(-\sin\sqrt{3}\rho-\sqrt{3}\cos\sqrt{3}\rho\right)\left(1-\int_0^1 \overline{P(x)}e^{2\rho x}dx\right)-\left(2\sin\sqrt{3}\rho-\sqrt{3}e^{3\rho}\right)\times$$

$$\times\left(1-\int_0^1 \overline{P(x)}e^{-\rho x}\cos\sqrt{3}\rho x dx\right)-\left(2\cos\sqrt{3}\rho+e^{3\rho}\right)\int_0^1 \overline{P(x)}e^{-\rho x}\sin\sqrt{3}\rho x dx. \qquad (5)$$

When $P(x)=0$ we get characteristic determinant of the "unperturbed" spectral problem for the operator L_0, given by the differential expression (1), with boundary value conditions:

$$u(0)=0,\ u'(0)=0,\ u(1)=0. \qquad (6)$$

We denote it by $\Delta_0(\rho)=-3\sin\sqrt{3}\rho-\sqrt{3}\cos\sqrt{3}\rho+\sqrt{3}e^{3\rho}$.

Following the results of [5, 6], we equate the determinant $\Delta_0(\rho)$ to zero, and obtain the condition for existence of nontrivial solutions $u(x)$:

$$\sqrt{3}\sin\sqrt{3}\rho+\cos\sqrt{3}\rho=e^{3\rho}. \qquad (7)$$

Equation (7) can have a solution if $\rho\leqslant 0$.

Equation (7) we reduce to the form

$$2\cos\frac{\pi}{6}\sin\sqrt{3}\rho+2\sin\frac{\pi}{6}\cos\sqrt{3}\rho=e^{3\rho},$$

i.e.

$$\sin\left(\frac{\pi}{6}+\sqrt{3}\rho\right)=\frac{1}{2}e^{3\rho},\ \ \rho<0 \qquad (8)$$

Eigenvalues of the "unperturbed" operator L_0 will be the roots of this equation. These roots are defined as the abscissa of the intersection points of the curves

$$y=\sin\left(\frac{\pi}{6}+\sqrt{3}\rho\right),\quad y=\frac{1}{2}e^{3\rho x},\ x\leqslant 0.$$

When $x = 0$ both curves have a common point $y = \frac{1}{2}$. Zero points of the functions $\sin\left(\frac{\pi}{6} + \sqrt{3}\rho\right)$ will be $\mu_k = -\frac{\pi}{6\sqrt{3}} - (k-1)\pi$, $k = 1, 2, \ldots$. Then for the eigenvalues ρ_k of the operator L_0, we obtain

$$\rho_k = \mu_k + (-1)^k \varepsilon_k, \; k = 1, 2, ... \tag{9}$$

moreover, $\lim\limits_{k\to\infty} \varepsilon_k = 0$.

Eigenfunctions of the operator L_0 will be functions

$$u_k(x) = e^{2\rho_k x} - \left(\cos\sqrt{3}\rho_k x + \sqrt{3}\sin\sqrt{3}\rho_k x\right) e^{-\rho_k x}. \tag{10}$$

The function (10) satisfies all conditions of the problem (1)–(6).

Conjugate operator L_0^* has the form:

$$L_0^* v = l_0^*(v) = v'''(x) - \bar{\lambda} v(x) = 0, \quad v(0) = 0, \; v(1) = 0, \; v'(1) = 0. \tag{11}$$

Eigenvalues of the operator L_0^* coincide with eigenvalues of the operator L_0 and the corresponding eigenfunctions of the operator L_0^* will be the functions

$$v_k(x) = e^{-2\rho_k x} - \left(\cos\sqrt{3}\rho_k x + N\sin\sqrt{3}\rho_k x\right) e^{\rho_k x}, \tag{12}$$

where $N = \dfrac{\cos\sqrt{3}\rho_k - e^{3\rho_k}}{\sin\sqrt{3}\rho_k}$.

The function $P(x)$ under the integral in (5) we represent in the form of a biorthogonal expansion in Fourier series by the system $\{v_k(x)\}$:

$$P(x) = \sum_{k=1}^{\infty} a_k v_k(x). \tag{13}$$

Calculating the integrals in (5), taking (12) and (13) into account, we will have the following form of the determinant $\Delta_1(\rho)$:

$$\begin{aligned}
\Delta_1(\rho) = &\left(-\sin\sqrt{3}\rho - \sqrt{3}\cos\sqrt{3}\rho\right) + \left(\sin\sqrt{3}\rho - \sqrt{3}\cos\sqrt{3}\rho\right) \times \\
&\times \left(\sum_{k=1}^{\infty} \alpha_k \left[\frac{e^{2(\rho-\rho_k)}}{2(\rho-\rho_k)} - \frac{1}{2(\rho-\rho_k)} + \frac{2\rho+\rho_k}{(2\rho+\rho_k)^2 + 3\rho_k^2} - \frac{e^{2\rho+\rho_k}}{(2\rho+\rho_k)^2 + 3\rho_k^2} \times \right.\right. \\
&\times \left(\sqrt{3}\rho_k \sin\sqrt{3}\rho_k + (2\rho+\rho_k)\cos\sqrt{3}\rho_k\right) + N\frac{\sqrt{3}\rho_k}{(2\rho+\rho_k)^2 + 3\rho_k^2} + \\
&\left.\left. + N\frac{e^{2\rho+\rho_k}}{(2\rho+\rho_k)^2 + 3\rho_k^2}\left((2\rho+\rho_k)\sin\sqrt{3}\rho_k - \sqrt{3}\cos\sqrt{3}\rho_k\right)\right]\right) - \\
&- \left(2\sin\sqrt{3}\rho - \sqrt{3}e^{3\rho}\right) + \left(2\sin\sqrt{3}\rho - \sqrt{3}e^{3\rho}\right)\left(\sum_{k=1}^{\infty}\alpha_k\left[\frac{2\rho+\rho_k}{(2\rho_k+\rho)^2 - \left(\sqrt{3}\rho\right)^2} + \right.\right. \\
&+ \frac{e^{-(2\rho+\rho_k)}}{(2\rho_k+\rho)^2 - \left(\sqrt{3}\rho\right)^2}\left(\sqrt{3}\rho\sin\sqrt{3}\rho - (2\rho_k+\rho)\cos\sqrt{3}\rho\right) + \frac{1}{8(\rho_k-\rho)} - \\
&- \frac{e^{\rho_k-\rho}}{8(\rho_k-\rho)} \cdot \left(\cos\sqrt{3}(\rho_k-\rho) + \sqrt{3}\sin\sqrt{3}(\rho_k-\rho)\right) - \frac{1}{2}\cdot\frac{\rho_k-\rho}{(\rho_k-\rho)^2 + 3(\rho_k+\rho)^2} + \\
&+ \frac{1}{2}\cdot\frac{e^{\rho_k-\rho}}{(\rho_k-\rho)^2 + 3(\rho_k+\rho)^2}\cdot\left(\sqrt{3}(\rho_k+\rho)\sin\sqrt{3}(\rho_k+\rho) + (\rho_k-\rho)\cos\sqrt{3}(\rho_k+\rho)\right) + \\
&+ \frac{N}{2}\cdot\frac{\sqrt{3}(\rho_k+\rho)}{(\rho_k-\rho)^2 + 3(\rho_k+\rho)^2} + \frac{N}{2}\cdot\frac{e^{\rho_k-\rho}}{(\rho_k-\rho)^2 + 3(\rho_k+\rho)^2}\left((\rho_k+\rho)\sin\sqrt{3}(\rho_k+\rho) - \right.
\end{aligned}$$

$$-\sqrt{3}\,(\rho_k+\rho)\cos\sqrt{3}\,(\rho_k+\rho)\Big)+\frac{1}{8}\cdot\frac{N}{\rho_k-\rho}+\frac{1}{8}\cdot\frac{N}{\rho_k-\rho}e^{\rho_k-\rho}\cdot\Big(\sin\sqrt{3}\,(\rho_k-\rho)-$$

$$-\cos\sqrt{3}\,(\rho_k+\rho)\Big)\Big]\Big)-\Big(2\cos\sqrt{3}\rho+e^{3\rho}\Big)\cdot\Bigg(\sum_{k=1}^{\infty}\alpha_k\Bigg[\frac{\sqrt{3}\rho}{\left(2\rho_k^2+\rho\right)^2+3\rho^2}-\frac{1}{4}\cdot\frac{\sqrt{3}}{\rho_k-\rho}+$$

$$+\frac{1}{4}\cdot\frac{\sqrt{3}\,(\rho_k+\rho)}{\rho_k+2\rho}-\frac{N}{8}\cdot\frac{1}{\rho_k-\rho}+\frac{1}{2}\cdot\frac{\rho_k-\rho}{3(\rho_k+\rho)^2+(\rho_k-\rho)^2}-\frac{e^{-(2\rho_k+\rho)}}{\left(2\rho_k^2+\rho\right)^2+3\rho^2}\times$$

$$\times\Big(\sqrt{3}\rho\cos\sqrt{3}\rho+(2\rho_k+\rho)\sin\sqrt{3}\rho\Big)+\frac{1}{4}\cdot\frac{e^{\rho_k-\rho}}{\rho_k-\rho}\Big(\sqrt{3}\cos\sqrt{3}\,(\rho_k-\rho)+\sin\sqrt{3}\,(\rho_k-\rho)\Big)+$$

$$+\frac{1}{4}\cdot\frac{e^{\rho_k-\rho}}{\rho_k+2\rho}\Big(\sin\sqrt{3}\,(\rho_k+\rho)-\sqrt{3}\,(\rho_k+\rho)\cos\sqrt{3}\,(\rho_k+\rho)\Big)+\frac{1}{8}N\frac{e^{\rho_k-\rho}}{\rho_k-\rho}\Big(\sqrt{3}\sin\sqrt{3}\,(\rho_k-\rho)+$$

$$+\cos\sqrt{3}\,(\rho_k-\rho)\Big)-\frac{1}{2}\cdot\frac{\rho_k-\rho}{3(\rho_k+\rho)^2+(\rho_k-\rho)^2}\Big((\rho_k-\rho)\cos\sqrt{3}\,(\rho_k+\rho)+$$

$$+\sqrt{3}\,(\rho_k+\rho)\sin\sqrt{3}\,(\rho_k+\rho)\Big)\Big]\Bigg), \tag{14}$$

where $N=\dfrac{\cos\sqrt{3}\rho_k-e^{3\rho_k}}{\sin\sqrt{3}\rho_k}$, which is an entire analytic function of the variable ρ, since at the points $\rho=\rho_k$ it has poles of first order, at these points the determinant $\Delta_0(\rho)$ has roots.

Therefore, we have proved

Theorem 1. *We represent the characteristic determinant of the "perturbed" spectral problem (1)–(2) in the form (14), where the determinant $\Delta_0(\rho)$ is the characteristic determinant of the "unperturbed" spectral problem (1)-(6), α_k are Fourier coefficients of the biorthogonal expansion (13) of the functions $P(x)$ by the system of eigenfunctions $\{v_k(x)\}$ of the conjugate unperturbed spectral problem (1)–(2).*

Remark 1. The "perturbed" spectral problem (1)–(2) is reduced to the study of zeros of the entire analytic function $\Delta_1(\rho)$. The study of zeros of an entire analytic function $\Delta_1(\rho)$ remains open.

The research has been funded by the grant (no. AP23485279) of the Science Commitee of the Ministry of Science and Higher Education of the Republic of Kazakhstan.

References

[1] G.D.Birkhoff, On the asymptotic character of the solutions of certain linear differential equations containing a parameter, *Transactions of the American Mathematical Society*, **9**(1908), no. 2, 219–231

[2] G.D.Birkhoff, Boundary value and expansion problems of ordinary linear differential equations, *Transactions of the American Mathematical Society*, **9**(1908), no. 4, 373–395

[3] M.A.Naimark, Linear Differential Operators, Moscow, Nauka, 1969 (in Russian)

[4] N.S.Imanbaev, B.E.Kangushin, B.T.Kalimbetov, On zepos of the characteristic determinant of the spectral problem for a third-order differential operator on a segment with nonlocal boundary conditions, *Advances in Difference Equations*, **2013**(2013), article 110. DOI: 10.1186/1687-1847-2013-110.

[5] N.S.Imanbaev, Distribution of eigenvalues of a third-order differential operator with strongly regular nonlocal boundary conditions, *AIP Conference Proceedings*, **1997**(2018), 020027. DOI: 10.1063/1/5049021

[6] N.S.Imanbaev, Ye.Kurmysh, On computation of eigenfunctions of composite type equations with regular boundary value conditions, *International journal of Applied Mathematics*, **34**(2021), no. 4, 681–692. DOI: 10.12732/ijam.v34i4.7

[7] A.I.Kozhanov, S.V.Potapova, Conjugation problem for a third-order equation with multiple characteristics, with an alternating function at the highest derivative, *Journal of Mathematical Sciences*, **215**(2016), no. 4, 510–516. DOI: 10.1007/s10958-016-2855-5

[8] V.P.Mikhailov, On Riesz basis in, *Doklady Mathematics*, **144**(1962), no. 5, 981–984 (in Russian)

[9] G.M.Kesel'man, About unconditional convergence of expansion by eigenfunctions of some differential operators, *Izv. Vyssh. Ucheb. Zaved. Math.*, **2**(1964), 82 (in Russian)

[10] A.A.Shkalikov, On the basis problem of the eigenfunctions of an ordinary differential operator, *Russian Mathematical Surveys*, **34**(1979), no. 5, 249–250. DOI: 10.1070/RM1979v034n05ABEH003901

[11] A.M.Minkin, Resolvent growth and Birkhoff-regularity, *Journal of Mathematical Analysis and Applications*, **323**(2006), no. 1, 387–402

[12] M.Kandemir, O.Sh.Mukhtarov, A method on solving irregular boundary value problems with transmission conditions, *Kuwait journal of Science and Eingeneering*, **36**(2A), 2009, 79–98

[13] N.S.Imanbaev, M.A.Sadybekov, Stability of basis property of a type of problems on eigenvalues with nonlocal perturbation of boundary conditions, *Ufa mathematical journal*, **3**(2011), no. 2, 27–32.

[14] N.S.Imanbaev, M.A.Sadybekov, Stability of basis property of a periodic problem with nonlocal perturbation of boundary conditions, *AIP Conference Proceedings*, **1759**(2016), 020080. DOI: 10.1063/1.4959694

[15] N.S.Imanbaev, On a problem that does not habe basis property of root vectors, associated with a perturbed regular operator of multiple differentiation, *Journal of Siberian Federal University. Mathematics and Physics*, **13**(2020), no. 5, 568–573. DOI: 10.17516/1997-1397-2020-13-5-568-573

[16] M.A.Sadybekov, N.S.Imanbaev, On system of root vectors of perturbed regular second-order differential operator not possessing basis property, *Mathematics*, **11**(2023), 4364. DOI: 10.3390/math11204364

[17] A.M.Sarsenbi, Criteria for the Riesz basis property of systems of eigen-and associated functions for higner-order differential operators on an interval, *Doklady Mathematics*, **77**(2008), no. 2, 290–292

[18] T.Dzh.Dzhuraev, On spectral problems for third-order equations of composite type, *Reports of the Academy of Sciences of the Republic of Uzbekistan*, **2**(2006), 5–8 (in Russian)

[19] T.Sh.Kal'menov, Selfadjoint boundare value problems for the Tricomi equation, *Differentsial'nye Uravneniya*, **19**(1983), no. 1, 66–75 (in Russian)

Характеристический определитель возмущенного регулярного дифференциального оператора третьего порядка на отрезке

Нурлан С. Иманбаев
Южно-Казахстанский педагогический университет имени Жанибекова
Шымкент, Казахстан
Институт математики и математического моделирования
Алматы, Казахстан

Аннотация. В работе рассматривается спектральная задача для оператора третьего порядка при интегральном возмущении одного из краевых условий, являющихся регулярными, одновременно усиленно регулярными, где особенностью задачи является, что сопряжённым оператором будет нагруженный дифференциальный оператор третьего порядка с регулярными (усиленно регулярными) краевыми условиями. Построен характеристический определитель спектральной задачи, на основании которой предпологаются выводы об собственных значениях возмущённого оператора.

Ключевые слова: дифференциальный оператор, интегральное возмущение, сопряженный оператор, нагруженный, собственные значение, собственные функции, характеристический определитель, целая функция.

Journal of Siberian Federal University. Mathematics & Physics 2025, 18(1), 32–40

EDN: CRYONU
УДК 524.8

Existence of Bianchi type Cosmological Phantom universe with Polytropic EoS Parameter in Lyra's Geometry

Suryanarayana Kornu*
Rajamahanthi Santhikumar† A. Lakshmana Rao‡
M. Ramanamurty§
P. Sudheer Kumar¶
K. Seetharam‖
B. Divya**
Aditya Aditya Institute of Technology and Management
K. Kotturo Tekkali
Srikakulam-532201, Andhra Pradesh, India

Received 30.05.2024, received in revised form 27.07.2024, accepted 10.09.2024

Abstract. This study explores the Bianchi type-V cosmological model with in the frame work of general relativity, featuring a perfect fluid governed by the polytropic equation in Lyra's Geometry, expressed $p = \alpha\rho + k\rho^n$, as proposed at [1]. We considered the case representing phantom universe for$(1 + \alpha + \mathrm{k}\rho^{n-1}) \leqslant 0$, $k < 0$, where ρ increases with the radius $a(t)$. The role of Lyra;s Geometry has been discussed/ The solution to Einsteins field equation have been derived, providing insights into the physical and cosmological attributes of this particular model.

Keywords: accelerated expansion, polytropic Eequation of state, perfect fluid, phantom universe, Lyra's geometry.

Citation: S. Kornu, R. Santhikumar, A.L. Rao, M. Ramanamurty, P.S. Kumar, K. Seetharam, B. Divya, Existence of Bianchi type Cosmological Phantom universe with Polytropic EoS Parameter in Lyra's Geometry, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 32–40. EDN: CRYONU.

Introduction

In modern cosmology, the present combination of the universe's energy is delineated as around 5% ordinary matter, 20% dark matter, and 75% dark energy [1] The universe's expansion initiated with a remarkable inflationary surge propelled by vacuum energy. Between 10^{-35} and 10^{-33} seconds after the onset of the Big Bang, the universe underwent a staggering expansion by a factor of 10^{30} [2–4]. However, inflation fails to provide an explanation for the time preceding the origin of the universe. As a result, the universe transitioned into the radiation era and, as the temperature decreased less than 10^3 K, the matter era advanced [5]. Currently, the universe undergoes accelerated expansion [6], attributed to either the cosmological constant or a form of dark energy with negative pressure violating the strong energy condition [7]. This phase represents a second inflationary period, which is distinct from the initial one. However, the nature of

*suryanarayana.kornu@gmail.com https://orcid.org/0000-0002-8793-3588
†skrmahanthi@gmail.com https://orcid.org/0000-0001-5122-3800
‡agatamudi111@gmail.com https://orcid.org/0000-0001-7651-9376
§ramanamurtymuddada@gmail.com
¶sudheerkumar9732@gmail.com https://orcid.org/0009-0009-4086-486X
‖karanamuseetharam@gmail.com https://orcid.org/0000-0002-1770-5811
**baratamdivya123@gmail.com

the pre-radiation era (very early universe), dark matter and dark energy has remained enigmatic and mysterious, prompting speculation. The nature of the early universe in the context of the Big Bang theory is not understood [8]. The characteristic s of the universe are not consistent with weather the universe was hot or cold state. The investigation of cosmic microwave background radiation (CMB) was achieved via the hot universe theory [9–11] therefore, we accepted the Big Bang theory. According to that theory, the early universe was combined with an ultra-relativistic classical gas can present a photon, electrons, positrons, quarks, antiquarks, etc. It is clearly observed that at the early stage of the universe, the scale factor vanishes while the energy density and temperature become infinite. This situation reveals the initial cosmological singularity commonly known as the Big Bang. The up-to-date studies on a cosmological model based on Bose–Einstein condensate dark matter (BEC DM). The Bose–Einstein condensate EoS [12–15] can be generalized as $p = \alpha\rho + k\rho^2$. If k $>$ 0 the model represents repulsive, if k $<$ 0, the model represents attractive self-interaction and when k = 0, the standard linear equation of state $p = \alpha\rho c^2$.

This paper delves into the examination of the modified equation of state, as proposed by [1].

$$p = \alpha\rho + k\rho^n \tag{1}$$

The author of [16, 17] reviewed different available sources and studied the likeness between the polytropic equation of state and a cosmological model, where the fluid that fills the universe has an effective bulk viscosity. The author of [18] suggested that a polytropic gas model. The authors of [19] used the available information from different sources and studied the polytropic inflationary model in brane world scenario. The author [20] studied Bianchi-I Dark with polytropic DE. The authors of [21] studied the Kantowski universe with the Polytropic EoS parameter. The authors of [22] conducted a remarkable study on anisotropic and homogeneous cosmological models with the polytropic EoS parameter. The authors of [23–25] performed an intended study on Bianchi-type-V cosmological models in modified theories of gravitation.

Einstein's field equations, integral to comprehending the uniformity and static model of the universe, impose constraints allowing only dynamic cosmological models for non-zero energy density. Consequently, Einstein's general theory of relativity becomes interpretable in terms of geometry. Following the advent of general relativity, Weyl [26] in 1918 expanded Riemannian geometry, applying it to physical contexts to formulate the initial unified theory encompassing gravity and electromagnetism. However, Einstein vehemently opposed Weyl's unified theory, leading to its neglect for over several years. Subsequently, Lyra [27] introduced a gauge function, effectively modifying Riemannian geometry and rendering it a structureless manifold. This alteration naturally gave rise to cosmological constants within the fabric of the universe's geometry. Since we have performed remarkable research, we are interested in continuing our research with this paper on the Bianchi type-V model and the energy-momentum tensor consisting of a perfect fluid with polytropic equation of state in Lyra's Geometry. In the case $(1 + \alpha + k\rho^{n-1}) \leqslant 0$, $k < 0$ the universe is considered to represent the phantom universe.

This document is structured as subsequent sections: In Section 2 Field equations of these metrics are obtained by Bianchi type-V metric in the existence of a perfect fluid with the polytropic EoS parameter in Lyra's Geometry. Section 3 is dedicated to solution of the model. Section 4 is devoted to the physical and geometric properties of the model. Section 5 — exact solution of the model.

1. Metric and field equations

Since Bianchi models are spatially homogeneous and anisotropic and are more appropriate for describing the universe because it has less symmetry than do standard FRW models. So, we considered here Bianchi type-V cosmological model.

$$ds^2 = dt^2 - A^2dx^2 - B^2e^{-2x}dy^2 - C^2e^{-2x}dz^2 \tag{2}$$

where A(t), B(t) and C(t) are the three anisotropic directions of expansion in normal three-dimensional space.
The average scale factor a(t), the spatial volume V and the average Hubble's parameter are

$$a\left(t\right) = \left(ABC\right)^{\frac{1}{3}} \tag{3}$$

$$V = a^3 = ABC \tag{4}$$

$$H = \frac{1}{3}(H_1 + H_2 + H_3 \tag{5}$$

Here, $H_1 = \frac{\acute{A}}{A}$, $H_2 = \frac{\acute{B}}{B}$ and $H_1 = \frac{\acute{C}}{C}$.

$$\frac{\acute{V}}{V} = \left(\frac{\dot{A}}{A} + \frac{\dot{B}}{B} + \frac{\dot{C}}{C}\right) = 3H \tag{6}$$

The Einstein modified the filed equation in normal gauge for Lyra's modified equation is given by (where $8\pi G = 1$, $C = 1$)

$$R_i^j - \frac{1}{2}g_i^jR + \left[\frac{3}{2}\phi_i\phi^j - \frac{3}{4}\phi_k\phi^kg_i^j\right] = -T_i^j \tag{7}$$

$$\phi_i = \left(\beta\left(t\right), 0, 0, 0\right) \tag{8}$$

where ϕ_i is the displacement vector Let T_{ij} be the energy- momentum tensor of the matter [9]. Additionally,

$$T^{ij}_{;j} \equiv 0 \tag{9}$$

The energy momentum tensor T_{ij}

$$T_{ij} = \left(p + \rho\right)u_iu_j - pg_{ij} \tag{10}$$

where ρ is the energy density, p is the pressure and u^i is the four velocity vectors satisfying $g_{ij}u^iu^j = 1$. The above perfect fluid obeys the polytropic equation of state.

$$p = \alpha\rho + k\rho^n \text{ with } \ k < 0 \tag{11}$$

The conservation equation $T^{ij}_{;j} \equiv 0$ leads to

$$\dot{\rho} + \left(p + \rho\right)\left(\frac{\dot{A}}{A} + \frac{\dot{B}}{B} + \frac{\dot{C}}{C}\right) = 0 \tag{12}$$

Using Eq. (11) and Eq. (12), we have

$$\dot{\rho} + 3H\rho(1 + \alpha + k\rho^{n-1}) = 0 \tag{13}$$

We considered the case here for the phantom universe is $(1 + \alpha + k\rho^{n-1}) \leqslant 0$, $k < 0$ [23], where $-1 \leqslant \alpha \leqslant 1$, k is the polytropic constant and n is the polytropic index. Moreover conservation of the L.H.S of Eq. (7) leads to

$$R_i^j - \frac{1}{2}g_i^jR + \left[\frac{3}{2}\phi_i\phi^j - \frac{3}{4}\phi_k\phi^kg_i^j\right] = 0 \tag{14}$$

$$\frac{3}{2}\phi_i\left[\frac{\partial\phi^j}{\partial x^j}+\phi^l\Gamma^j_{lj}\right]+\frac{3}{2}\phi^j\left[\frac{\partial\phi_i}{\partial x^j}-\phi_l\Gamma^l_{ij}\right]-\frac{3}{4}g^j_i\phi_k\left[\frac{\partial\phi^k}{\partial x^j}+\phi^l\Gamma^k_{lj}\right]-\frac{3}{4}g^j_i\phi^k\left[\frac{\partial\phi_k}{\partial x^j}-\phi_l\Gamma^l_{kj}\right]=0 \quad (15)$$

Eq. (10) leads to

$$\frac{3}{2}\beta\dot{\beta}+\frac{3}{2}\beta^2\left(\frac{\dot{A}}{A}+\frac{\dot{B}}{B}+\frac{\dot{C}}{C}\right)=0 \quad (16)$$

2. Solution and model

In a co-moving coordinate system, by Eqs. (7) and (10) we have

$$\frac{\ddot{B}}{B}+\frac{\ddot{C}}{C}+\frac{\dot{B}\dot{C}}{BC}-\frac{1}{A^2}+\frac{3}{4}\beta^2=p \quad (17)$$

$$\frac{\ddot{A}}{A}+\frac{\ddot{C}}{C}+\frac{\dot{A}\dot{C}}{AC}-\frac{1}{A^2}+\frac{3}{4}\beta^2=p \quad (18)$$

$$\frac{\ddot{A}}{A}+\frac{\ddot{B}}{B}+\frac{\dot{A}\dot{B}}{AB}-\frac{1}{A^2}+\frac{3}{4}\beta^2=p \quad (19)$$

$$\frac{\dot{A}\dot{B}}{AB}+\frac{\dot{B}\dot{C}}{BC}+\frac{\dot{A}\dot{C}}{AC}-\frac{3}{A^2}-\frac{3}{4}\beta^2=-\rho \quad (20)$$

$$\frac{2\dot{A}}{A}-\frac{\dot{B}}{B}+\frac{\dot{C}}{C}=0 \quad (21)$$

Integrating Eq. (21), we obtain

$$A^2=BC \quad (22)$$

Without loss of generality, subtracting (17) from (18) and (18) from (19) and taking the second integral of each, we obtain the following three relations.

$$\frac{A}{B}=c_1e^{\int\frac{d_1}{a^3}dt},\quad \frac{A}{C}=c_2e^{\int\frac{d_2}{a^3}dt}\quad \text{and}\quad \frac{B}{C}=c_3e^{\int\frac{d_3}{a^3}dt} \quad (23)$$

where c_1,c_2,c_3,d_1,d_2 and d_3 are constants of integration. Since, $a(t)=(ABC)^{\frac{1}{3}}$, by Eq. (17) and Eqs. (23) we obtain

$$A(t)=l_1ae^{\left(m_1\int a^{-3}dt\right)},\quad B(t)=l_2ae^{\left(m_2\int a^{-3}dt\right)},\quad C(t)=l_3ae^{\left(m_3\int a^{-3}dt\right)} \quad (24)$$

where $l_1=(d_1d_2)^{\frac{1}{3}}$, $m_1=\dfrac{k_1+k_2}{3}$, $l_2=\left(\dfrac{d_3}{d_1}\right)^{\frac{1}{3}}$, $m_2=\dfrac{k_3-k_1}{3}$, $l_3=(d_3d_2)^{-\frac{1}{3}}$, $m_3=\dfrac{-(k_3+k_2)}{3}$
The constants m_1,m_2,m_3 and l_1,l_2,l_3 will satisfy these two conditions:

$$m_1+m_2+m_3=0\quad\text{and}\quad l_1l_2l_3=1 \quad (25)$$

Using Eq. (22) in Eqs. (24), we obtain

$$l_1=1,\quad l_2=l_3^{-1}=k_1(say),\quad m_1=0,\quad m_2=-m_3=k_2(say) \quad (26)$$

By using Eq. (26) in Eqs. (25), we have

$$A(t)=a(t),\quad B(t)=k_1a(t).e^{\left(k_2\int(a(t))^{-3}dt\right)},\quad C(t)=\frac{a(t)}{k_1}.e^{\left(-k_2\int(a(t))^{-3}dt\right)} \quad (27)$$

By using Eq. (27) in Eqs. (2) we have

$$ds^2=dt^2-a^2dx^2-\left[k_1a(t).e^{\left(k_2\int(a(t))^{-3}dt\right)}\right]^2e^{-2x}dy^2-\left[\frac{a(t)}{k_1}.e^{\left(-k_2\int(a(t))^{-3}dt\right)}\right]^2e^{-2x}dz^2 \quad (28)$$

This represents the Bianchi type-V cosmological model with an average scale factor.

3. Physical and geometric properties

Solve the differential equation Eq. (13) in case of $(1+\alpha+k\rho^{n-1}) \leqslant 0$, $k<0$, represents the phantom universe, where the density increases with the radius. We obtain density, average scale factor and pressure as follows

$$\rho = \left[\frac{1+\alpha}{a^{-3(1+\alpha)(n-1)} - k}\right]^{\frac{1}{n-1}} \text{ for } (1+\alpha) > 0 \tag{29}$$

$$a\left(t\right) = \left[\frac{1+\alpha}{\rho^{n-1}} + k\right]^{3(1+\alpha)(n-1)} \tag{30}$$

$$p = \alpha\left[\frac{1+\alpha}{a^{-3(1+\alpha)(n-1)} - k}\right]^{\frac{1}{n-1}} + k\left[\frac{1+\alpha}{a^{-3(1+\alpha)(n-1)} - k}\right]^{\frac{n}{n-1}} \tag{31}$$

Clearly, $\rho\left(t\right) \propto \left[a(t)\right]^3$ indicates that the universe is considered to be the phantom universe.

$$\omega\left(t\right) = \frac{p}{\rho} = \alpha + k\left[\frac{1+\alpha}{a^{-3(1+\alpha)(n-1)} - k}\right] \tag{32}$$

When $(1+\alpha) > 0$ or $\alpha > -1$ and $k < 0$, the EoS parameter $\omega\left(t\right) < -1$ representing the model is the phantom universe, which leads to no Big Rip singularity. Solving Eq. (16) for the displacement vector $\beta(t)$, we have

$$\beta(t) = ca^{-3} \tag{33}$$

c is the integration constant Considering c=1 and substituting $a(t) = \beta^{-1/3}(t)$ in eqs. (28), (31) and (32), we have

$$\rho = \left[\frac{1+\alpha}{\beta^{(1+\alpha)(n-1)} - k}\right]^{\left(\frac{1}{n-1}\right)} \tag{34}$$

$$p = \alpha\left[\frac{1+\alpha}{\beta^{(1+\alpha)(n-1)} - k}\right]^{\left(\frac{1}{n-1}\right)} + k\left[\frac{1+\alpha}{\beta^{(1+\alpha)(n-1)} - k}\right]^{\left(\frac{n}{n-1}\right)} \tag{35}$$

$$\omega\left(t\right) = \alpha + k\left[\frac{1+\alpha}{\beta^{(1+\alpha)(n-1)} - k}\right. \tag{36}$$

$$ds^2 = dt^2 - \frac{dx^2}{\beta^{\frac{2}{3}}\left(t\right)} - \left[k_1 . \frac{e^{\left(k_2 \int \beta(t)dt\right)}}{\beta^{\frac{1}{3}}\left(t\right)}\right]^2 e^{-2x}dy^2 - \left[\frac{e^{\left(-k_2 \int \beta(t)dt\right)}}{k_1 \beta^{\frac{1}{3}}\left(t\right)}\right]^2 e^{-2x}dz^2 \tag{37}$$

Since, $\rho(t) \propto \left[a(t)\right]^3$ and $\rho(t) \propto \dfrac{1}{\beta(t)}$, it can be described with this displacement vector $\beta(t)$ in the late universe. Since the ρ increases with respect to a(t) the universe accelerating rapidly. At $n=-1$, the analytical model of phantom universe represents bouncing universe "disappearing" at t=0.

4. Exact solutions of the model

The significance of the hyperbolic function as an exponential component is that the derived deceleration parameter demonstrates time dependence, highlighting that the universe's is in the acceleration phase. Therefore, opting for this average scale factor is deemed physically acceptable. Consider an average scale factor $a(t)$ as by [26]

$$a\left(t\right) = \left(\sin ht\right)^{\frac{1}{3}} \tag{38}$$

Therefore, by Eqs. (32) and (35)

$$A(t) = (\sin ht)^{\frac{1}{3}} . \tag{39}$$

$$B(t) = k_1 (\sin ht)^{\frac{1}{3}} (\cot ht - \operatorname{cosec} ht)^{k_2} \tag{40}$$

$$C(t) = \frac{(\sin ht)^{\frac{1}{3}} (\cot ht - \operatorname{cosec} ht)^{k_2}}{k_1}. \tag{41}$$

Substituting Eqs. (38)–(41) in Eq. (2) we get

$$\begin{aligned} ds^2 = dt^2 - t^{\frac{2}{3}} dx^2 - \left[(\sin ht)^{\frac{1}{3}} (\cot ht - \operatorname{cosec} ht)^{k_2}\right]^2 e^{-2x} dy^2 - \\ - \left[\frac{(\sin ht)^{\frac{1}{3}} (\cot ht - \operatorname{cosec} ht)^{k_2}}{k_1}\right]^2 e^{-2x} dz^2. \end{aligned} \tag{42}$$

This represents the field equation of the Bianchi type-V cosmological model with polytropic EoS parameter in Lyra's geometry which is in phantom era.
We obtain the spatial volume V(t), density (ρ), pressure (p), EoS parameter $\omega(t)$, Hubble's parameter H(t), energy density parameter $\Omega(t)$, scalar expansion $\theta(t)$, deceleration parameter $q(t)$, anisotropy parameter, and shear scalar for the model are

$$V(t) = a^3 = \sin ht. \tag{43}$$

$$\rho = \left[\frac{1+\alpha}{(\sin ht)^{(1+\alpha)(n-1)} - k}\right]^{\left(\frac{1}{n-1}\right)} \tag{44}$$

$$p = \alpha \left[\frac{1+\alpha}{(\sin ht)^{(1+\alpha)(n-1)} - k}\right]^{\left(\frac{1}{n-1}\right)} + k \cdot \left[\frac{1+\alpha}{(\sin ht)^{(1+\alpha)(n-1)} - k}\right]^{\left(\frac{n}{n-1}\right)} \tag{45}$$

$$\omega(t) = \alpha + k \cdot \left[\frac{1+\alpha}{(\sin ht)^{(1+\alpha)(n-1)} - k}\right]^{\left(\frac{-1}{n-1}\right)} \tag{46}$$

$$H(t) = \cot ht \tag{47}$$

$$\Omega(t) = \frac{\rho}{3H} = \frac{\left[\frac{1+\alpha}{(\sin ht)^{(1+\alpha)(n-1)} - k}\right]^{\frac{1}{n-1}}}{3 \cot ht} \tag{48}$$

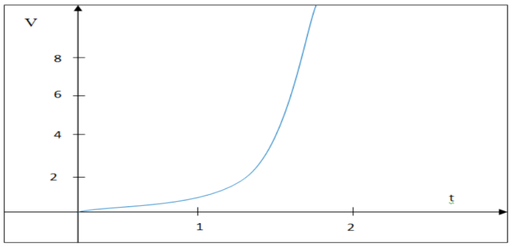
$$\theta(t) = 3H = 3 \cot ht \tag{49}$$

$$q(t) = \frac{d}{dt}\left(\frac{1}{H}\right) - 1 = \sec h^2 t - 1 \tag{50}$$

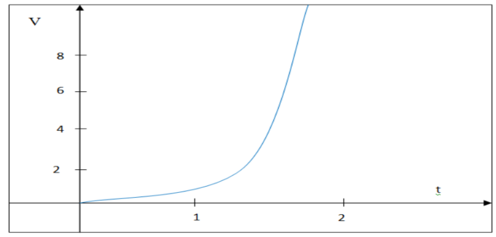
$$A_m = \frac{1}{3}\left(\frac{3}{\cot ht \cdot \sin ht}\right)^2. \tag{51}$$

$$\sigma^2 = \left(\frac{1}{\sin ht}\right)^2 \tag{52}$$

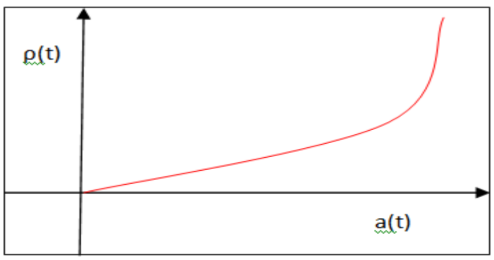
Here, we can note that spatial volume is zero for t=0 and that the scalar expansion is infinite, which is the big bang scenario. Additionally pressure (p), density (ρ), Hubble's(H) and shear scalar (σ) diverge in the early universe. As $t \to \infty$ increases the, volume becomes infinity where the pressure (p), density (ρ), Hubble parameter (H) and shear scalar (σ) approach zero. Since, $\lim_{t\to\infty} \dfrac{\rho}{\theta^2} = constant$ represents the anisotropic nature of the model.



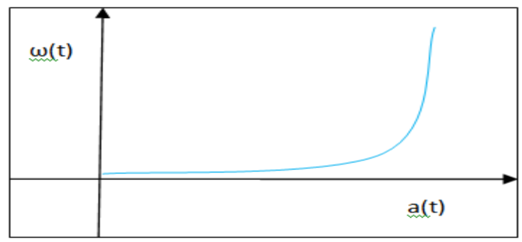
Volume V v/s Cosmic time t



displacement vector $\beta(t)$ v/s average scale factor a(t)



density ρ v/s average scale factor a(t)



EoS parameter $\omega\,(t)$ ρ v/s average scale factor a(t)

Conclusion

In this paper, we studied the Bianchi type-V cosmological model in the presence of a perfect fluid with polytropic equation of state. We considered a this case $(1+\alpha+k\rho^{n-1})<0$, $k<0$ representing that the model of the universe is phantom stage with increasing density with respect to average scale factor. We observe that the spatial volume is zero for t=0 and that the scalar expansion is infinite, which shows that the universe starts evolving with zero volume at t=0 which is the big bang scenario. We also, observed in this model that as time increases, the volume increases and becomes infinitely large $t\to\infty$. Clearly, $\rho\,(t)\propto[a\,(t)]^3$ and $\rho(t)\propto\dfrac{1}{\beta(t)}$ \$ indicates that the universe is considered to be the phantom universe, where the density increases with the radius (average scale factor). Moreover pressure increases with respect to the scale factor. The EoS parameter $\omega\,(t)<-1$ representing the model is the phantom universe and leads to no Big Rip singularity. When $n=-1$ this is the analytical model of the phantom bouncing universe "disappearing" at t=0. $\lim_{t\to\infty}\dfrac{\rho}{\theta^2}=constant$ indicates that this model is an anisotropic model.

References

[1] S.Nojiri, S.D.Odintsov, Inhomogeneous equation of state of the universe: Phantom era, future singularity, and crossing the phantom barrier, *Physical Review D*, **72**(2005), no. 2, 023003. DOI: 10.1103/PhysRevD.72.023003

[2] A.H.Guth, Inflationary universe: A possible solution to the horizon and flatness problems, *Physical Review D*, **23**(1981), no. 2, 347.

[3] A.D.Linde, A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems, *Physics Letters B*, **108**(1982), no. 6, 389–393.

[4] A.Albrecht, et al., Reheating an inflationary universe, *Physical Review Letters*, **48**(1982), no. 20, 1437.

[5] S.Weinberg, Principles and applications of the general theory of relativity, Gravitation and cosmology, (1972).

[6] A.G.Riess, et al., Observational evidence from supernovae for an accelerating universe and a cosmological constant, The astronomical journal, **116**(1998), no. 3 009.

[7] E.J.Copeland, S.Mohammad, T.Shinji, Dynamics of dark energy, *International Journal of Modern Physics D*, **15**(2006), no. 11 1753–1935.

[8] A.Friedmann, G.Lema?tre, E.de Cosmologie, Preceded by L'inventions du Big Bang by J.P. Luminent, Source du Savoir Seuil, 1997.

[9] G.Gamow, The origin of elements and the separation of galaxies, *Physical Review*, **74**(1948), no. 505.

[10] A.G.Doroshkevich, I.D.Novikov, Mean radiation density in Metagalaxy and some problems of relativistic cosmology, *Dokl. Akad. Nauk Nauk SSSR*, **154**(1964), no. 4, 809–811 (in Russian).

[11] P.J.Edwin.Peebles, The Black-Body Radiation Content of the Universe and the Formation of Galaxies,*Astrophysical Journal*, **142**(1965), 1317.

[12] P.-H.Chavanis, Models of universe with a polytrophic equation of state: II. The late universe, *The European physical journal plus*, **129**(2014), no. 10, 222.

[13] P.-H.Chavanis, Growth of perturbations in an expanding universe with Bose-Einstein condensate dark matter, *Astronomy and Astrophysics*, **537**(2012), no. A127.

[14] P.H.Chavanis, A simple model of universe with a polytropic equation of state, 2012, arXiv preprint arXiv:1208.1192.

[15] P.H.Chavanis, Models of universe with a polytropic equation of state: III. The phantom universe, 2012, arXiv preprint arXiv:1208.1185.

[16] J.D.Barrow, String-driven inflationary and deflationary cosmological models, *Nuclear Physics B*, **310**(1988), 743–763.

[17] P.-H.Chavanis, Predictive model of BEC dark matter halos with a solitonic core and an isothermal atmosphere, *Physical Review D*, **100**(2019), 083022.
DOI: 10.1103/PhysRevD.100.083022

[18] K.Karami, S.Ghaffari, J.Fehri, Interacting polytropic gas model of phantom dark energy in non-flat universe, *The European Physical Journal C*, **64**(2009), 85–88.

[19] M.Setare, A.Ravanpak, H.Farajollahi, Polytropic Inspired Inflation on the Brane, *Gravitation and Cosmology*, **24**(2018), 52–56.

[20] S.D.Katore, D.V.Kapse, Bianchi Type-I Dark Energy Cosmological Model With Polytropic Equation of State In Barber's Second Self-Creation Cosmology, *International Journal of Mathematics Trends and Technology-IJMTT*, **53**(2018).

[21] K.S.Adav, P.R.Agarwal, R.R.Saraogi, Kantowski Cosmological Model with Polytropic Equation of State, *The Afr. Rev. phys*, 12:0001 92017.

[22] K.S.Adav, P.R.Agarwal, R.R.Saraogi, Anisotropic and Cosmological Models with polytropic equation of state in General Relativity, *Bulg J. Phy.*, **43**(2016), 171–183

[23] D. R. K.Reddy, et al., Bianchi type-V bulk viscous string cosmological model in Saez-Ballester scalar-tensor theory of gravitation, *Astrophysics and Space Science*, **349**(2014) 473–477.

[24] M.P.V.V.Bhaskara Rao, D.R.K.Reddy, K.Sobhan Babu, Bianchi type-V bulk viscous string cosmological model in a self-creation theory of gravitation, *Astrophysics and Space Science*, **359**(2015), 1–5.

[25] N.Ahmed, A.Pradhan, Bianchi type-V cosmology in f (R,T) gravity with $\Lambda(T)$, *International Journal of Theoretical Physics*, **53**(2014), 289–306. DOI: 10.1007/s10773-013-1809-7

[26] H.Weyl, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin, 465, 1918.

[27] G.Lyra, Über eine modifikation der riemannschen geometrie, *Mathematische Zeitschrift*, **54**(1951), no 1, 52–64.

Существование космологического фантома типа Бьянки с политропным параметром уравнения состояния в геометрии Лиры

Сурьянараяна Корну
Раджамаханти Сантикумар
А. Лакшмана Рао
М. Раманамурти
П. Судхир Кумар
К. Ситхарам
Б. Дивья
Институт технологий и менеджмента Адитьи
К. Коттуро Теккали
Шрикакулам-532201, Андхра-Прадеш, Индия

Аннотация. Эта работа исследует космологическую модель Bianchi Type-V с в рамках общей теории относительности, в которой есть идеальная жидкость, управляемая политропическим уравнением в геометрии, выраженной $p = \alpha\rho + k\rho^n$, как предложено в citens. Мы рассмотрели случай, представляющий фантомную вселенную для $(1 + \alpha + \mathrm{k}\ \rho^{n-1}) \leqslant 0$, $k < 0$, где ρ увеличивается с радиусом $a(t)$. Была получена роль геометрии Лиры;

Ключевые слова: ускоренное расширение, политропическое уравнение состояния, идеальная жидкость, фантомная вселенная, геометрия Лиры.

Journal of Siberian Federal University. Mathematics & Physics 2025, 18(1), 41–50

EDN: GXJKAY
УДК 512.554

On Spectra and Minimal Polynomials in Finite Semifields

Olga V. Kravtsova*
Ilya K. Kuzmin†
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 10.08.2024, received in revised form 26.09.2024, accepted 01.11.2024

Abstract. We apply the notion of a one-side-ordered minimal polynomial to investigations in finite semifields. A proper finite semifield has non-associative multiplication, that leads to the anomalous properties of its left and right spectra. We obtain the sufficient condition when the right (left) order of a semifield element is a divisor of the multiplicative loop order. The interrelation between the minimal polynomial of non-zero element and its right (left) order is described using the spread set. This relationship fully explains the most interesting and anomalous examples of small-order semifields.

Keywords: semifield, right order, right spectrum, right-ordered minimal polynomial, spread set.

Citation: O.V. Kravtsova, I.K. Kuzmin, On Spectra and Minimal Polynomials in Finite Semifields, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 41–50. EDN: GXJKAY.

1. Introduction and preliminaries

The weakening of the field axioms leads to more general algebraic systems such as near-fields, semifields and quasifields. According to [1], a *semifield* is a set Q with two binary algebraic operations $+$ and $*$ such that:

1) $\langle Q, +\rangle$ is an abelian group with neutral element 0;

2) $\langle Q^*, *\rangle$ is a loop ($Q^* = Q \setminus \{0\}$);

3) both distributivity laws hold, $a*(b+c) = a*b + a*c$, $(b+c)*a = b*a + c*a$ for all $a, b, c \in Q$.

The first examples of non-trivial semifields (not the fields) were constructed by L. E. Dickson in 1906, the multiplicative law in a proper semifield is non-associative. By replacing the two-sided distributivity with a one-sided one, we get the concept of a *quasifield* (left or right). A quasifield with associative multiplication is a *near-field*. Unlike the finite near-fields, which were completely classified by H. Zassenhaus in 1936, neither semifields nor even quasifields have received an exhaustive classification by now.

The absence of associativity even in a finite semifield and a finite quasifield leads to it having a number of specific properties, which are poorly studied. The identification of structural features and anomalous properties is an important step in solving the classification problem of finite quasifields. The most complete review is presented by N. L. Johnson at al. in Handbook [2].

The following problems for finite proper quasifields were presented in 2013 by V. M. Levchuk at research seminar of chair of algebra of Moscow State University, see also [3].

*ol71@bk.ru https://orcid.org/0000-0002-6005-2393
†ilyabarinovy@gmail.com

(A) *Enumerate maximal subfields and their possible orders.*
(B) *Find the finite quasifields Q with not-one-generated loop Q^*.*
(C) *What loop spectra Q^* of finite semifields and quasifields are possible?*
(D) *Find the automorphism group $\mathrm{Aut}\, Q$.*

The notion of *spectrum* is used for quasifields and semifields taking into account the abcense of associativity. The product of m multipliers is said to be *m-th degree* of a fixed element $a \in Q^*$, if every multiplier coincides with a. The smallest integer $m \geqslant 1$ such that there exists the m-th degree of a, which is equal to the identity, is called *the order of a* and denoted by $|a|$. The set of orders of all elements is called the spectrum of multiplicative loop Q^*.

Similarly, using the right-ordered and the left-ordered m-th degrees

$$a^{m)} = a^{m-1)} * a, \quad a^{(m} = a * a^{(m-1}, \quad a^{1)} = a = a^{(1},$$

we define the right order $|a|_r$ and the left order $|a|_l$ of a and *the right and the left spectra* of Q^* respectively.

Even the weakened associativity of multiplication allows us to obtain important results about loops and, consequently, semifields and quasifields. Thus, Lagrange's theorem and some other classical group-theoretic theorems can be transferred to binary associative loops or *Moufang loops* (A. N. Grishkov, A. V. Zavarnitsin) [4]. In general, Lagrange's theorem is not valid for a multiplicative loop of a semifield or quasifield. In particular, even the semifields of the minimal order 16 contain the elements of the right and left order 6, which do not divide the order of the loop. In the exceptional non-primitive Knuth–Rúa semifield of order 32, all elements except 0 and 1 have the same right and left order 21.

To identify the patterns of the right and left spectra, we apply the classical concept of a minimal polynomial of a nonzero element to the study of finite semifields. Let Q be a semifield of order p^n, p be prime. The *right-ordered minimal polynomial* of an element $a \in Q$ is said to be a monic polynomial

$$\mu_a^r(x) = x^m + c_1 x^{m-1} + \dots + c_{m-1}x + c_m \in \mathbb{Z}_p[x] \tag{1}$$

of minimal degree such that

$$a^{m)} + c_1 a^{m-1)} + \dots + c_{m-1}a + c_m = 0.$$

The *left-ordered minimal polynomial* $\mu_a^l(x)$ is defined likewise. Some useful properties of one-sided-ordered minimal polynomials see in [5].

The main result of the paper is the following theorem, where «lcm» is a least common multiple of some numbers.

Theorem 1. *Let Q be a non-associative semifield of order p^n (p be prime), the right-ordered minimal polynomial of an element $a \in Q^*$ has the canonical decomposition into irreducible factors:*

$$\mu_a^r(x) = \varphi_1^{s_1}(x)\varphi_2^{s_2}(x)\dots\varphi_s^{s_d}(x) \in \mathbb{Z}_p[x].$$

Then the right order of an element a is a divisor of the number

$$\mathrm{lcm}(p^{m_1}-1, p^{m_2}-1, \dots, p^{m_d}-1, k_1, k_2, \dots, k_d),$$

where m_i is the degree of irreducible polynomial $\varphi_i(x)$, the number k_i equals to 1 if $s_i = 1$, otherwise k_i is the minimal with conditions

$$C_{k_i}^1 \vdots\, p, \quad C_{k_i}^2 \vdots\, p, \ \dots, \ C_{k_i}^{s_i-1} \vdots\, p,$$

for all $i = 1, 2, \dots, d$.

As a corollary, we indicate the important special cases of small-rank semifields: for orders p^3, p^4, p^5. Moreover, we can say that our results are true for left orders and left-ordered minimal polynomials in finite semifields. Also, for right and left quasifield, respectively.

The research method is closely related to linear spaces and spread sets, is based on multiplication recording in a quasifield as a linear transformation in the associated linear space. The matrix operations allow us to effectively apply the method to prove the theoretical result and to illustrate it by the examples of some semifields of orders 2^4, 2^5, 2^6, 3^4, 5^4, 13^4.

2. Spread set and minimal polynomials

It is well-known, that the order of finite semifield or quasifield is the prime number degree p^n [1]. A finite quasifield may be constructed on the basis of a linear space over an appropriate finite field. Let Q be a n-dimensional linear space over the field $\mathbb{Z}_p$, θ is a bijective mapping from Q to $GL_n(p) \cup \{0\}$ such that:

1) $\det(\theta(u) - \theta(v)) \neq 0 \ \forall u, v \in Q, \ u \neq v$,

2) $\theta(0, 0, \ldots, 0) = 0$ is zero matrix, $\theta(1, 0, \ldots, 0) = E$ is identity matrix.

Define the multiplication law on Q by the rule

$$u * v = u\theta(v), \qquad u, v \in Q,$$

then $\langle Q, +, * \rangle$ is a right quasifield of order p^n. The multiplicative neutral element $\theta^{-1}(E)$ is denoted as e. The image

$$R = \{\theta(u) \mid u \in Q\} \subset GL_n(p) \cup \{0\} \tag{2}$$

is called a *spread set*. And inversely, the right multiplication $R_a : x \to x * a$ in a right quasifield Q is a linear transformation of the linear space Q over the prime subfield $\mathbb{Z}_p$. The set of R_a for all $a \in Q$ is the spread set of Q. For more information see [6], the well-known properties is presented by following preposition:

1) Q is a semifield iff its spread set R is closed under addition;

2) Q is a semifield iff R is closed under multiplication;

3) Q is a field iff R is a field.

Evidently, the matrix representation of the spread set depends on the base of Q as a vector space. Another base choice with the transition matrix T leads to the new spread set TRT^{-1}, so different spread sets can define the isomorphic quasifields. Next, we will choose the appropriate matrix representation of a spread set up to the matrices conjugation. As a rule, we will assume the first basic vector $e_1 = e$ and we will construct the base of Q such that the matrix $\theta(a)$ (for the chosen element a) be of more convenient form – Jordan normal form or close to it.

Some properties of one-side-ordered minimal polynomials in a finite semifield Q correspond to similar results in finite fields, see [5]. The right- or the left-ordered minimal polynomial of an element $a \in Q^*$ is not necessarily irreducible, but $\mu_a^r(0) \neq 0$, $\mu_a^l(0) \neq 0$. The right-(left-)ordered minimal polynomial is a factor of the polynomial $x^k - 1$, where $k = |a|_r$ ($k = |a|_l$). The minimal polynomial of a has the degree 1 or 2 iff a belongs to a subfield of order p^2 in Q, see [7].

Let $a \in Q^*$ and $A = \theta(a)$ is the corresponding matrix from the spread set $R \subset GL_n(p) \cup \{0\}$. Then the right-ordered minimal polynomial of an element a is factor of the minimal polynomial of the matrix A. Moreover, the right order of a is a factor of the order of the matrix A in the general linear group $GL_n(p)$ (proved in [8]).

For completeness, we will prove the following simple but useful result.

Lemma 1. *Let Q be a semifield of order p^n with the spread set R (2). If an elelment $a \in Q$ does not belong to the prime subfield $\mathbb{Z}_p$ then the characteristic polynomial of the matrix $A = \theta(a) \in R$ has no linear factors over $\mathbb{Z}_p$.*

Proof. Assume that the statement is false and the polynomial $\det(A - \lambda E)$ has the factor $\lambda - \alpha$, $\alpha \in \mathbb{Z}_p$. Then the linear transformation with the matrix A has an eigenvector $v \in Q^*$ with the eigenvalue α:

$$v\theta(a) = \alpha v \Rightarrow v * a = v * \alpha,$$

it contradicts the definition of a loop Q^*. □

Evidently that the statement is true for any (right) quasifield too, if $\mathbb{Z}_p \subset Z(Q)$.

Remind that for any square mathix A the characteristic matrix $A - \lambda E$ can be transform, by equivalent tranformations, to the normal diagonal form:

$$A - \lambda E \sim \begin{pmatrix} E_1(\lambda) & 0 & \dots & 0 \\ 0 & E_2(\lambda) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & E_n(\lambda) \end{pmatrix},$$

where the non-zero invariant factors $E_i(\lambda) \in \mathbb{Z}_p[\lambda]$ are monic polynomials, and $E_i(\lambda)$ is a divisor of $E_{i+1}(\lambda)$, $1 \leqslant i < n$. Moreover, the characteristic polynomial of A is

$$\det(A - \lambda E) = (-1)^n E_1(\lambda) E_2(\lambda) \dots E_n(\lambda),$$

and the last invariant factor equals to the minimal polynomial of A: $E_n(\lambda) = \mu_A(\lambda)$.

3. Main results

We will prove the main Theorem 1.1 by the sequence of lemmas each of them can be considered as an independent result. These lemmas represent the necessary partial cases, and the theorem proof can be constructed by evident induction.

Consider the right-ordered minimal polynomial $\mu_a^r(x)$ for an element $a \in Q^*$, this polynomial is a divisor of $\mu_A(x)$ for $A = \theta(a)$. It is clear that the right order of a is uniquely defined by the polynomial $\mu_a^r(x)$; $|a|_r$ equals to the length of the neutral element orbit under the linear transformation $\psi = R_a : y \to y * a$. When the degree of the polynomial $\mu_a^r(x)$ is $m < n$, we can consider the map ψ, instead of n-dimensional linear space Q, in the m-dimensional linear sub-space $\mathcal{L}_a \subset Q$ with the base $e, a, a^2, a^{3)}, \dots, a^{m-1)}$.

Lemma 2. *If the right-ordered minimal polynomial $\mu_a^r(x) \in \mathbb{Z}_p[x]$ of an element $a \in Q^*$ is an irreducible polynomial of the degree m then the right order of a is a divisor of the number $p^m - 1$.*

Proof. Consider the right-ordered polynomial $\mu_a^r(x)$ (1) and construct the matrix A of the linear transformation $\psi : y \to y * a$ of the linear space $\mathcal{L}_a$ using the base above:

$$\begin{aligned}
&e^\psi = e * a = a = (0, 1, 0, 0, \dots, 0),\\
&a^\psi = a * a = a^2 = (0, 0, 1, 0, \dots, 0),\\
&(a^2)^\psi = a^2 * a = a^{3)} = (0, 0, 0, 1, \dots, 0),\\
&\dots,\\
&(a^{m-1)})^\psi = a^{m-1)} * a = a^{m)} = -c_m - c_{m-1}a - \dots - c_1 a^{m-1)} = (-c_m, -c_{m-1}, \dots, -c_1);
\end{aligned}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 \\ -c_m & -c_{m-1} & -c_{m-2} & \dots & -c_1 \end{pmatrix}.$$

It is the companion matrix of $\mu_a^r(x)$, and the set

$$F = \mathbb{Z}_p(A) = \{b_0E + b_1A + b_2A^2 + \dots + b_{m-1}A^{m-1} \mid b_i \in \mathbb{Z}_p,\ i = 0, 1, \dots, m-1\}$$

is the field of order p^m, see [9]. So, the orbit length of the element $e \in \mathcal{L}_a$ under ψ equals to the order of the matrix A in the cyclic group F^*, $|a|_r$ is a divisor of $p^m - 1$. □

As can be seen, the lemma proven generalizes the corollary from Lagrange's theorem that the element order is a divisor of the finite group order. For any nonzero element a of an arbitrary finite semifield Q, the result is incorrect, see examples below. The result of the lemma is trivial when a belongs to the simple subfield $\mathbb{Z}_p$: the minimal polynomial is linear and the right (and left) order of the element divides $p-1$. It is clear that the result is also valid for an element from any subfield of a finite semifield Q.

Note that the transition from the semifield $Q = (Q, +, *)$ to the *opposite semifield* $Q^{op} = (Q, +, \circ)$ with the multiplication $x \circ y = y * x$ interchanges the right order and the left order of a, also the right-ordered minimal polynomial and the left-ordered minimal polynomial. Thus, all results proved for the right spectrum can be transferred to the left spectrum.

Lemma 3. *If the right-ordered minimal polynomial of an element $a \in Q^*$ is $\mu_a^r(x) = \varphi^2(x)$, where $\varphi(x)$ is irreducible polynomial of degree m, $n = 2m$, then the right order of a is a divisor of the number $p(p^m - 1)$.*

Proof. Clear that the normal diagonal form of the matrix $\theta(a) - \lambda E$ is $\mathrm{diag}(1, 1, \dots, 1, \varphi^2(\lambda))$. Choose the base of Q such that the matrix $\theta(a)$ be of the form

$$A = \begin{pmatrix} B & E \\ 0 & B \end{pmatrix},$$

where all the blocks are $(m \times m)$-dimensional, so the normal diagonal form of $B - \lambda E$ is $\mathrm{diag}(1, 1, \dots, 1, \varphi(\lambda))$. For instance, we can write the matrix B as the companion matrix of the polynomial $\varphi(x)$ by the manner above. Such the base choice is possible because the matrices A and $\theta(a)$ are conjugated, see the previous section.

Evidently, for any $k \in \mathbb{N}$ we have

$$A^k = \begin{pmatrix} B^k & kB^{k-1} \\ 0 & B^k \end{pmatrix}.$$

The image of the neutral element $e = (1, 0, 0, \dots, 0)$ under the linear transformation $\psi^k : y \to yA^k$ coincides to e iff $k \equiv 0 \pmod{p}$ and $B^k = E$. The second condition follows from the irreducibility of the polynomial $\varphi(x)$, because the set $\mathbb{Z}_p(B)$ of $(m \times m)$-matrices is the field of order p^m. So, the order of matrix A in the group $GL_n(p)$ is a divisor of $p(p^m - 1)$, the lemma is proved. □

Additionally, we note that this reasoning shows the need for the divisibility of $|a|_r$ by the number p. We will not focus on this condition because of the complexity in the general case.

Lemma 4. *If the right-ordered minimal polynomial of an element $a \in Q^*$ is the product of two different irreducible polynomials $\mu_a^r(x) = \varphi_1(x)\varphi_2(x)$ of orders m_1 and m_2, $n = m_1 + m_2$, then the right order of a is a divisor of the least common multiple of numbers $p^{m_1} - 1$ and $p^{m_2} - 1$.*

Proof. The normal diagonal form of the matrix $\theta(a) - \lambda E$ is $\mathrm{diag}(1, \dots, 1, \varphi_1(\lambda)\varphi_2(\lambda))$, so, up to conjugation, the matrix $\theta(a)$ can be chosen as

$$A = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}.$$

Here the block B is $(m_1 \times m_1)$-matrix and $B - \lambda E \sim \mathrm{diag}(1, 1, \dots, 1, \varphi_1(\lambda))$, the block C is $(m_2 \times m_2)$-matrix and $C - \lambda E \sim \mathrm{diag}(1, 1, \dots, 1, \varphi_2(\lambda))$. The order of the matrix A evidently equals to the least common multiple of the orders of B and C in general linear groups $GL_{m_1}(p)$ and $GL_{m_2}(p)$, or, more precisely, in cyclic multiplicative groups of associated fields

$$F_1 = \{f(B) \mid f(x) \in \mathbb{Z}_p[x]\} \simeq GF(p^{m_1}) \quad \text{and} \quad F_2 = \{f(C) \mid f(x) \in \mathbb{Z}_p[x]\} \simeq GF(p^{m_2}).$$

The lemma is proved. □

Remark 1. *It is clear that the case of more than two irreducible factors in the polynomial $\mu_a^r(x)$ decomposition is considered by induction. Moreover, in the case when $m_1 + m_2 < n$, we must replace the linear space Q with its linear subspace $\mathcal{L}_a$.*

It remains to consider the case when the irreducible polynomial $\varphi(x)$ is s-times factor of $\mu_a^r(x)$, $s > 2$. It is easy to show, that in this case, the choice of the base allows us to write the corresponding $(ms \times ms)$-dimensional block in the form:

$$A = \begin{pmatrix} B & E & 0 & 0 & \dots & 0 \\ 0 & B & E & 0 & \dots & 0 \\ 0 & 0 & B & E & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & E \\ 0 & 0 & 0 & 0 & \dots & B \end{pmatrix}.$$

Now we can raise it to the k-th degree using Newton's binomial:

$$A^k = \begin{pmatrix} B^k & C_k^1 B^{k-1} & C_k^2 B^{k-2} & C_k^3 B^{k-3} & \dots & 0 \\ 0 & B^k & C_k^1 B^{k-1} & C_k^2 B^{k-2} & \dots & 0 \\ 0 & 0 & B^k & C_k^1 B^{k-1} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & C_k^1 B^{k-1} \\ 0 & 0 & 0 & 0 & \dots & B^k \end{pmatrix}.$$

The image of the neutral element $e = (1, 0, \dots, 0)$ equals to $eA^k = e$ when two condition hold:

1) the order of the matrix B in $GL_m(p)$ (or in multiplicative group of the associated field $GF(p^m)$) is a divisor of the number k and

2) the characteristic p is a divisor of the binomial coefficients C_k^1, $C_k^2, \dots$, C_k^{s-1}.

These arguments, together with the lemmas and the remark, complete the proof of the Theorem 1.1.

Remark 2. *The result of the theorem remains valid for the right order and right-ordered minimal polynomial in a finite right quasifield, as well as for the left order and left-ordered minimal polynomial in a finite left quasifield (including a semifield).*

The following corollary represents some important cases of small-rank semifield.

Corollary 1. *Let Q be non-associative semifield of order p^n, $a \in Q^*$. The right order and the left order of an element a are divisors of:*

1) $p^3 - 1$, when $n = 3$;

2) $p^4 - 1$ or $p(p^2 - 1)$, when $n = 4$;

3) $p^5 - 1$ or $(p^2 - 1)(p^3 - 1)$, when $n = 5$.

Thus, any three-dimensional finite semifield satisfies to the corollary of Lagrange's theorem. We can not guarantee it for arbitrary four- and five-dimensional semifield. In the case of $n = 6$ the listing of all the variants is too complicated.

4. Examples

1. Illustrate the results by the example of a semifield of order 16. It is known that there exist 23 pairwise non-isomorphic semifields of order 16, see enumeration by E. Kleinfeld and results of P K. Shtukkert and V. M. Levchuk, see [3]. The detailed table in that review contains the information on spectra, subfields and automorphisms. All the semifields of order 16 are *right and left primitive*, that is the multiplicative loop Q^* is the set of left-ordered and right-ordered degrees of some element a. So, the right and left spectra contains the number 15, these spectra are the following (for different semifields): $\{1, 3, 15\}$, $\{1, 3, 6, 15\}$, $\{1, 3, 5, 6, 15\}$, $\{1, 5, 6, 15\}$. The number 6 in the spectra is not the divisor of $|Q^*| = 15$, but from corollary we have $p(p^2 - 1) = 2 \cdot 3 = 6$, in this case we see the right- or left-ordered minimal polynomial $(x^2 + x + 1)^2$.

2. The results on 3-primitive semifield projective planes of order 81 are presented in [10]. There exist exactly 8 non-isomorphic semifield planes of order 81 that admit an involution automorphism which fixes pointwise a subplane of order 9. Corresponding 8 non-isotopic semifields of order 81 have the right and left spectra containing only divisors of $|Q^*| = 80$: $\{1, 2, 4, 8, 16, 40, 80\}$ or $\{1, 2, 4, 8, 16, 80\}$.

Another example of semifields of order 81 is the commutative Cohen–Ganley semifield [11]

$$Q = \{(x, y) \mid x, y \in F \simeq GF(9)\}$$

with the multiplication

$$(x, y) \circ (u, v) = (xv + yu + x^3u^3, yv + \eta xu + \eta^{-1}xu), \qquad x, y, u, v \in F,$$

η is non-square in F. The spread set of this semifield considered as 4-dimensional linear space over $\mathbb{Z}_3$ consists of matrices

$$\theta(x_1, x_2, x_3, x_4) = x_1E + x_2 \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 1 & 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 2 & 2 \\ 1 & 2 & 0 & 1 \end{pmatrix} + x_4 \begin{pmatrix} 0 & 0 & 0 & 1 \\ 2 & 0 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

$x_1, x_2, x_3, x_4 \in \mathbb{Z}_3$. The next two tables present the elements $a \in Q^*$, their minimal polynomials $\mu_a^r(x) = \mu_a^l(x)$ and their right (left) orders $|a|_r = |a|_l$ calculated by the second author.

Note that the elements with minimal polynomials of degree 1 and 2

$$\{(1, 0, 0, 0), (2, 0, 0, 0), (0, 1, 2, 0), (0, 2, 1, 0), (1, 1, 2, 0), (1, 2, 1, 0), (2, 1, 2, 0), (2, 2, 1, 0)\},$$

together with zero vector form the subfield of order 9. It is so-called *middle nucleus* of Q:

$$N_m = \{b \in Q \mid (a * b) * c = a * (b * c) \ \ \forall b, c \in Q\}.$$

Table 1. Right order is a divisor of $p^4 - 1 = 80$

Element $a \in Q^*$	$\mu_a^r(x)$	$\lvert a\rvert_r$
(1, 0, 0, 0)	$x-1$	1
(2, 0, 0, 0)	$x-2$	2
(0, 1, 2, 0), (0, 2, 1, 0)	x^2+1	4
(2, 0, 1, 0), (2, 1, 0, 2), (2, 1, 1, 0), (2, 1, 1, 1)	$x^4+x^3+x^2+x+1$	5
(1, 1, 2, 0), (1, 2, 1, 0)	x^2+x+2	8
(2, 1, 2, 0), (2, 2, 1, 0)	x^2+2x+2	8
(1, 0, 2, 0), (1, 2, 0, 1), (1, 2, 2, 0), (1, 2, 2, 2)	$x^4+2x^3+x^2+2x+1$	10
(0, 0, 0, 1), (0, 0, 0, 2), (0, 1, 2, 2), (0, 2, 1, 1)	x^4+x^2+2	16
(0, 0, 1, 0), (0, 1, 0, 2), (0, 1, 1, 0), (0, 1, 1, 1)	x^4+x^2+x+1	40
(0, 0, 2, 0), (0, 2, 0, 1), (0, 2, 2, 0), (0, 2, 2, 2)	x^4+x^2+2x+1	40
(1, 0, 0, 1), (1, 0, 0, 2), (1, 1, 2, 2), (1, 2, 1, 1)	$x^4+2x^3+x^2+1$	40
(2, 0, 0, 1), (2, 0, 0, 2), (2, 1, 2, 2), (2, 2, 1, 1)	$x^4+x^3+x^2+1$	40
(1, 0, 1, 0), (1, 1, 0, 2), (1, 1, 1, 0), (1, 1, 1, 1)	$x^4+2x^3+x^2+x+2$	80
(2, 0, 2, 0), (2, 2, 0, 1), (2, 2, 2, 0), (2, 2, 2, 2)	$x^4+x^3+x^2+2x+2$	80

Table 2. Right order is a divisor of $p(p^2 - 1) = 24$

Element $a \in Q^*$	$\mu_a^r(x)$	$\lvert a\rvert_r$
(0, 1, 1, 2), (0, 2, 2, 1)	$(x^2+2x+2)(x^2+x+2)$	8
(1, 1, 1, 2), (1, 2, 2, 1)	$(x^2+1)(x^2+2x+2)$	8
(2, 1, 1, 2), (2, 2, 2 ,1)	$(x^2+1)(x^2+x+2)$	8
(0, 0, 1, 1), (0, 0, 1, 2), (0, 0, 2, 1), (0, 0, 2, 2), (0, 1, 0, 0), (0, 1, 0, 1), (0, 1, 2, 1), (0, 2, 0, 0), (0, 2, 0, 2), (0, 2, 1, 2)	$(x^2+1)^2$	12
(1, 0, 1, 1), (1, 0, 1, 2), (1, 0, 2, 1), (1, 0, 2, 2), (1, 1, 0, 0), (1, 1, 0, 1), (1, 1, 2, 1), (1, 2, 0, 0), (1, 2, 0, 2), (1, 2, 1, 2)	$(x^2+x+2)^2$	24
(2, 0, 1, 1), (2, 0, 1, 2), (2, 0, 2, 1), (2, 0, 2, 2), (2, 1, 0, 0), (2, 1, 0, 1), (2, 1, 2, 1), (2, 2, 0, 0), (2, 2, 0, 2), (2, 2, 1, 2)	$(x^2+2x+2)^2$	24

The feature of this example is the number of «right roots» of the polynomials. This number equals m for irreducible polynomials of degree m (see Tab. 1), and it does not equal m for reducible ones (see Tab. 2).

Question. *How many «right roots» and «left roots» does a polynomial $f(x) \in \mathbb{Z}_p[x]$ have in a semifield Q of order p^n, if deg$(f) = m$?*

3. The results of the first author on the semifield planes of order p^4 with the special automorphisms subgroup $H \simeq Q_8$ in [12] were illustrated by the examples of semifield planes and semifields of order 5^4 and 13^4. It was proved that all the coordinatizing semifields are both left and right primitive, non-commutative. Each of them have 1, 2 or $p+2$ maximal subfields of order p^2, the automorphism group is $\mathbb{Z}_2$ or $\mathbb{Z}_{p+1}$.

Let M_n be the set of all divisors of integer n. According to the corollary, the right spectrum of semifields of order 625 above is contained in

$$M_{5^4-1} \cup \{15, 30, 40, 60, 120\} \subset M_{5^4-1} \cup M_{5\cdot(5^2-1)},$$

for the semifields of order 13^4 the right spectrum is the subset of

$$M_{13^4-1} \cup \{21, 91, 104, 182, 273, 312, 364, 546, 728, 1092, 2184\} \subset M_{13^4-1} \cup M_{13\cdot(13^2-1)}.$$

4. Consider two exceptional non-primitive semifields, for more information see [3]. In 1991 G.P. Wene wrote the hypothesis: *any finite semifield is right or left primitive*. In 2004 I.F. Rúa gave the counter-example to Wene's conjecture, using a Knuth semifield $\mathcal{R}$ of order 32. This commutative *Knuth-Rúa semifield* is neither right nor left primitive. The second counter-example is *Hentzel-Rúa semifield* $\mathcal{H}$ of order 64, which was constructed in 2007. These semifields have no elements of one-sided order 31 and 63 respectively. Another counter-examples are still unknown.

Note that even non-primitive Knuth–Rúa and Hentzel–Rúa semifields are *right-cyclic*, these semifields admit a $\mathbb{Z}_p$-base

$$\{e,\ a,\ a^{2)},\ldots,a^{n-1)}\},$$

for some element a.

It is known that any element $a\in\mathcal{R}\setminus\{0,1\}$ has the right (and left) order 21. The direct calculation presented in [5] shows that the right-ordered minimal polynomial $\mu_a^r(x)$ is

$$x^5+x^4+1=(x^2+x+1)(x^3+x+1) \text{ or } x^5+x+1=(x^2+x+1)(x^3+x^2+1).$$

So, by the corollary, we obtain $(p^2-1)(p^3-1)=21$, which is consistent with earlier results [3].

Now consider the Hentzel–Rúa semifield $\mathcal{H}$ of order 64, using the information from [5]. Note that the right-ordered minimal polynomial of $a\in\mathcal{H}$ is not necessarily equal to the minimal polynomial of the associated matrix $A=\theta(a)$.

The most interesting situation we see when the right-ordered minimal polynomial of a is $(x^2+x+1)^3$. According the main theorem 1.1 for $m_1=2$ and $s_1=3$, the right order of a must be a divisor of the number $\mathrm{lcm}(2^2-1,k_1)$, where k_1 is the minimal with the conditions $C_{k_1}^1\,\vdots\,2$, $C_{k_1}^2\,\vdots\,2$. From Pascal's triangle

$$\begin{array}{c} 1\\ 1\ 1\\ 1\ 2\ 1\\ 1\ 3\ 3\ 1\\ 1\ \underline{4\ 6}\ 4\ 1 \end{array}$$

we see that $k_1=4$, $\mathrm{lcm}(2^2-1,4)=12=|a|_r$. One can check the rest of the cases in the Tab. 3.

Table 3. Orders and minimal polynomials in $\mathcal{H}$

$\lvert a\rvert_l=\lvert a\rvert_r$	$m_a^l(x)=m_a^r(x)$	$m_A(x)$
7	$(x^3+x+1)(x^3+x^2+1)$	$(x^3+x+1)(x^3+x^2+1)$
12	$(x^2+x+1)^3$	$(x^2+x+1)^3$
15	x^4+x+1	$(x^4+x+1)(x^2+x+1)$
6	$(x^2+x+1)^2$	$(x^2+x+1)^3$
7	x^3+x+1 or x^3+x^2+1	$(x^3+x+1)^2$ or $(x^3+x^2+1)^2$
3	x^2+x+1	x^2+x+1

This work is supported by the Krasnoyarsk Mathematical Center and financed by the Ministry of Science and Higher Education of the Russian Federation (Agreement no. 075-02-2024-1429).

References

[1] D.R.Hughes, F.C.Piper, Projective planes, Springer–Verlag New–York Inc., 1973.

[2] N.L.Johnson, V.Jha, M.Biliotti, Handbook of finite translation planes, Pure and applied mathematics. Chapman&Hall/CPC, 2007.

[3] V.M.Levchuk, O.V.Kravtsova, Problems on structure of finite quasifields and projective translation planes, *Lobachevskii Journal of Mathematics*, **38**(2017), no. 4, 688–698. DOI: 10.1134/S1995080217040138

[4] A.N.Grishkov, A.V.Zavarnitsyn, Lagrange's theorem for Moufang loops, *Math. Proc. Phil. Soc.*, **139**(2005), 41–57.

[5] O.V.Kravtsova, Minimal polynomials in finite semifields, *Journal of Siberian Federal University. Mathematics & Phisics*, **11**(2018), no. 5, 588–596. DOI: 10.17516/1997-1397-2018-11-5-588-596

[6] O.V.Kravtsova, D.S.Skok, The spread set method for the construction of finite quasifields, *Trudy Inst. Mat. i Mekh. UrO RAN*, **28**(2022), no. 1, 164–181. DOI: 10.21538/0134-4889-2022-28-1-164-181

[7] O.V.Kravtsova, Minimal proper quasifields with additional conditions, *Journal of Siberian Federal University. Mathematics & Physics*, **13**(2020), no. 1, 104–113. DOI: 10.17516/1997-1397-2020-13-1-104-113

[8] O.V.Kravtsova, V.S.Loginova, Questions of the structure of finite Hall quasifields, *Trudy Inst. Mat. i Mekh. UrO RAN*, **30**(2024), no. 1, 128–141. DOI: 10.21538/0134-4889-2024-30-1-128-141

[9] R.Lidl, G.Pilz, Applied Abstract Algebra, Springer–Verlag New York, 1984.

[10] O.V.Kravtsova, I.V.Sheveleva, On some 3-primitive projective planes, *Chebyshevskii Sb.*, **20**(2019), no. 3, 316–332. DOI: 10.22405/2226-8383-2018-20-3-316-332

[11] S.D.Cohen, M.J.Ganley, Commutative semifields, two dimensional over their middle nuclei, *Journal of Algebra*, **75**(1982), Is. 2, 373–385.

[12] O.V.Kravtsova, Semifield planes admitting the quaternion group Q_8, *Algebra and Logic*, **59**(2020), no. 1, 71–81. DOI: 10.1007/s10469-020-09583-y

О спектрах и минимальных многочленах в конечных полуполях

Ольга В. Кравцова
Илья К. Кузьмин
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. Для исследования конечных полуполей применяется понятие одностороннеупорядоченного минимального многочлена. Отсутствие ассоциативности умножения в собственном полуполе приводит к аномальным свойствам его левого и правого спектра. Получено достаточное условие делимости порядка мультипликативной лупы на правый (левый) порядок элемента. С использованием регулярного множества полуполя описана связь минимального многочлена ненулевого элемента и его правого (левого) порядка. Эта взаимосвязь дает исчерпывающее объяснение наиболее интересным аномальным примерам полуполей малых порядков.

Ключевые слова: полуполе, правый порядок, правый спектр, правоупорядоченный минимальный многочлен, регулярное множество.

EDN: FVEGFC
УДК 517.55

On the Aris-Amundson model

Alexander M. Kytmanov*
Olga V. Khodos†
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 09.09.2024, received in revised form 05.10.2024, accepted 0.11.2024

Abstract. The work is devoted to the study of the real roots of the system of transcendental Aris–Amundson equations. It is shown that the number of real roots is related to the number of real roots of some entire function (resultant). The number of complex roots is investigated.

Keywords: systems of transcendental equations, resultant, simple root.

Citation: A.M. Kytmanov, O.V. Khodos, On the Aris-Amundson model, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 51–58. EDN: FVEGFC.

Introduction

Finding the number of real roots of polynomials is a classical algebraic problem. The Hermite method of quadratic forms, the Sturm method, the Descartes sign rule, and the Byudan–Fourier theorem are devoted to this problem (see, for example, [1]). Further development of these methods for polynomials can be found in the work [2] and the monograph [3]. For entire functions, the question of localization of real positive roots was considered in the classical works of N. G. Chebotarev [4] (pp. 28–56), as well as in the work of [5] (we refer to the collected works of N. G. Chebotarev, since his original works are hardly accessible).

For systems of equations, the number of real roots was studied in the articles [6–8]. In the article [9], the number of real roots was related to the number of real roots of the resultant.

The monographs [10,11] consider algebraic and transcendental systems of equations. Systems of transcendental equations arise, for example, in the study of equations of chemical kinetics [12]. One of the problems that arise there is the problem of the number of real positive roots of a system of equations in a reaction polyhedron. As an example, the Aris-Amundson system has been studied.

1. Multiple roots of the resultant

Let us consider one of the models of a continuous perfectly stirred reactor, the so-called Aris-Amundson model in the dimensionless form (see [12, ch. 2])

$$\frac{dx}{d\tau} = f(y)(1-x) - x = f_1(x,y), \quad \frac{dy}{d\tau} = \beta f(y)(1-x) - s(y-1) = f_2(x,y), \tag{1}$$

where $f(y) = Dae^{\gamma(1-1/y)}$. All constants are positive.

The stationary states of the system (1) are solutions of the stationarity system

$$f_1(x,y) = 0, \quad f_2(x,y) = 0, \tag{2}$$

*akytmanov@sfu-kras.ru https://orcid.org/0000-0002-7394-1480
†khodos_olga@mail.ru

which can be written as

$$Dae^{\gamma(1-1/y)}(1-x)-x=0,$$

$$\beta Dae^{\gamma(1-1/y)}(1-x)-s(y-1)=0.$$

Denoting $Da=b$, $t=\gamma\,(1-1/y)$, we get the system

$$be^t(1-x)-x=0, \quad \beta be^t(1-x)-s\frac{t}{\gamma-t}=0. \tag{3}$$

Obviously, the system (3) has no roots with zero coordinates.

Earlier in the work [13], the Zeldovich–Semenov model was studied in a similar way. The main idea of the study is the application of the multidimensional theory of residues, the study of power sums of roots and residue integrals (see [10, 11]).

Then we get

$$be^t\cdot\frac{\beta(\gamma-t)-st}{\beta(1-t)}-\frac{st}{\beta(\gamma-t)}=0.$$

Thus, the entire function of the first order of growth can serve as the resultant of the system (2)

$$F(t)=be^t(\beta\gamma-t(\beta+s))-st=0.$$

Let us check it for multiple zeros. We convert it to the form

$$\varphi(t)=be^t-\frac{st}{\beta\gamma-t(\beta+s)}.$$

Calculating the derivative, we get

$$\varphi'(t)=be^t-\frac{s\beta\gamma}{(\beta\gamma-t(\beta+s))^2}.$$

Obviously, if $F(t)=0$ and $F'(t)=0$ at some point t, then $\varphi(t)=0$ and $\varphi'(t)=0$ at this point. The converse is also true.

Then from the equalities $\varphi(t)=0$, $\varphi'(t)=0$ we get

$$t_{1,2}=\frac{\beta\gamma\mp\sqrt{\beta^2\gamma^2-4\beta^2\gamma-4\beta\gamma s}}{2(\beta+s)}.$$

Substituting these values, for example, into the first equation, we get

$$b\cdot\exp\left(\frac{\beta\gamma\mp\sqrt{\beta^2\gamma^2-4\beta^2\gamma-4\beta\gamma s}}{2(\beta+s)}\right)=\frac{s}{\beta+s}\cdot\frac{\beta\gamma\mp\sqrt{\beta^2\gamma^2-4\beta^2\gamma-4\beta\gamma s}}{\beta\gamma\pm\sqrt{\beta^2\gamma^2-4\beta^2\gamma-4\beta\gamma s}}. \tag{4}$$

Thus, in equality (4), there is an exponential function on the left, and a power function on the right. Therefore, they cannot match for almost all parameter values. Then for almost all parameter values there are no multiple roots of the function $\varphi(t)$ (and therefore $F(t)$).

Proposition 1. *For almost all parameter values the function $\varphi(t)$ (and therefore $F(t)$) has no multiple roots.*

2. The number of real roots of the resultant

Next, we use the following statement (see [14]).

Theorem 1. *If the system (2) with real coefficients is such that it has no roots with zero coordinates and all zeros of the resultant $F(t)$ are simple, then the number of real roots of the system (2) coincides with the number of real roots of the function $F(t)$.*

From the system (2) we get $1 - x = 1 + \dfrac{s}{\beta}(1-y)$, $x = \dfrac{s}{\beta}(y-1)$. We substitute it into the first equation

$$be^{\gamma(1-1/y)} \cdot \left(1 + \frac{s}{\beta}(1-y)\right) + \frac{s}{\beta}(1-y) = 0.$$

The resultant looks like

$$\varphi(y) = be^{\gamma(1-1/y)} \cdot \frac{\beta - s(y-1)}{s(y-1)} - 1,$$

and we find the number of roots of $\varphi(y)$.

First, we find the intervals of increase and decrease of $\varphi(y)$.

$$\varphi'(y) = \frac{be^{\gamma(1-1/y)}}{s} \cdot \frac{-(\gamma s + \beta)y^2 + \gamma(\beta + 2s)y - \gamma(\beta + s)}{y^2(y-1)^2}.$$

The derivative $\varphi'(y) = 0$ if and only if

$$-(\gamma s + \beta)y^2 + \gamma(\beta + 2s)y - \gamma(\beta + s) = 0.$$

Solving the resulting quadratic equation, we find the discriminant

$$D = \gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s.$$

Solutions to the quadratic equation are

$$y_{1,2} = \frac{\gamma(\beta + 2s) \mp \sqrt{\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s}}{2(s\gamma + \beta)}.$$

If $D > 0$, that is, $\gamma\beta - 4(\beta + s) > 0$, then

$$\psi(y) = -(\gamma s + \beta)y^2 + \gamma(\beta + 2s)y - \gamma(\beta + s)$$

has two real roots $y_1 < y_2$.

Since the graph of the function $\psi(y)$ is a parabola with branches down, then $\psi(y) < 0$ on the interval $(-\infty; y_1) \cup (y_2; \infty)$ and $\psi(y) > 0$ in the interval $(y_1; y_2)$.

If $D = 0$, that is, $\gamma\beta - 4(\beta + s) = 0$, then $\psi(y)$ has one real root y_0 and $\psi(y) < 0$ on the interval $(-\infty; y_0) \cup (y_0; \infty)$.

If $D < 0$, that is, $\gamma\beta - 4(\beta + s) < 0$, then $\psi(y)$ has no real roots and $\psi(y) < 0$ on the entire real line.

Let us show that if $D \geqslant 0$, then the roots of $\psi(y)$ lie to the right of 1, that is, $1 < y_1 \leqslant y_2$. Note beforehand that if $D \geqslant 0$, then $\gamma > 4$.

Indeed, $D \geqslant 0$ is equivalent to the inequality $\gamma \geqslant 4 + \dfrac{4s}{\beta}$, which implies that $\gamma > 4$ $(\beta, s > 0)$.

Assume that $y_1 = \dfrac{\gamma(\beta + 2s) - \sqrt{D}}{2(s\gamma + \beta)} > 1$. This inequality is equivalent to $\beta(\gamma - 2) > \sqrt{D}$. Since $\gamma > 4$, the left and right sides of the last inequality are non-negative, which means it is equivalent to $\beta^2(\gamma-2)^2 > \gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s$. Simplifying it, we get the equivalent condition $4\beta(\beta+\gamma s) > 0$, which is always true, since $\beta, \gamma, s > 0$. Thus, our assumption that $y_1 > 1$ is correct. That is, for $D \geqslant 0$, the condition $1 < y_1 \leqslant y_2$ is satisfied.

It follows from the above that if $D > 0$, that is, $\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s > 0$, then $\varphi'(y)$ has two real roots $1 < y_1 < y_2$ and $\varphi'(y) < 0$ on the set $(-\infty;0)\cup(0;1)\cup(1;y_1)\cup(y_2;+\infty)$, $\varphi'(y) > 0$ in the interval $(y_1;y_2)$. So $\varphi(y)$ decreases on the set $(-\infty;0)\cup(0;1)\cup(1;y_1)\cup(y_2;+\infty)$ and $\varphi'(y)$ increases in the interval $(y_1;y_2)$.

It also follows from the above that if $D > 0$, that is, $\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s > 0$, then $\varphi'(y)$ has two real roots $1 < y_1 < y_2$ and $\varphi'(y) < 0$ on the set $(-\infty;0)\cup(0;1)\cup(1;y_1)\cup(y_2;+\infty)$, $\varphi'(y) > 0$ in the interval $(y_1;y_2)$. So $\varphi(y)$ decreases on the set $(-\infty;0)\cup(0;1)\cup(1;y_1)\cup(y_2;+\infty)$ and $\varphi'(y)$ increases in the interval $(y_1;y_2)$.

If $D = \gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s = 0$, then $\varphi'(y)$ has one real root $y_0 > 1$ and $\varphi'(y) < 0$ on the set $(-\infty;0)\cup(0;1)\cup(1;y_0)\cup(y_0;+\infty)$. So $\varphi(y)$ decreases on the set $(-\infty;0)\cup(0;1)\cup(1;y_0)\cup(y_0;+\infty)$. If $D = \gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s = 0$, then $\varphi'(y)$ has one real root $y_0 > 1$ and $\varphi'(y) < 0$ on the set $(-\infty;0)\cup(0;1)\cup(1;y_0)\cup(y_0;+\infty)$. So $\varphi(y)$ decreases on the set $(-\infty;0)\cup(0;1)\cup(1;y_0)\cup(y_0;+\infty)$.

If $D = \gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s < 0$, then $\varphi'(y)$ has no real roots and $\varphi'(y) < 0$ over the entire domain of $\varphi'(y)$, which means $\varphi(y)$ decreases over the entire domain of definition of $\varphi(y)$.

For a more accurate understanding of the behavior of the function $\varphi(y)$, we find the limits of $\varphi(y)$ at $\pm\infty$ and at the break points: $\lim\limits_{y\to-\infty}\varphi(y) = -be^{\gamma} - 1 < 0$, $\lim\limits_{y\to 0-0}\varphi(y) = -\infty$, $\lim\limits_{y\to 0+0}\varphi(y) = -1$, $\lim\limits_{y\to 1-0}\varphi(y) = -\infty$, $\lim\limits_{y\to 1+0}\varphi(y) = +\infty$, $\lim\limits_{y\to+\infty}\varphi(y) = -be^{\gamma} - 1 < 0$

Now we find the number of roots of the function $\varphi(y)$.

1. If
$$\begin{cases} D > 0, \\ \varphi(y_1) < 0, \\ \varphi(y_2) > 0, \end{cases}$$
or more precisely
$$\begin{cases} \gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s > 0, \\ be^{\frac{\gamma\beta - \sqrt{\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s}}{2(\beta+s)}} \cdot \dfrac{\gamma\beta + \sqrt{\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s} - 2(\beta+s)}{2s} - 1 < 0, \\ be^{\frac{\gamma\beta + \sqrt{\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s}}{2(\beta+s)}} \cdot \dfrac{\gamma\beta - \sqrt{\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s} - 2(\beta+s)}{2s} - 1 > 0, \end{cases}$$
then $\varphi(y)$ has three real roots $1 < Y_1 < Y_2 < Y_3$.

For example, if $b = 0.04$, $\gamma = 10$, $\beta = 1$, $s = 1$, we get the discriminant $D = 20 > 0$, $y_1 = \dfrac{15-\sqrt{5}}{11} \approx 1.16035745659093 > 1$, $y_2 = \dfrac{15+\sqrt{5}}{11} \approx 1.5669152706818 > y_1$, $\varphi(y_1) \approx -0.16584745271763 < 0$, $\varphi(y_2) \approx 0.13869366143044 > 0$ and the function $\varphi(y)$ has three real roots $Y_1 \approx 1.073488201 > 1$, $Y_2 \approx 1.356686984 > Y_1$, $Y_3 \approx 1.733497054 > Y_2$ (see Fig. 1).

Another example: for $b = 0.001$, $\gamma = 10$, $\beta = 10$, $s = 1$ we get the discriminant $D = 5600 > 0$, $y_1 = =3 - \dfrac{\sqrt{14}}{2} \approx 1.129171306 > 1$, $y_2 = 3 + \dfrac{\sqrt{14}}{2} \approx 4.870828694 > y_1$, $\varphi(y_1) \approx -0.76011792742972 < 0$, $\varphi(y_2) \approx 3.4763004785113 > 0$ and the function $\varphi(y)$ has three real roots $Y_1 \approx 1.011153756 > 1$, $Y_2 \approx 1.812214562 > Y_1$, $Y_3 \approx 9.890609328 > Y_2$ (see Fig. 2).

2. When the following conditions are met
$$\begin{cases} \gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s > 0, \\ be^{\frac{\gamma\beta - \sqrt{\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s}}{2(\beta+s)}} \cdot \dfrac{\gamma\beta + \sqrt{\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s} - 2(\beta+s)}{2s} - 1 = 0 \end{cases}$$

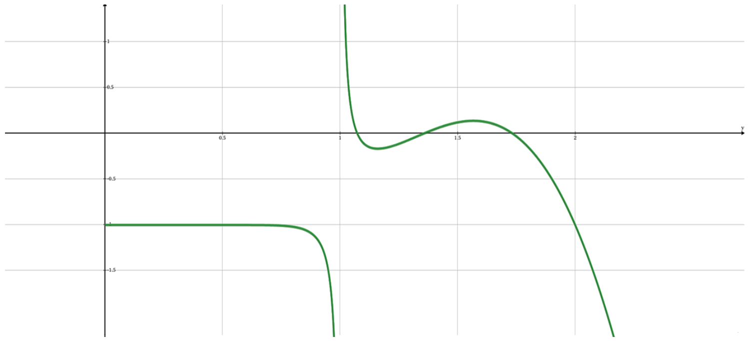


Fig. 1. $\varphi(y)$ has 3 real roots

Fig. 2. $\varphi(y)$ has 3 real roots

(the case when y_1 is a multiple real root of $\varphi(y)$) or

$$\begin{cases} \gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s > 0, \\ be^{\frac{\gamma\beta+\sqrt{\gamma^2\beta^2-4\gamma\beta^2-4\gamma\beta s}}{2(\beta+s)}} \cdot \dfrac{\gamma\beta - \sqrt{\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s} - 2(\beta+s)}{2s} - 1 = 0 \end{cases}$$

(the case when y_2 is a multiple real root of $\varphi(y)$), the function $\varphi(y)$ has two real roots $1 < Y_1 < Y_2$.

An example when Y_1 is a multiple real root ($Y_1 = y_1$) is the following: for $\gamma = 10$, $\beta = 10$, $s = 1$, $b = = \dfrac{(4-\sqrt{14}) \cdot e^{\frac{40-10\sqrt{14}}{-6+\sqrt{14}}}}{16+\sqrt{14}}$ we get the discriminant $D = 5600 > 0$, $y_1 = 3 - \dfrac{\sqrt{14}}{2} \approx 1.129171307 > 1$, $y_2 = 3 + \dfrac{\sqrt{14}}{2} \approx 4.870828694 > y_1$, $\varphi(y_1) = 0$, $\varphi(y_2) \approx 17.6604210584 > 0$ and the function $\varphi(y)$ has two real roots $Y_1 = y_1 = 3 - \dfrac{\sqrt{14}}{2} \approx 1.129171307 > 1$, $Y_2 \approx 10.73089616 > Y_1$ (see Fig. 3).

3. If

$$\gamma^2\beta^2 - 4\gamma\beta^2 - 4\gamma\beta s < 0,$$

the function $\varphi(y)$ has one real root $Y_1 > 1$.

Proposition 2. *The resultant has no more than 3 real roots, therefore, the system (2) has, according to Theorem 1, no more than 3 real roots.*

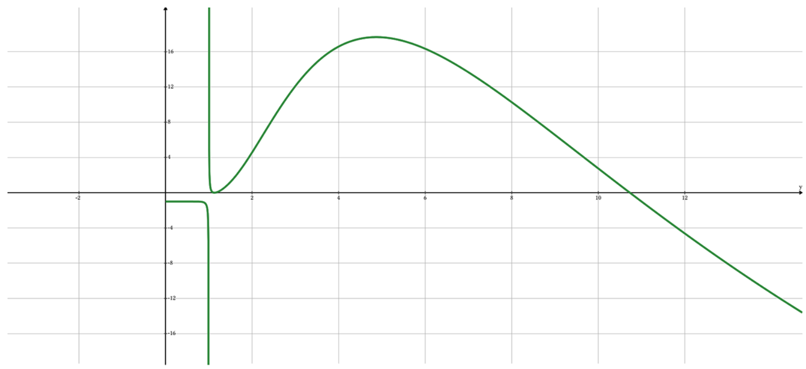


Fig. 3. $\varphi(y)$ has 2 real roots

3. Complex roots of the system

Recall Hadamard's theorem for entire functions of finite order of growth (see, for example, [15]). Expressions $E(u,0) = 1-u$, $E(u,p) = (1-u)e^{u+\frac{u^2}{2}+\cdot+\frac{u^p}{p}}$, $p = 1,2,\dots$ are called *approximate multipliers.*

If a function $f(t)$ on the complex plane has a finite order of growth ρ and $t_1,\dots,t_n,\dots$ its zeros, then there exists an integer $p \leqslant \rho$ independent of n such that the product

$$\prod_{n=1}^{\infty} E\left(\frac{t}{t_n}, p\right) \tag{5}$$

converges for all t if the series converges

$$\sum \left(\frac{r}{r_n}\right)^{p+1},$$

where $r_1, r_2, \dots$ are the absolute values of the zeros of the function $f(t)$, and this series converges for all values of r if $p+1 \geqslant \rho$.

The product (5) with the smallest of the integers p for which the series converges is called the *canonical product* constructed from zeros $f(t)$, and this smallest p is called its *genus.*

Theorem 2 (Hadamard)**.** *If the an entire function $f(t)$ of order ρ has zeros t_1, t_2, ..., and $f(0) \neq 0$, then*

$$f(t) = e^{Q(t)}P(t),$$

where $P(t)$ is the canonical product constructed from zeros $f(t)$, and $Q(t)$ is a polynomial of degree no higher than ρ.

Consider the resultant

$$F(t) = be^t(\beta\gamma - t(\beta+s)) - st.$$

This is a entire function of the first order of growth.

If it has a finite number of zeros, then according to Hadamard's theorem it will have the form

$$F(t) = e^t \cdot P_m(t),$$

where $P_m(t)$ is a certain polynomial. From here

$$e^t = \frac{st}{b(\beta\gamma - t(\beta+s))} \cdot P_m(t),$$

which is impossible, since there is a transcendental function on the left, and a rational one on the right.

Thus, the resultant $F(t)$ has an infinite number of complex roots t_k, $|t_k| \to +\infty$ as $k \to \infty$.

From the system (3), we express x and y in terms of t and get

$$y = \frac{\gamma}{\gamma - t}, \quad x = \frac{st}{\beta(\gamma - t)}.$$

Then at the points t_k we have $x_k = \dfrac{st_k}{\beta(\gamma - t_k)}$, $y_k = \dfrac{\gamma}{\gamma - t_k}$. Therefore, $x_k \to -\dfrac{s}{\beta}$, $y_k \to 0$ as $k \to \infty$.

The work was supported by the Russian Science Foundation, project no. 24-21-00023.

References

[1] F.R.Gantmakher, Theory of Matrices, New York, Chelsea Pub. Co., 1959.

[2] M.G.Krein, M.A.Naimark, The Method of Symmetric and Hermitian Forms in the Theory of the Separation of the Roots of Algebraic Equation, *Linear Multilin. Algebra*, **10**(1981), no. 4, 265–308.

[3] E.I.Jury, Inners and stability of dynamic systems, New York-London-Sydney-Toronto, Wiley, 1974.

[4] N.G.Chebotarev, Work Collection, Moscow-Leningrad, AN SSSR, Vol. 2, 1949 (in Russian).

[5] A.M.Kytmanov, O.V.Khodos, On localization of the zeros of an entire function of finite order of growth, *Journal Complex Analysis and Operator Theory*, **11**(2017), 393–416. DOI:10.1007/s11785-016-0606-8

[6] L.A.Ajzenberg, V.A.Bolotov, A.K.Tsikh, On the solution of systems of nonlinear algebraic equations using the multidimensional logarithmic residue. On the solvability in radicals, *Sov. Math., Dokl.*, **21**(1980), 645–648.

[7] N.N.Tarkhanov, Calculation of the Poincare index, Izv. Vyssh. Uchebn. Zaved. Mat., (1984), no. 9, 47–50 (in Russian).

[8] A.M.Kytmanov, On the number of real roots of systems of equations, *Soviet Math. (Iz. VUZ)*, **35**(1991), no. 6, 19–22.

[9] A.M.Kytmanov, O.V.Khodos, On the roots of systems of transcendental equations, *Probl. Anal.*, **13**(2024), no. 1, 37–49. DOI:10.15393/j3.art.2024.14430

[10] V.Bykov, A.Kytmanov, M.Lazman, M.Passare (ed), Elimination Methods in Polynomial Computer Algebra, Springer science+business media, Dordreht, 1998.

[11] A.M.Kytmanov, Algebraic and trascendental systems of equations and transcendental systems of equations, Krasnoyarsk, Siberian Federal University, 2019 (in Russian).

[12] V.I.Bykov, S.B.Tsybenova, Nonlinear models of chemical kinetics, Moscow, KRASAND, 2011 (in Russian).

[13] O.V.Khodos, On Some System of Non-algebraic Equation in $\mathbb{C}^n$, *J. Sib. Fed. Univ. Math. Phys.*, **7**(2014), no. 4, 455–465.

[14] A.M.Kytmanov, O.V.Khodos, On the Real Roots of Systems of Transcendental Equations, *J. Sib. Fed. Univ. Math. Phys.*, **17**(2024), no, 3, 328–335. EDN: HKGXLG

[15] A.I.Markushevich, Theory of Functions of a Complex Variable, Vol. 2, Prentice-Hall, 1965.

О модели Ариса-Амундсона

Александр М. Кытманов
Ольга В. Ходос
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. Работа посвящена исследованию вещественных корней системы трансцендентных уравнений Ариса–Амундсона. Показано, что число вещественных корней связано с числом вещественных корней некоторой целой функции (результанта). Исследовано число комплексных корней.

Ключевые слова: системы трансцендентных уравнений, результант, простой корень.

EDN: CUZOVF
УДК 512.577+519.68:007.5

On some Commutative and Idempotent Finite Groupoids Associated with Subnets of Multilayer Feedforward Neural Networks

Andrey V. Litavrin*
Tatyana V. Moiseenkova†
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 05.07.2024, received in revised form 23.08.2024, accepted 06.10.2024

Abstract. The work studies commutative and idempotent finite groupoids that are associated with subnetworks of multilayer feedforward neural networks (hereinafter simply neural networks). Previously, the concept of a neural network subnet was introduced. This paper introduces the concept of a generalized subnetwork of a neural network. This concept generalizes the previously introduced concept. The resulting groupoids are called additive and multiplicative groupoids of generalized subnets of a given neural network. These groupoids model the union and intersection of generalized subnets of a neural network. The conditions that the neural network architecture must satisfy in order for the additive groupoid of generalized subnets to be associative are identified. The conditions that the neural network architecture must satisfy in order for the multiplicative groupoid of generalized subnets to be associative are obtained. Subgroupoids of the constructed groupoids are studied.

Keywords: groupoid, multilayer neural network of feedforward signal propagation, subnetwork of multilayer neural network of feedforward signal propagation, additive groupoid of generalized subnetworks, multiplicative groupoid of generalized subnetworks, generalized subnetwork.

Citation: A.V. Litavrin, T.V. Moiseenkova, On some Commutative and Idempotent Finite Groupoids Associated with Subnets of Multilayer Feedforward Neural Networks, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 59–70. EDN: CUZOVF.

Introduction

In this work, only multilayer feedforward neural networks are considered (therefore, we will further call them simply neural networks or networks). Information about neural networks can be found in the works [1–4]. The work is a continuation of the works [1, 5, 6]. In the work [1] for each network $\mathcal{N}$, a commutative and idempotent groupoid is constructed $\mathrm{AGS}(\mathcal{N})$. This groupoid is called the additive groupoid of neural network subnets of the neural network $\mathcal{N}$. In the work [6] a multiplicative groupoid of subnets $\mathrm{MGS}(\mathcal{N})$ is constructed. The supports of the groupoids $\mathrm{AGS}(\mathcal{N})$ and $\mathrm{MGS}(\mathcal{N})$ coincide.

The connection between elements of groupoids $\mathrm{AGS}(\mathcal{N})$ and $\mathrm{MGS}(\mathcal{N})$ with neural network subnets $\mathcal{N}$ is discussed in [1,6]. Article [1] introduces the concept of a subnetwork of a multilayer feedforward neural network (see Definition 4 of [1]). Subnet data is obtained from the original network by disabling a certain set of neurons. After switching off the selected neurons, the synaptic

*anm11@rambler.ru https://orcid.org/0000-0001-6285-0201
†tanya-mois11@yandex.ru https://orcid.org/0009-0009-2216195X

connections that connect the excluded neurons to any other neurons disappear. The remaining neurons and synaptic connections have the same architectural parameters as in the original network. That is, the activation functions, threshold values, weights of synaptic connections for the neurons and synaptic connections remaining in the subnetwork do not change. Elements of a groupoid $\mathrm{AGS}(\mathcal{N})$ (hence, $\mathrm{MGS}(\mathcal{N})$) contain information about the neurons remaining after switching off. The operation in the groupoid $\mathrm{AGS}(\mathcal{N})$ allows you to model the merging (i.e. unioning) of two subnets into one network, whenever possible. The groupoid operation $\mathrm{MGS}(\mathcal{N})$ allows you to model the intersection of two subnets when possible.

Objectives of the work. Introduction of new groupoids that allow modeling of various processes associated with neural networks. Studying the properties of a neural network depending on the algebraic properties of groupoids built on this neural network.

Main results. This work expands the concept *multilayer feedforward neural network*. By virtue of Definition 3 of [1], a neural network must have at least two layers of neurons. The latter seemed justified in the context that it is networks with at least 2 layers that are of practical value. But this led to excessive formalism. Thus, some elements of the groupoid $\mathrm{AGS}(\mathcal{N})$ could be associated with subnets of the neural network $\mathcal{N}$, but many others could not. At the same time, in practice, situations arise when it is convenient to carry out various manipulations with layers of neurons. In other works, neural networks were composed of neurons (see, for example, [2]), which were associated with abstract automata.

Definition 3 of [1] in this work has been modified so that a neural network can have one layer of neurons (see Definition 1.1). A single neuron can now also be considered a neural network by Definition 1.1. In this work, the concept of a neural network subnet was expanded (see Definition 1.3). Now a neural network subnet can consist of neurons of one layer (see Definition 2.1). One neuron can now also be considered a subnetwork. In Definition 4 of [1] subnetworks were required to contain neurons of at least two layers.

The Definition 2.1 introduces the concept of a *generalized neural network subnet*. This concept allows us to consider generalized subnetworks in which a certain selected set of synaptic connections has been disconnected. The disconnection of a synaptic link is modeled by assigning a weight of zero to that synaptic link. The introduction of this feature is justified from a practical point of view. In practice, it can be convenient to remove weak synaptic connections from a trained neural network (that is, weight connections that are small enough and have little effect on the operation of the network). The latter leads to improved performance of the algorithm built on neural network principles.

The introduction of the concept of a generalized subnetwork of a neural network leads to the appearance of *additive groupoid of generalized subnets* and *multiplicative groupoid of generalized subnets*: $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$ (see Definition 2.2). The elements of these groupoids now carry information about the neurons that remain after removing all other neurons, and about the synaptic connections that will be disconnected. Operations in these groupoids will continue to model the union and intersection of neural network subnets.

Let $n(\mathcal{N})$ denote the number of layers of neurons in the network $\mathcal{N}$. The main results of the work are formulated in the form of Theorems 3.1, 3.2 and 4.1. The groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ is associative iff $n(\mathcal{N}) = 1$ or $n(\mathcal{N}) = 2$. The groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$ is associative iff in the neural network $\mathcal{N}$ only the first and last layers have more than one neuron. In particular, gruppoid $\widehat{\mathrm{MGS}}(\mathcal{N})$ is associative if $n(\mathcal{N}) = 1$ or $n(\mathcal{N}) = 2$; in these cases, there are no restrictions on the layers of neurons. Thus, we see that the associativity condition for the groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$ imposes restrictions on architecture (i.e. structure) of the neural network $\mathcal{N}$.

Theorem 4.1 reveals the connection between the generalized subnetwork $\mathcal{N}'$ networks $\mathcal{N}$ and subgroupoids of groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$.

Algebraic properties of groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$ are closely related to the structure of the graph of the neural network $\mathcal{N}$ (this is confirmed by Theorems 3.1, 3.2).

1. Basic definitions

Further, $\mathbb{R}$ is the set of real numbers and $F(\mathbb{R}) := \mathrm{Hom}(\mathbb{R}, \mathbb{R})$ is the set of all mappings from $\mathbb{R}$ to $\mathbb{R}$.

Definition 1.1. Let the following objects be given:
1) the tuple $(M_1, \ldots, M_n)$ of length $n \geqslant 1$ of finite non-empty sets, where for $i \neq j$ the condition $M_i \cap M_j = \varnothing$ is satisfied;
2) the set $S := (M_1 \times M_2) \cup (M_2 \times M_3) \cup \cdots \cup (M_{n-1} \times M_n)$;
3) the mapping $f : S \to \mathbb{R}$;
4) the set $A := M_1 \cup \cdots \cup M_n$;
5) the mapping $g : A \to F(\mathbb{R})$;
6) the mapping $l : A \to \mathbb{R}$.

Then the tuple $\mathcal{N} = (M_1, \ldots, M_n, f, g, l)$ will be called *a multilayer feedforward neural network.*

Neural network operation. Each neural network $\mathcal{N} = (M_1, \ldots, M_n, f, g, l)$ and each two bijections

$$i : M_1 :\to \{1, \ldots, |M_1|\}, \quad o : M_n \to \{1, \ldots, |M_n|\}$$

corresponds to the mapping $F_{i,o,\mathcal{N}} : \mathbb{R}^{|M_1|} \to \mathbb{R}^{|M_n|}$, which implements the operation of a neural network as a computing circuit. The mapping $F_{i,o,\mathcal{N}}$ is defined using an artificial neuron (McCulloch–Pitts; see [2]) model. If compositions of neural networks are studied, then the bijections i and o must be written into the definition of definition 1.1 (see [7]).

Standard notations associated with neural networks. We will associate the following notations with each neural network $\mathcal{N} = (M_1, \ldots, M_n, f, g, l)$:

$$n(\mathcal{N}) = n, \quad A(\mathcal{N}) = \bigcup_{i=1}^{n} M_i, \quad Syn(\mathcal{N}) = \bigcup_{i=1}^{n-1} M_i \times M_{i+1}.$$

Thus, $n(\mathcal{N})$ is the number of all layers of the neural network, $A(\mathcal{N})$ is the set of all neurons, and $Syn(\mathcal{N})$ is the set all synaptic connections. We will call the tuple $(M_1, \ldots, M_n)$ *the main tuple of neurons* of the network $\mathcal{N}$.

A tuple of empty sets will be denoted by the symbol $\overline{\varnothing} := (\varnothing, \ldots, \varnothing)$ (the length of such a tuple will always be clear from the context). Let two tuples $\overline{X} = (X_1, \ldots, X_n)$ and $\overline{Y} = (Y_1, \ldots, Y_n)$ of finite sets be given. Then we will use the notation

$$\overline{X} \cup \overline{Y} := (X_1 \cup Y_1, \ldots, X_n \cup Y_n); \quad \overline{X} \cap \overline{Y} := (X_1 \cap Y_1, \ldots, X_n \cap Y_n);$$

$$\overline{X} \subseteq \overline{Y} \Leftrightarrow (X_1 \subseteq Y_1) \wedge (X_2 \subseteq Y_2) \wedge \cdots \wedge (X_n \subseteq Y_n).$$

Definition 1.2. Let $(X_1, \ldots, X_n)$ be some tuple composed of finite sets. We will say that the tuple is *continuous* if for any distinct $i, j \ in \{1, \ldots, n\}$ the following implication holds: if $X_i \neq \varnothing$ and $X_j \neq \varnothing$ and $i < j$, then for any $s \in \{i, i+1, \ldots, j-1, j\}$ the inequality $X_s \neq \varnothing$ holds. The tuple $\overline{\varnothing}$ is considered continuous by definition.

For a tuple of sets to be continuous, it must not contain an alternation of a non-empty set with an interval of empty sets, and then again with a non-empty set.

Let us introduce a definition of *subnet*, similar to Definition 4 from [1]. But it differs from it in that single-layer networks can now also be subnets.

Definition 1.3. Let the neural network be defined $\mathcal{N} = (M_1, \ldots, M_n, f, g, l)$ and a continuous tuple $(X_1, \ldots, X_n)$ is given such that the conditions are satisfied $(X_1, \ldots, X_n) \subseteq (M_1, \ldots, M_n)$ and $(X_1, \ldots, X_n) \neq \overline{\varnothing}$. We assume that $(Y_1, \ldots, Y_m)$ is a tuple obtained from a tuple $(X_1, \ldots, X_n)$ by deleting components equal to the empty set, where $m \leqslant n$. If f' is the restriction of the function f on the set $S' := (Y_1 \times Y_2) \cup (Y_2 \times Y_3) \cup \cdots \cup (Y_{m-1} \times Y_m)$ and g', l' are the restriction of the function g and the restriction of the function l on the set $A' := Y_1 \cup \cdots \cup Y_m$, then the object $\mathcal{N}' := (Y_1, \ldots, Y_m, f', g', l')$ will be called *subnet* of the network $\mathcal{N}$. We will say that the tuple $(X_1, \ldots, X_n)$ *induces* a subnetwork $\mathcal{N}'$. The tuple $(Y_1, \ldots, Y_m)$ is the main tuple of neurons of the subnetwork $\mathcal{N}'$. In general, the tuples $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ can be different.

Groupoids AGS($\mathcal{N}$) are introduced into [1], and groupoids MGS($\mathcal{N}$) are introduced into [6]. For the convenience of the reader, we give an explicit definition below.

Definition 1.4. Let a neural network $\mathcal{N} = (M_1, \ldots, M_n, f, g, l)$ be defined with a main tuple of neurons $\overline{M} = (M_1, \ldots, M_n)$. The set of all possible continuous tuples $\overline{X} \subseteq \overline{M}$ will be denoted by the symbol AGS($\mathcal{N}$). Further, $\overline{X}$ and $\overline{Y}$ are two arbitrary element from AGS($\mathcal{N}$). Let us define binary algebraic operations $(+)$ and $(*)$ on the set AGS($\mathcal{N}$):

$$\overline{X} + \overline{Y} := \begin{cases} \overline{X} \cup \overline{Y}, & \text{if } \overline{X} \cup \overline{Y} \in \mathrm{AGS}(\mathcal{N}), \\ \overline{\varnothing}, & \text{if } \overline{X} \cup \overline{Y} \notin \mathrm{AGS}(\mathcal{N}); \end{cases} \quad \overline{X} * \overline{Y} := \begin{cases} \overline{X} \cap \overline{Y}, & \text{if } \overline{X} \cap \overline{Y} \in \mathrm{AGS}(\mathcal{N}), \\ \overline{\varnothing}, & \text{if } \overline{X} \cap \overline{Y} \notin \mathrm{AGS}(\mathcal{N}). \end{cases}$$

Then the groupoid $\mathrm{AGS}(\mathcal{N}) := (\mathrm{AGS}(\mathcal{N}), +)$ will be called *additive groupoid of neural network subnets* $\mathcal{N}$. The groupoid $\mathrm{MGS}(\mathcal{N}) := (\mathrm{AGS}(\mathcal{N}), *)$ will be called *the multiplicative groupoid of neural network subnets* $\mathcal{N}$.

Remark 1.1. Each tuple $\overline{X} \neq \overline{\varnothing}$ of AGS($\mathcal{N}$) induces some subnetwork. Two different tuples from AGS($\mathcal{N}$) induce different subnets of the network $\mathcal{N}$ (this follows trivially from the definition of 1.3). Each subnet of the network $\mathcal{N}$ is induced by some tuple from AGS($\mathcal{N}$). Thus, there is a bijection between the set of all subnets of the network $\mathcal{N}$ and the set $\mathrm{AGS}(\mathcal{N}) \setminus \{\overline{\varnothing}\}$.

Remark 1.2. The changes made to Definitions 3 and 4 from [1] does not change the contents of the set AGS($\mathcal{N}$). Additionally, these changes do not affect definitions of operation: $(+)$ and $(*)$. These changes allow more elements in AGS($\mathcal{N}$) to be associated with subnets. Continuous tuples with only one layer different from the empty set did not induce any subnetworks due to Definition 4 of [1]. Because the subnets from that definition had at least two layers. If $\mathcal{G}_1(\mathcal{N})$ is the set of all subnets by Definition 4 of [1] and $\mathcal{G}_2(\mathcal{N})$ is the set of all subnets by Definition 1.3, then the inclusion $\mathcal{G}_1(\mathcal{N}) \subset \mathcal{G}_2(\mathcal{N})$.

2. Generalized subnets

The concept of a neural network subnet, introduced by the Definition 1.3, describes objects obtained from the original network by switching off a certain set of neurons and the deleting of synaptic connections associated with disconnected neurons. Let's build a model of a generalized subnetwork that describes objects that can be obtained by turning off a certain set of neurons and resetting the weights of a given set of synaptic connections to zero.

The set of all subsets of the set X, as usual, will be denoted by 2^X. Let $\mathcal{N} = (M_1, \ldots, M_n, f, g, l)$. Then we introduce the set

$$\widehat{\mathrm{AGS}}(\mathcal{N}) := \mathrm{AGS}(\mathcal{N}) \times 2^{Syn(\mathcal{N})}.$$

Elements from $\widehat{\mathrm{AGS}}(\mathcal{N})$ will be denoted by capital Latin letters with a cap.

Definition 2.1. Let $\mathcal{N}' = (Y_1, \ldots, Y_m, f', g', l')$ is subnet of network $\mathcal{N}$, which is induced by the tuple $\overline{X}$ from $\mathrm{AGS}(\mathcal{N})$. We assume that $S' := (Y_1 \times Y_2) \cup (Y_2 \times Y_3) \cup \cdots \cup (Y_{m-1} \times Y_m)$ and Q is a certain subset of set $Syn(\mathcal{N})$. Let us introduce the mapping

$$f''(s) := \begin{cases} f'(s), & s \notin Q, \\ 0, & s \in Q \end{cases} \qquad (s \in S').$$

Then the object $\mathcal{N}' := (Y_1, \ldots, Y_m, f'', g', l')$ will be called a *generalized subnetwork* of the network $\mathcal{N}$. We will say that the tuple $\widehat{U} = (\overline{X}, Q)$ *induces* a generalized subnet $\mathcal{N}'$. Cortege $(Y_1, \ldots, Y_m)$ is the main tuple of neurons of the generalized subnet $\mathcal{N}'$.

Remark 2.1. A generalized subnet $\mathcal{N}'$ is an object that satisfies the definition of 1.1. Various tuples from $\widehat{\mathrm{AGS}}(\mathcal{N})$ can induce one generalized subnet $\mathcal{N}'$ of the network $\mathcal{N}$ (an important difference with the case of simple subnets, see remark 1.1). A tuple $\widehat{U}$ induces a subnet of the network $\mathcal{N}$ if and only if it contains in the set

$$\widehat{\mathrm{AGS}}(\mathcal{N}) \setminus \{(\overline{\varnothing}, W) \mid W \subseteq Syn(\mathcal{N})\}.$$

Definition 2.2. We assume that the neural network $\mathcal{N} = (M_1, \ldots, M_n, f, g, l)$ is defined. Next, $\widehat{U}_1 = (\overline{X}_1, Q_1)$, $\widehat{U}_2 = (\overline{X}_2, Q_2)$ — these are two arbitrary elements from $\widehat{\mathrm{AGS}}(\mathcal{N})$. Let us define binary algebraic operations $(+)$ and $(*)$ on the set $\widehat{\mathrm{AGS}}(\mathcal{N})$:

$$\widehat{U}_1 + \widehat{U}_2 := \begin{cases} (\overline{X}_1 \cup \overline{X}_2, Q_1 \cup Q_2), & \text{if } \overline{X}_1 \cup \overline{X}_2 \in \mathrm{AGS}(\mathcal{N}), \\ (\overline{\varnothing}, \varnothing), & \text{if } \overline{X}_1 \cup \overline{X}_2 \notin \mathrm{AGS}(\mathcal{N}); \end{cases} \tag{1}$$

$$\widehat{U}_1 * \widehat{U}_2 := \begin{cases} (\overline{X}_1 \cap \overline{X}_2, Q_1 \cap Q_2), & \text{if } \overline{X}_1 \cap \overline{X}_2 \in \mathrm{AGS}(\mathcal{N}), \\ (\overline{\varnothing}, \varnothing), & \text{if } \overline{X}_1 \cap \overline{X}_2 \notin \mathrm{AGS}(\mathcal{N}). \end{cases} \tag{2}$$

Then the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N}) := (\widehat{\mathrm{AGS}}(\mathcal{N}), +)$ will be called *the additive groupoid of generalized neural network subnets* $\mathcal{N}$ and groupoid $\widehat{\mathrm{MGS}}(\mathcal{N}) := (\widehat{\mathrm{AGS}}(\mathcal{N}), *)$ we will call *multiplicative groupoid of generalized neural network subnets* $\mathcal{N}$.

Remark 2.2. Operations in groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$ are also denoted as in the groupoids $\mathrm{AGS}(\mathcal{N})$ and $\mathrm{MGS}(\mathcal{N})$, respectively. In practice this does not lead to confusion. It is always clear from the context what operation is meant. Further, for tuples from $\widehat{\mathrm{AGS}}(\mathcal{N})$ it will be convenient to use the operation of componentwise union and intersection. If $\widehat{U}_1 = (\overline{X}_1, Q_1)$ and $\widehat{U}_2 = (\overline{X}_2, Q_2)$ then $\widehat{U}_1 \cup \widehat{U}_2 := (\overline{X}_1 \cup \overline{X}_2, Q_1 \cup Q_2)$ and $\widehat{U}_1 \cap \widehat{U}_2 := (\overline{X}_1 \cap \overline{X}_2, Q_1 \cap Q_2)$.

Remark 2.3. The additive generalized subnet groupoid models the merging of two subnets into one when possible and returns the tuple $(\overline{\varnothing}, \varnothing)$ when this is not possible. The multiplicative groupoid of generalized subnets models the intersection of two subnets (i.e., returns a subnet that is contained in both networks) when possible and returns the tuple $(\overline{\varnothing}, \varnothing)$ when this is not possible.

3. Basic algebraic properties

The main result of this section is expressed in the form of Theorems 3.1 and 3.2. First, we formulate and prove some algebraic properties of additive and multiplicative groupoids of generalized subnets (see Properties 3.1, 3.2 and 3.3).

Property 3.1. *For any neural network $\mathcal{N}$ the following statements are satisfied:*
1) groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$ are commutative and idempotent;
2) the tuple $(\overline{\varnothing}, \varnothing)$ is a neutral element of the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$;
3) the tuple $(\overline{\varnothing}, \varnothing)$ has the multiplicative zero property in the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$;
4) the tuple $((M_1, \dots, M_n), Syn(\mathcal{N}))$ is a neutral element in the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$, where $(M_1, \dots, M_n)$ is the main tuple of neurons of the network $\mathcal{N}$;
5) the tuple $((M_1, \dots, M_n), Syn(\mathcal{N}))$ has the multiplicative zero property in the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$.

Proof. Commutativity and idempotency of the groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$ is trivial follows from (1) and (2).

Statements 2) – 5) follow from the definitions of the operations $(+)$ and $(*)$. Indeed, let $\widehat{U} = (\overline{X}, Q)$ be an arbitrary element of the set $\widehat{\mathrm{AGS}}(\mathcal{N})$ (hence, it is an element of groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$). We assume that $(M_1, \dots, M_n)$ is the main tuple of neurons in the network $\mathcal{N}$. Then the equalities

$$\widehat{U} + (\overline{\varnothing}, \varnothing) = (\overline{X} \cup \overline{\varnothing}, Q \cup \varnothing) = \widehat{U}, \quad \widehat{U} * (\overline{\varnothing}, \varnothing) = (\overline{X} \cap \overline{\varnothing}, Q \cap \varnothing) = (\overline{\varnothing}, \varnothing),$$

$$\widehat{U} * ((M_1, \dots, M_n), Syn(\mathcal{N})) = (\overline{X} \cap (M_1, \dots, M_n), Q \cap Syn(\mathcal{N})) = (\overline{X}, Q),$$

$$\widehat{U} + ((M_1, \dots, M_n), Syn(\mathcal{N})) = (\overline{X} \cup (M_1, \dots, M_n), Q \cup Syn(\mathcal{N})) = ((M_1, \dots, M_n), Syn(\mathcal{N}))$$

show the validity of statements 2)–5). □

Property 3.2. *If $\overline{X}_1 \cup \overline{X}_2, \overline{Y}_1 \cap \overline{Y}_2 \in \mathrm{AGS}(\mathcal{N})$, then for elements $\widehat{U}_1 = (\overline{X}_1, Q_1)$, $\widehat{U}_2 = (\overline{X}_2, Q_2)$ of the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ and elements $\widehat{U}_3 = (\overline{Y}_1, W_1)$, $\widehat{U}_4 = (\overline{Y}_2, W_2)$ of the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$ the equalities hold*

$$\widehat{U}_1 + \widehat{U}_2 = (\overline{X}_1 + \overline{X}_2, Q_1 \cup Q_2), \quad \widehat{U}_3 * \widehat{U}_4 = (\overline{Y}_1 * \overline{Y}_2, W_1 \cap W_2). \tag{3}$$

Proof. Since on the left side of the equalities (3) the operations $(+)$ and $(*)$ are operations of groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$, and on the right side these are groupoid operations $\mathrm{AGS}(\mathcal{N})$ and $\mathrm{MGS}(\mathcal{N})$, then by equalities (1) and (2) the equalities are satisfied

$$(\overline{X}_1 + \overline{X}_2, Q_1 \cup Q_2) = \begin{cases} (\overline{X}_1 \cup \overline{X}_2, Q_1 \cup Q_2), & \text{if } \overline{X}_1 \cup \overline{X}_2 \in \mathrm{AGS}(\mathcal{N}) \\ (\overline{\varnothing}, \varnothing), & \text{if } \overline{X}_1 \cup \overline{X}_2 \notin \mathrm{AGS}(\mathcal{N}) \end{cases} = \widehat{U}_1 + \widehat{U}_2,$$

$$(\overline{X}_1 * \overline{X}_2, Q_1 \cap Q_2) = \begin{cases} (\overline{X}_1 \cap \overline{X}_2, Q_1 \cap Q_2), & \text{if } \overline{X}_1 \cap \overline{X}_2 \in \mathrm{AGS}(\mathcal{N}) \\ (\overline{\varnothing}, \varnothing), & \text{if } \overline{X}_1 \cap \overline{X}_2 \notin \mathrm{AGS}(\mathcal{N}) \end{cases} = \widehat{U}_1 * \widehat{U}_2,$$

which give equalities (3). □

We define mapping $\Psi : \widehat{\mathrm{AGS}}(\mathcal{N}) \to \mathrm{AGS}(\mathcal{N})$ as follows $\Psi((\overline{X}, Q)) := \overline{X}$, where $(\overline{X}, Q) \in \widehat{\mathrm{AGS}}(\mathcal{N})$ and $\overline{X} \in \mathrm{AGS}(\mathcal{N})$.

Property 3.3. *The following statements are true:*
1) the mapping Ψ *is a homomorphism of the groupoid* $\widehat{\mathrm{AGS}}(\mathcal{N})$ *into the groupoid* $\mathrm{AGS}(\mathcal{N})$*;*
2) the mapping Ψ *is a homomorphism of the groupoid* $\widehat{\mathrm{MGS}}(\mathcal{N})$ *into the groupoid* $\mathrm{MGS}(\mathcal{N})$*;*
3) the sets $\Phi(\widehat{\mathrm{AGS}}(\mathcal{N})) = \mathrm{AGS}(\mathcal{N})$ *are equal.*

Proof. Let
$$\widehat{U}_1 = (\overline{X}_1, Q_1),\ \widehat{U}_2 = (\overline{X}_2, Q_2),\ \widehat{U}_3 = (\overline{Y}_1, W_1),\ \widehat{U}_4 = (\overline{Y}_2, W_2)$$
these are arbitrary elements of the groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$. We assume that $\overline{X}_1 \cup \overline{X}_2$ and $\overline{Y}_1 \cap \overline{Y}_2$ are continuous tuples (i.e. tuples from $\mathrm{AGS}(\mathcal{N})$). Then, by virtue of the equalities (3), the equalities
$$\Psi(\widehat{U}_1+\widehat{U}_2) = \Psi((\overline{X}_1+\overline{X}_2, Q_1\cup Q_2)) = \overline{X}_1+\overline{X}_2 = \Psi((\overline{X}_1, Q_1))+\Psi((\overline{X}_2, Q_2)) = \Psi(\widehat{U}_1)+\Psi(\widehat{U}_2),$$
$$\Psi(\widehat{U}_3 * \widehat{U}_4) = \Psi((\overline{Y}_1 * \overline{Y}_2, W_1 \cap W_2)) = \overline{Y}_1 * \overline{Y}_2 = \Psi((\overline{Y}_1, W_1)) * \Psi((\overline{Y}_2, W_2)) = \Psi(\widehat{U}_3) * \Psi(\widehat{U}_4).$$
Let now the tuples $\overline{X}_1 \cup \overline{X}_2$ and $\overline{Y}_1 \cap \overline{Y}_2$ not belong to $\mathrm{AGS}(\mathcal{N})$. Then we have equalities
$$\Psi(\widehat{U}_1 + \widehat{U}_2) = \Psi((\overline{\varnothing}, \varnothing)) = \overline{\varnothing} = \overline{X}_1 + \overline{X}_2 = \Psi((\overline{X}_1, Q_1)) + \Psi((\overline{X}_2, Q_2)) = \Psi(\widehat{U}_1) + \Psi(\widehat{U}_2),$$
$$\Psi(\widehat{U}_3 * \widehat{U}_4) = \Psi((\overline{\varnothing}, \varnothing)) = \overline{\varnothing} = \overline{Y}_1 * \overline{Y}_2 = \Psi((\overline{Y}_1, Q_1)) * \Psi((\overline{Y}_2, Q_2)) = \Psi(\widehat{U}_3) * \Psi(\widehat{U}_4).$$
These equalities show that Ψ is a homomorphism from statements 1) and 2). Statements 1) and 2) have been proven. Statement 3) follows from the definition of the set $\widehat{\mathrm{AGS}}(\mathcal{N})$. The property is proved. □

Theorem 3.1. *For any neural network* $\mathcal{N}$ *the following statements are equivalent.*
1) The condition $n(\mathcal{N}) \in \{1, 2\}$ *is satisfied.*
2) The groupoid $\mathrm{AGS}(\mathcal{N})$ *is associative.*
3) The groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ *is associative.*

Proof. Let us show that statement 1) is equivalent to statement 2). Let statement 1) be true. Then for any $\overline{X}_1, \overline{X}_2 \in \mathrm{AGS}(\mathcal{N})$ tuple $\overline{X}_1 + \overline{X}_2$ will be a continuous tuple (since there is no way to get a discontinuous tuple). Therefore, for any $\overline{Y}_1, \overline{Y}_2, \overline{Y}_3 \in \mathrm{AGS}(\mathcal{N})$ the relations
$$(\overline{Y}_1+\overline{Y}_2)+\overline{Y}_3 = (\overline{Y}_1\cup\overline{Y}_2)\cup\overline{Y}_3 = \overline{Y}_1\cup\overline{Y}_2\cup\overline{Y}_3, \quad \overline{Y}_1+(\overline{Y}_2+\overline{Y}_3) = \overline{Y}_1\cup(\overline{Y}_2\cup\overline{Y}_3) = \overline{Y}_1\cup\overline{Y}_2\cup\overline{Y}_3.$$
These relations show that the groupoid $\mathrm{AGS}(\mathcal{N})$ is associative. Therefore, the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ is associative. Thus, from 1) it follows 2).

On the other hand, suppose that statement 2) holds and statement 1) does not hold. Then $n(\mathcal{N}) > 2$. In this case, for any network $\mathcal{N}$ it is always possible to specify three tuples $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$ for which the condition is satisfied $\overline{X}_1 + (\overline{X}_2 + \overline{X}_3) \neq (\overline{X}_1 + \overline{X}_2) + \overline{X}_3$. For example, you can take tuples:
$$\overline{X}_1 = (\{a\}, \varnothing, \varnothing, \ldots, \varnothing),\ \overline{X}_2 = (\varnothing, \{b\}, \varnothing, \ldots, \varnothing),\ \overline{X}_3 = (\varnothing, \varnothing, \{c\}, \ldots, \varnothing).$$
A contradiction has been obtained. It shows that from 2) follows (1). This means that statements 1) and 2) are equivalent.

Let us show that statements 2) and 3) are equivalent. Let statement 2) be true. The groupoid $\mathrm{AGS}(\mathcal{N})$ is associative if and only if $n(\mathcal{N}) \in \{1, 2\}$. Then, as noted above, for any $\overline{X}_1, \overline{X}_2 \in \mathrm{AGS}(\mathcal{N})$ tuple $\overline{X}_1 \cup \overline{X}_2$ is a continuous tuple. Therefore, by virtue of the equality (3) for any elements $\widehat{U}_1 = (\overline{Y}_1, Q_1)$, $\widehat{U}_2 = (\overline{Y}_2, Q_2)$, $\widehat{U}_3 = (\overline{Y}_3, Q_3)$ groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ the following relations hold:

$$(\widehat{U}_1 + \widehat{U}_2) + \widehat{U}_3 = (\overline{Y}_1 + \overline{Y}_2, Q_1 \cup Q_2) + (\overline{Y}_3, Q_3) = ((\overline{Y}_1 + \overline{Y}_2) + \overline{Y}_3, (Q_1 \cup Q_2) \cup Q_3) =$$

$$= (\overline{Y}_1 \cup \overline{Y}_2 \cup \overline{Y}_3, Q_1 \cup Q_2 \cup Q_3) = (\overline{Y}_1 + (\overline{Y}_2 + \overline{Y}_3), Q_1 \cup (Q_2 \cup Q_3)) = \widehat{U}_1 + (\widehat{U}_2 + \widehat{U}_3).$$

From this we obtain the associativity of the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$. Thus, from 2) follows 3).

Statement 3) implies statement 2). Indeed, since Ψ is a homomorphism of $\widehat{\mathrm{AGS}}(\mathcal{N})$ into $\mathrm{AGS}(\mathcal{N})$ and $\Phi(\widehat{\mathrm{AGS}}(\mathcal{N})) = \mathrm{AGS}(\mathcal{N})$ (see property 3.3), then from the associativity of the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ implies the associativity of the groupoid $\mathrm{AGS}(\mathcal{N})$.

Statements 2) and 3) are equivalent. Since 2) is equivalent to 1), then 3) is equivalent to 1). The theorem is proved. $\square$

Remark 3.1. Statement 2 of [1] states that the groupoid $\mathrm{AGS}(\mathcal{N})$ is associative if and only if $\mathcal{N}$ is a two-layer neural network. The discrepancy with the results of Theorem 3.1 is due to the fact that in the work [1] single-layer neural networks were not considered. Taking this fact into account, it can be argued that the results of Statement 2 of [1] and Theorem 3.1 are consistent.

Theorem 3.2. *For any neural network $\mathcal{N}$ the following statements are equivalent.*
1) In a neural network $\mathcal{N} = (M_1, \ldots, M_n, f, g, l)$, only the input layer M_1 and the output layer M_n can contain more than one neuron.
2) The groupoid $\mathrm{MGS}(\mathcal{N})$ is associative.
3) The groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$ is associative.

Proof. Let us show that statements 1) and 2) are equivalent. Let statement 1) be true. If $n(\mathcal{N}) \in \{1, 2\}$, then for any $\overline{X} = (X_1, \ldots, X_n)$ and $\overline{Y} = (Y_1, \ldots, Y_n)$ from $\mathrm{MGS}(\mathcal{N})$ the tuple $\overline{X} \cap \overline{Y}$ is continuous. In this case, $\overline{X} * \overline{Y} = \overline{X} \cap \overline{Y}$. Due to the associativity of the operation $(\cap)$ on sets, we have the associativity of the operation $(\cap)$ on tuples from $\mathrm{MGS}(\mathcal{N})$. Therefore the groupoid $\mathrm{MGS}(\mathcal{N})$ is associative.

We assume that $n(\mathcal{N}) > 2$ and $\overline{X} = (X_1, \ldots, X_n)$, $\overline{Y} = (Y_1, \ldots, Y_n)$ are two elements of the groupoid $\mathrm{MGS}(\mathcal{N})$ such that the tuple $\overline{X} \cap \overline{Y}$ is not continuous. This means that the following conditions are met:
(c.1) there is an index $i \in \{1, \ldots, n\}$ such that $X_i \cap Y_i = \varnothing$;
(c.2) there are indices $u, v \in \{1, \ldots, n\}$ such that the conditions are satisfied

$$u < i < v, \ X_u \cap Y_u \neq \varnothing, \ X_v \cap Y_v \neq \varnothing.$$

From (c.2) the conditions follow: $i \neq 1$ and $i \neq n$. Condition (c.1) cannot be satisfied. Indeed, since Statement 3) holds, then $X_i = Y_i = \{a\}$, where a is an element of layer M_i. Thus, we have shown that for any $\overline{X}$ and $\overline{Y}$ from The $\mathrm{MGS}(\mathcal{N})$ tuple $\overline{X} \cap \overline{Y}$ is continuous. Consequently, the identity $\overline{X} * \overline{Y} = \overline{X} \cap \overline{Y}$ holds. Therefore the groupoid $\mathrm{MGS}(\mathcal{N})$ is associative. Statement 1) gives statement 2).

Let us show that statement 2) implies statement 1). Let the groupoid $\mathrm{MGS}(\mathcal{N})$ is associative and statement 1) does not hold. Since statement 1) does not hold, then $n(\mathcal{N}) > 2$ and there is an index $i \notin \{1, n\}$ such that layer M_i contains more than one neuron. For any tuple $\overline{Y}$ from $\mathrm{MGS}(\mathcal{N})$ we denote by $K_s(\overline{Y})$ the s-th component of the tuple $\overline{Y}$. For any network $\mathcal{N}$ with the specified conditions, we can define tuples $\overline{Y}_1$, $\overline{Y}_2$ and $\overline{Y}_3$ from $\mathrm{MGS}(\mathcal{N})$ so that the following conditions are satisfied:

$$K_{i-1}(\overline{Y}_1) = \{a\}, \ K_i(\overline{Y}_1) = \{b\}, \ K_{i+1}(\overline{Y}_1) = \{c\}, \ K_s(\overline{Y}_1) = \varnothing \ \ (s \notin \{i-1, i, i+1\});$$

$$K_{i-1}(\overline{Y}_2) = \{a\},\ K_i(\overline{Y}_2) = \{m\},\ K_{i+1}(\overline{Y}_2) = \{c\},\ K_s(\overline{Y}_2) = \varnothing\ \ (s \notin \{i-1, i, i+1\});$$
$$K_{i-1}(\overline{Y}_3) = \varnothing,\ K_i(\overline{Y}_3) = \varnothing,\ K_{i+1}(\overline{Y}_3) = \{c\},\ K_s(\overline{Y}_3) = \varnothing\ \ (s \notin \{i-1, i, i+1\})$$
$$(a \in M_{i-1},\ b, m \in M_i,\ c \in M_{i+1}).$$

Then the equalities hold $(\overline{Y}_1 * \overline{Y}_2) * \overline{Y}_3 = \overline{\varnothing}$, $\overline{Y}_1 * (\overline{Y}_2 * \overline{Y}_3) = \overline{Y}_3$. The equality data shows the lack of associativity in the groupoid $\mathrm{MGS}(\mathcal{N})$ if $|M_i| > 2$ and $i \neq 1, n$. This contradiction shows that statement 1) must be true if statement 2) is true. Statements 1) and 2) are equivalent.

Let us show that from statement 2) and 3) are equivalent. Let 2 be fulfilled The groupoid $\mathrm{MGS}(\mathcal{N})$ is associative if and only if Statement 1) holds. Therefore, for any $\overline{X}_1, \overline{X}_2 \in \mathrm{MGS}(\mathcal{N})$ tuple $\overline{X}_1 \cap \overline{X}_2$ is a continuous tuple. Therefore, by virtue of equalities (3) for any elements $\widehat{U}_1 = (\overline{Y}_1, Q_1)$, $\widehat{U}_2 = (\overline{Y}_2, Q_2)$, $\widehat{U}_3 = (\overline{Y}_3, Q_3)$ groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$ the following relations hold:

$$(\widehat{U}_1 * \widehat{U}_2) * \widehat{U}_3 = (\overline{Y}_1 * \overline{Y}_2, Q_1 \cap Q_2) * (\overline{Y}_3, Q_3) = ((\overline{Y}_1 * \overline{Y}_2) * \overline{Y}_3, (Q_1 \cap Q_2) \cap Q_3) =$$
$$= (\overline{Y}_1 \cap \overline{Y}_2 \cap \overline{Y}_3, Q_1 \cap Q_2 \cap Q_3) = (\overline{Y}_1 * (\overline{Y}_2 * \overline{Y}_3), Q_1 \cap (Q_2 \cap Q_3)) = \widehat{U}_1 * (\widehat{U}_2 * \widehat{U}_3).$$

From this we obtain the associativity of the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$. Thus, from 2) follows 3).

Statement 3) implies statement 2). Indeed, since the groupoid $\mathrm{MGS}(\mathcal{N})$ is a homomorphic image of the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$ (see property 3.3), then the associativity of $\widehat{\mathrm{MGS}}(\mathcal{N})$ implies the associativity of $\mathrm{MGS}(\mathcal{N})$. Statements 2) and 3) are equivalent. Since 2) is equivalent to 1), then 3) is equivalent to 1). The theorem is proved. □

4. Generalized subnets and subgroupoids

Theorem 4.1. *Let $\mathcal{N}'$ be a generalized subnet of the neural network $\mathcal{N}$. Then the set $\widehat{\mathrm{AGS}}(\mathcal{N})$ has a subset $T(\mathcal{N}')$ such that this subset is a subgroupoid in the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ and a subgroupoid in the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$. In this case, the isomorphisms hold*

$$(T(\mathcal{N}'), +) \cong \widehat{\mathrm{AGS}}(\mathcal{N}'), \quad (T(\mathcal{N}'), *) \cong \widehat{\mathrm{MGS}}(\mathcal{N}').$$

Proof. Let the tuple $\widehat{U} = (\overline{X}, Q)$ induce a generalized subnet $\mathcal{N}'$. Let's build a set

$$T(\mathcal{N}') := \{(\overline{V}, W) \in \widehat{\mathrm{AGS}}(\mathcal{N}) \mid \overline{V} \subseteq \overline{X},\ W \subseteq Syn(\mathcal{N}')\}.$$

From the construction it is clear that $T(\mathcal{N}') \subseteq \widehat{\mathrm{AGS}}(\mathcal{N})$ does not depend on the set Q. The set $T(\mathcal{N}')$ contains the tuple $(\overline{\varnothing}, \varnothing)$. Moreover, $T(\mathcal{N}')$ is closed under the operation $(+)$ in the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$. Indeed, if $\widehat{T_1}, \widehat{T_2}$ are two arbitrary elements from $T(\mathcal{N}')$, then at least one of the conditions is satisfied: $\widehat{T_1} + \widehat{T_2} = (\overline{\varnothing}, \varnothing)$, $\widehat{T_1} + \widehat{T_2} = \widehat{T_1} \cup \widehat{T_2}$. In both cases we have $\widehat{T_1} + \widehat{T_2} \in T(\mathcal{N}')$. Thus, $T(\mathcal{N}')$ is a subgroupoid of the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$. Similarly, we obtain that $T(\mathcal{N}')$ is a subgroupoid of the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$.

Let us show that $(T(\mathcal{N}'), +)$ is isomorphic to $\widehat{\mathrm{AGS}}(\mathcal{N}')$. Since $\widehat{U} = (\overline{X}, Q)$ induces a generalized subnet $\mathcal{N}'$, then $\overline{X}$ is a continuous tuple. We assume that the first non-empty component of the tuple $\overline{X}$ has number u, and the last non-empty component has number v. Since $\overline{X}$ is a continuous tuple, the neural network $\mathcal{N}'$ has exactly $v - u + 1$ layers (follows from the definition). As before, let $K_s(\overline{Y})$ denote the s-th component of the tuple $\overline{Y}$ from $\mathrm{AGS}(\mathcal{N})$. Let us define a mapping $\alpha : T(\mathcal{N}') \to \widehat{\mathrm{AGS}}(\mathcal{N}')$ so that for an arbitrary element $(\overline{Y}, W) \in T(\mathcal{N}')$ and arbitrary $s \in \{1, \ldots, v - u + 1\}$ the equalities hold

$$\alpha((\overline{Y}, W)) := (\alpha(\overline{Y}), W), \quad K_s(\alpha(\overline{Y})) := K_{u+s-1}(\overline{Y}), \tag{4}$$

where $\alpha(\overline{Y})$ is the first component of the tuple $\alpha((\overline{Y},W))$ by definition. Since the tuple $\overline{X}$ is continuous and by virtue of the construction of the set $T(\overline{Y})$, then for any $\overline{Y}\in T(\mathcal{N}')$ and an arbitrary index $d\notin\{u,u+1,\dots,v\}$ we have $K_d(\overline{Y})=\varnothing$. Therefore the α-images of two distinct elements from $T(\mathcal{N}')$ are different (α is injective). The surjectivity of α follows easily from the definitions of the set $T(\mathcal{N}')$ and $\widehat{\mathrm{AGS}}(\mathcal{N}')$. Thus, α is a bijection of the set $T(\mathcal{N}')$ onto the set $\widehat{\mathrm{AGS}}(\mathcal{N}')$.

In what follows, operations in the groupoids $\widehat{\mathrm{AGS}}(\mathcal{N}')$ and $\mathrm{AGS}(\mathcal{N}')$ will be denoted by $(+')$. Let $\widehat{U}_1=(\overline{Y}_1,W_1)$ and $\widehat{U}_2=(\overline{Y}_2,W_2)$ be two arbitrary elements of $T(\mathcal{N}')$. There are possible cases: either $\overline{Y}_1\cup\overline{Y}_2$ is a continuous tuple, or $\overline{Y}_1\cup\overline{Y}_2$ is not a continuous tuple. Let the first case be true. Then the identity $\overline{Y}_1+\overline{Y}_2=\overline{Y}_1\cup\overline{Y}_2$ is true, due to (4) tuple $\alpha(\overline{Y}_1\cup\overline{Y}_2)$ is a continuous tuple. In addition, for any index $s\in\{1,\dots,v-u+1\}$ the following equalities hold:

$$\alpha(\widehat{U}_1+\widehat{U}_2)=\alpha((\overline{Y}_1,W_1)+(\overline{Y}_2,W_2))=\alpha(\overline{Y}_1+\overline{Y}_2,W_1\cup W_2)=(\alpha(\overline{Y}_1\cup\overline{Y}_2),W_1\cup W_2),$$

$$K_s(\alpha(\overline{Y}_1\cup\overline{Y}_2))=K_{u+s-1}(\overline{Y}_1\cup\overline{Y}_2)=K_{u+s-1}(\overline{Y}_1)\cup K_{u+s-1}(\overline{Y}_2)=K_s(\alpha(\overline{Y}_1))\cup K_s(\alpha(\overline{Y}_2)).$$

From the last chain of equalities we obtain the condition $\alpha(\overline{Y}_1)\cup\alpha(\overline{Y}_2)\in\mathrm{AGS}(\mathcal{N}')$ and the relations

$$\alpha(\overline{Y}_1\cup\overline{Y}_2)=\alpha(\overline{Y}_1)\cup\alpha(\overline{Y}_2)=\alpha(\overline{Y}_1)+'\alpha(\overline{Y}_2).$$

Therefore we have the relations

$$\alpha(\widehat{U}_1+\widehat{U}_2)=(\alpha(\overline{Y}_1\cup\overline{Y}_2),W_1\cup W_2)=(\alpha(\overline{Y}_1)\cup\alpha(\overline{Y}_2),W_1\cup W_2)=(\alpha(\overline{Y}_1)+'\alpha(\overline{Y}_2),W_1\cup W_2)=$$

$$=(\alpha(\overline{Y}_1),W_1)+'(\alpha(\overline{Y}_2),W_2)=\alpha(\widehat{U}_1)+'\alpha(\widehat{U}_2).$$

Now let $\overline{Y}_1\cup\overline{Y}_2$ be not a continuous tuple. Then, by virtue of (4), we have the relations $(\alpha(\overline{\varnothing}),\varnothing)=(\overline{\varnothing},\varnothing)'$, where $(\overline{\varnothing},\varnothing)'$ is a tuple in the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N}')$.

Let there exist parameters $d,m,k\in\{u,u+1\dots,v\}$ such that the conditions

$$d<m<k,\quad K_d(\overline{Y}_1\cup\overline{Y}_2)=K_d(\overline{Y}_1)\cup K_d(\overline{Y}_2)\neq\varnothing,\quad K_k(\overline{Y}_1\cup\overline{Y}_2)=K_k(\overline{Y}_1)\cup K_k(\overline{Y}_2)\neq\varnothing,$$

$$K_m(\overline{Y}_1\cup\overline{Y}_2)=K_m(\overline{Y}_1)\cup K_m(\overline{Y}_2)=\varnothing.$$

The last statement is a necessary and sufficient condition for the fact that $\overline{Y}_1\cup\overline{Y}_2\notin\mathrm{AGS}(\mathcal{N})$. From the given equalities we derive the conditions

$$K_{d-u+1}(\alpha(\overline{Y}_1))\cup K_{d-u+1}(\alpha(\overline{Y}_2))=K_d(\overline{Y}_1)\cup K_d(\overline{Y}_2)\neq\varnothing,$$

$$K_{k-u+1}(\alpha(\overline{Y}_1))\cup K_{k-u+1}(\alpha(\overline{Y}_2))=K_k(\overline{Y}_1)\cup K_k(\overline{Y}_2)\neq\varnothing,$$

$$K_{m-u+1}(\alpha(\overline{Y}_1))\cup K_{m-u+1}(\alpha(\overline{Y}_2))=K_m(\overline{Y}_1)\cup K_m(\overline{Y}_2)=\varnothing,\;\; d-u+1<m-u+1<k-u+1.$$

Therefore, the equality $\alpha(\widehat{U}_1)+'\alpha(\widehat{U}_2)=(\overline{\varnothing},\varnothing)'$ holds. From here we get

$$\alpha(\widehat{U}_1+\widehat{U}_2)=\alpha((\overline{Y}_1,W_1)+(\overline{Y}_2,W_2))=\alpha((\overline{\varnothing},\varnothing))=(\alpha(\overline{\varnothing}),\varnothing)=(\overline{\varnothing},\varnothing)'=\alpha(\widehat{U}_1)+'\alpha(\widehat{U}_2).$$

This means that α is an isomorphism of the groupoid $T(\mathcal{N}')$ and $\widehat{\mathrm{AGS}}(\mathcal{N}')$.

Similarly, it is proved that α is an isomorphism between the groupoid $(T(\mathcal{N}'),*)$ and $\widehat{\mathrm{MGS}}(\mathcal{N}')$. Indeed, in the above reasoning, the operation $(\cup)$ must be replaced by the operation $(\cap)$, and the operation $(+')$ by the operation $(*')$, which is an operation in the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N}')$. The theorem is proved. □

The problems below are of interest.

Problems 4.1. Describe all subgroupoids H of the groupoid $X(\mathcal{N})$ such that $H \cong Y(\mathcal{N}')$ for a suitable generalized subnet $\mathcal{N}'$ networks $\mathcal{N}$, where:

$$(a)\ X(\mathcal{N}) := \widehat{\mathrm{AGS}}(\mathcal{N}),\ Y(\mathcal{N}') := \widehat{\mathrm{AGS}}(\mathcal{N}'); \quad (b)\ X(\mathcal{N}) := \widehat{\mathrm{MGS}}(\mathcal{N}),\ Y(\mathcal{N}') := \widehat{\mathrm{MGS}}(\mathcal{N}').$$

Problems 4.2. Describe all subgroupoids H of the groupoid $X(\mathcal{N})$ such that H is not isomorphic $Y(\mathcal{N}')$ for any generalized subnet $\mathcal{N}'$ of the $\mathcal{N}$ network, where:

$$(a)\ X(\mathcal{N}) := \widehat{\mathrm{AGS}}(\mathcal{N}),\ Y(\mathcal{N}') := \widehat{\mathrm{AGS}}(\mathcal{N}'); \quad (b)\ X(\mathcal{N}) := \widehat{\mathrm{MGS}}(\mathcal{N}),\ Y(\mathcal{N}') := \widehat{\mathrm{MGS}}(\mathcal{N}').$$

Solutions to problems 4.1 (a) and 4.2 (a) will give a description of all subgroupoids of the groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ (similar, problems 4.1 (b) and 4.2 (b) give a description of all subgroupoids of the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$).

Problems 4.3. Give a description of all subgroupoids of the H groupoid $\widehat{\mathrm{AGS}}(\mathcal{N})$ such that H is isomorphic $\widehat{\mathrm{AGS}}(\mathcal{N}')$, where $\mathcal{N}'$ is the appropriate neural network. A similar question for the groupoid $\widehat{\mathrm{MGS}}(\mathcal{N})$.

In the above problems, a description of subgroupoids is understood as a description that provides information about what elements a subgroupoid with the desired property contains.

Problems 4.4. Describe all pairs of neural networks $(\mathcal{N}, \mathcal{K})$ for which isomorphism holds: a) $\widehat{\mathrm{AGS}}(\mathcal{N}) \cong \widehat{\mathrm{MGS}}(\mathcal{K})$; b) $\widehat{\mathrm{AGS}}(\mathcal{N}) \cong \widehat{\mathrm{AGS}}(\mathcal{K})$; c) $\widehat{\mathrm{MGS}}(\mathcal{N}) \cong \widehat{\mathrm{MGS}}(\mathcal{K})$.

Problems 4.5. Give an element-by-element description of the monoids of all endomorphisms of the groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$.

Problems 4.6. Give an element-by-element description of the sets of all congruences of groupoids $\widehat{\mathrm{AGS}}(\mathcal{N})$ and $\widehat{\mathrm{MGS}}(\mathcal{N})$.

Problems 4.5 and 4.6 are closely related (the connection between endomorphisms and congruences of universal algebras is well known; see, for example, the homomorphism theorem).

This work is supported by the Krasnoyarsk Mathematical Center and financed by the Ministry of Science and Higher Education of the Russian Federation (Agreement no. 075-02-2024-1429).

References

[1] A.V.Litavrin, Endomorphisms of finite commutative groupoids associated with multilayer neural networks of direct distribution, *Proceedings of the Institute of Mathematics and Mechanics. Ural Branch of the Russian Academy of Sciences*, **27**(2021), no. 1, 130–145. DOI: 10.21538/0134-4889-2021-27-1-130-145

[2] A.N.Gorban, Generalized approximation theorem and computational capabilities of neural networks, *Sib. magazine Comput. mat.*, **1**(1998), no. 1, 11–24.

[3] I.I.Slepovichev, Algebraic properties of abstract neural networks, *Izv. Sarat. un-ta. New ser. Ser. Mathematics. Mechanics. Computer Science*, **16**(2016), no. 1, 96–103. DOI: 10.18500/1816-9791-2016-16-1-96-103

[4] A.V.Sozykin, Review of methods for training deep neural networks,*Bulletin of the South Ural State University. Series "Computational Mathematics and Information Science"*, **6**(2017), no. 3, 28–59. DOI: 10.14529/cmse170303

[5] A.V.Litavrin, On endomorphisms of the additive monoid of subnets of a two-layer neural network, *The Bulletin of Irkutsk State University. Series Mathematics*, **39**(2022), 111–126. DOI: 10.26516/1997-7670.2022.39.111

[6] A.V.Litavrin, On some finite commutative groupoids associated with multilayer feed-forward neural networks, Proceedings of the XXI International Conference dedicated to the 85th anniversary of the birth of A. A. Karatsuba. Tula: TSPU im. L. N. Tolstoy, (2022), 106–109.

[7] A.V.Litavrin, T.V.Moiseenkova, About one groupoid associated with the composition of multilayer feedforward neural networks, *Zhurnal Srednevolzhskogo matematicheskogo obshchestva*, **26**(2024), no. 2, 111–122. DOI: 10.15507/2079-6900.26.202402.111-122

О некоторых коммутативных и идемпотентных конечных группоидах, связанных с подсетями многослойных нейронных сетей прямого распространения сигнала

Андрей В. Литаврин
Татьяна В. Моисеенкова
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. В работе изучаются коммутативные и идемпотентные конечные группоиды, которые связанны с подсетями многослойных нейронных сетей прямого распространения сигнала (далее, просто нейронные сети). Ранее вводилось понятие подсети нейронной сети. В данной работе вводится понятие обобщенной подсети нейронной сети. Это понятие обобщает ранее введенное понятие. Полученные группоиды получают название обобщенных подсетей заданной нейронной сети. Данные группоиды моделируют объединение и пересечение обобщенных подсетей некоторой нейронной сети. Выявлены условия, которым должна удовлетворять архитектура нейронной сети, чтобы аддитивный группоид обобщенных подсетей был ассоциативен. Получены условия, которым должна удовлетворять архитектура нейронной сети, чтобы мультипликативный группоид обобщенных подсетей был ассоциативен. Изучаются подгруппоиды построенных группоидов.

Ключевые слова: группоид, многослойная нейронная сеть прямого распространения сигнала, подсеть многослойной нейронной сети прямого распространения сигнала, аддитивный группоид обобщенных подсетей, мультипликативный группоид обобщенных подсетей, обобщенная подсеть.

EDN: JZZUNK
УДК 539.3

Solving Cauchy Problem for Elasticity Equations in a Plane Dynamic Case

Sergei I. Senashov*
Irina L. Savostyanova†
Reshetnev Siberian State University of Science and Technology
Krasnoyarsk, Russian Federation
Olga N. Cherepanova‡
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 05.09.2024, received in revised form 24.10.2024, accepted 04.11.2024

Abstract. Equations of elasticity in a plane dynamic case are considered in this paper. The system of equations is replaced by system of first-order differential equations with the same solution. The solution-equivalent system is group fibration of the original system of equations. It is a combination of the resolving and automorphic systems. Special classes of conservation laws are found for the resolving system of equations. These laws allow one to find the solution of the original equations in the form of surface integrals over the boundary of an elastic body.

Keywords: equations of elasticity in a plane dynamic case, Cauchy problem, conservation laws, exact solutions.

Citation: S.I. Senashov, I.L. Savostyanova, O.N. Cherepanova, Solving Cauchy Problem for Elasticity Equations in a Plane Dynamic Case, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 71–80. EDN: JZZUNK.

Introduction

Equations of linear elasticity theory were presented in the works of A.Cauchy, L. Navier, B. Saint–Venant and others as early as in the 19 century. Since then, attempts have been made to build solutions of the initial and boundary value problems. General solutions for equations of elasticity theory in a dynamic were built by G. Lame, P. F. Papkovich, H. Neuber, M. Yakovak, N. I. Ostrosablin and some others [1–3]. But according to the words of S. L. Sobolev " . . . the knowledge of general solutions, with rare exception, gives nothing for solving important particular problems, . . . , because we get, while solving these particular problems, a system of so complex functional relations for arbitrary functions that their finding is practically impossible [4]". To solve the elasticity theory problems a greate variety of contemporary mathematical methods are used. Thus, methods of group analysis of differential equations were used [5–8 and the references therein]. The theory of symmetries allowes one to build vast classes of invariant and partially-invariant solutions which describe stress-strain state of elastic medium.
Symmetries, by virtue of their locality, are not appropriate for solving initial and boundary value

*sen@mail.sibsau.ru
†ruppa@inbox.ru https://orcid.org/0000-0002-9675-7109
‡cheronik@mail.ru

problems. Here, conservation laws are more suitable for solving boundary value problems [9, 10 and the references therein]. In fact, conservation laws were used for solving linear equations by B. Riemann and V. Volterra [11]. It is known [12] that equations of elasticity theory can be presented with the use of group fibration in the form of a combination of two solution-equivalent systems of first-order differential equations: resolving system and automorphic system. This fact turned out to be very useful for constructing conservation laws and solving Cauchy problems with their use.
In this article the conservation laws are built for the resolving system of differential equations of elasticity theory which gave an opportunity to solve Cauchy problem for this system in the form of surface integrals over the boundary of an elastic body. Further, Cauchy problem for the automorphic system is solved. This allows one to build the solution of the initial problem for the equations of elasticity theory in a dynamic case.

1. Preliminaries

Let us consider the equations of elasticity in a plane case

$$\begin{aligned} w^1_{tt} &= (\lambda+2\mu)w^1_{xx} + \mu w^1_{yy} + (\lambda+\mu)w^2_{xy},\\ w^2_{tt} &= (\lambda+2\mu)w^2_{yy} + \mu w^1_{xx} + (\lambda+\mu)w^1_{xy}, \end{aligned} \tag{1}$$

where λ, μ are Lame constants, w^1, w^2 are components of displacement vector, density is equal to one. On the plane $t = 0$, the Cauchy problem is set

$$\begin{aligned} &w^1|_{t=9} = f^1(x,y), w^2|_{t=9} = f^2(x,y).\\ &w^1_t|_{t=9} = g^1(x,y), w^2_t|_{t=9} = g^2(x,y). \end{aligned} \tag{2}$$

If functions f^i, g^i are continuous together with their derivatives on the plane $t = 0$ then all derivatives of functions w^1, w^2 in any direction are known on this plane. It is known that system of equations (1) is of hyperbolic type and it has characteristic surfaces defined as $\omega(t,x,y) = 0$ which satisfy the following equation [10]

$$[(\lambda+2\mu)(\omega_x^2+\omega_y^2) - \omega_t^2][\mu(\omega_x^2+\omega_y^2) - \omega_t^2] = 0. \tag{3}$$

It is known [4, 5] that system of equations (1) allows a group of point symmetries generated by operators

$$\begin{aligned} &X_1 = \partial_x,\ \ X_2 = \partial_y,\ \ X_0 = \partial_t,\\ &Z = y\partial_x - x\partial_y + w^2\partial_{w^1} - w^1\partial_{w^2},\\ &P_0 = w^1\partial_{w^1} + w^2\partial_{w^2},\ \ P_w = h^1\partial_{w^1} + h^2\partial_{w^2},\ \ R = x\partial_x + y\partial_y + t\partial_t, \end{aligned} \tag{4}$$

where h^1, h^2 — arbitrary solution of equations (1). The presence of operator $P_w = h^1\partial_{w^1} + h^2\partial_{w^2}$ allows one to perform group fibration of system of equations (1) [4, 11], that is, to present it in the form of automorphic and resolving systems of equations. Let us consider operator $P_w = h_x\partial_{w^1} + h_y\partial_{w^2}$, where h is arbitrary harmonic function. Invariants of operator P_w are t, x, y.
Let us extend operator P_w on the first-order derivatives [4]

$$\underset{1}{p_w} = h\partial_{w^1} + h\partial_{w^2} + h_x(\partial_{w^1_x} - \partial_{w^2_y}) + h_y(\partial_{w^1_y} + \partial_{w^2_x}).$$

Differential invariants of the extended operator are

$$w^1_t,\ w^2_t,\ w^1_x + w^2_y,\ w^2_x - w^1_y.$$

Assigning differential invariants to be functions of invariants, one can obtain the automorphic system

$$w^1_t = u(t,x,y),\ \ w^2_t = v(t,x,y),\ \ \theta(t,x,y) = w^1_x + w^2_y,\ \ \omega(t,x,y) = w^2_x - w^1_y. \tag{5}$$

Conditions of compatibility of equations (5) lead to the resolving system

$$u_t = (\lambda+2\mu)\theta_x - \mu\omega_y,\ \ v_t = (\lambda+2\mu)\theta_y + \mu\omega_x,\ \ \theta_t = u_x + v_y,\ \ \omega_t = v_x - u_y. \tag{6}$$

Solution of Lame system of equations (1) is equivalent to solution of systems (5), (6) [5, 6]. Using initial conditions for equations (1), it is not difficult to obtain initial conditions for the functions included in equations (5) and (6):

$$\theta|_{t=0} = \partial_x f^1 + \partial_y f^2,\ \ \omega|_{t=0} = \partial_x f^2 - \partial_y f^1,\ \ u|_{t=0} = g^1,\ \ v|_{t=0} = g^2. \tag{7}$$

2. Problem formulation

Let us find the conservation laws for the resolving system of equations. This allows one to solve Cauchy problem (7) for equations (6). Further on, using (5), one can solve Cauchy problem (2) for equations (1).

3. Conservation laws for resolving system

Let us consider system of equations (6) in the form

$$\begin{aligned} &F_1 = u_t - (\lambda+2\mu)\theta_x + \mu\omega_y = 0,\ \ F_2 = v_t - (\lambda+2\mu)\theta_y - \mu\omega_x = 0,\\ &F_3 = \theta_t - u_x - v_y = 0,\ \ F_4 = \omega_t - v_x + u_y = 0. \end{aligned} \tag{8}$$

Definition. Expression of the form

$$A_t + B_x + C_y = \sum_{i=1}^{4} \rho^i F_i \tag{9}$$

is called the conservation law for system of equations (8). Here ρ^i are some linear differential operators that are simultaneously not identically zero. Vector (A, B, C) is called conserved current for conservation law (9).
More general definitions of conservation laws can be found in [8, 9 and the references therein].
Let us assume that conserved current is written as

$$\begin{aligned} A &= \alpha^1 u + \beta^1 v + \gamma^1\theta + \delta^1\omega,\\ B &= \alpha^2 u + \beta^2 v + \gamma^2\theta + \delta^2\omega,\\ C &= \alpha^3 u + \beta^3 v + \gamma^3\theta + \delta^3\omega, \end{aligned} \tag{10}$$

where $\alpha^i, \beta^i, \gamma^i, \delta^i$ are smooth functions that depend only on t, x, y.

Note. System of equations (8) also has other conservation laws by virtue of linearity. However, for our purposes it is sufficient to have conservation laws with conserved current in form (10).
Let us substitute (10) into (9). Then a first-degree polynomial with respect to derivatives $u_t, u_x, \ldots, \omega_y$ and required functions u, v, θ, ω is obtained. Setting coefficients at these variables equal to zero,one can obtain

$$\begin{aligned}&\alpha^1 = \rho^1,\ \ \alpha^2 = -\rho^2,\ \ \alpha^3 = -\rho^4,\ \ \beta^1 = \rho^2,\ \ \beta^2 = -\rho^4,\ \ \beta^3 = -\rho^3,\\ &\gamma^1 = \rho^3,\ \ \gamma^2 = -(\lambda+2\mu)\rho^1,\ \ \gamma^3 = -(\lambda+2\mu)\rho^2,\ \ \delta^1 = \rho^4,\ \ \delta^2 = -\mu\rho^2,\ \ \delta^3 = -\mu\rho^1.\end{aligned} \tag{11}$$

$$\begin{aligned}&\alpha^1_t - \gamma^1_x + \delta^1_y = 0, \beta^1_t - \delta^1_x - \gamma^1_y = 0,\\ &\gamma^1_t - (\lambda+2\mu)\alpha^1_x - (\lambda+2\mu)\beta^1_y = 0,\\ &\delta^1_t - \mu\beta^1_x + \mu\alpha^1_y = 0.\end{aligned} \tag{12}$$

It follows from (10)—(12) that conserved current is written as

$$\begin{aligned}A &= \alpha^1 u + \beta^1 v + \gamma^1\theta + \delta^1\omega,\\ B &= -\gamma^1 u - \delta^1 v - (\lambda+2\mu)\alpha^1\theta - \mu\beta^1\omega,\\ C &= \delta^1 u - \gamma^1 v - (\lambda+2\mu)\beta^1\theta + \mu\alpha^1\omega.\end{aligned} \tag{13}$$

It follows from (12) that (γ^1, δ^1) is an arbitrary solution of equations of elasticity (1).
Let us find the solution of equations (1) in the form of Lame

$$\gamma^1 = \Phi_x + \Psi_y,\ \ \delta^1 = \Phi_y - \Psi_x, \tag{14}$$

where Φ, Ψ are arbitrary solutions of equations

$$(\lambda+2\mu)(\Phi_{xx} + \Phi_{yy}) - \Phi_{tt} = 0, \tag{15}$$

$$\mu(\Psi_{xx} + \psi_{yy}) - \psi_{tt} = 0. \tag{16}$$

First, let us find the solution of equations (1) in the form

$$\gamma^1 = \Phi_x,\ \ \delta^1 = \Phi_y, \tag{17}$$

Then it follows from (12) that

$$\alpha^1_t = 0, \quad \beta^1_t = \Phi_{tt}/(\lambda+2\mu).$$

Further on, it is assumed that

$$\alpha^1 = 0, \quad \beta^1 = \Phi_t/(\lambda+2\mu). \tag{18}$$

Let us find the solution of equation (15) in the form of Kirchhoff

$$\Phi = \frac{1}{r}(G_1(t - t_0 + (\sqrt{\lambda+2\mu})^{-1}\, r) + G_2(t - t_0 - (\sqrt{\lambda+2\mu})^{-1}\, r),$$

where $r = \sqrt{(x-x_0)^2 + (y-y_0)^2}$, (t_0, x_0, y_0) is some point such that $t_0 \neq 0$. Let us assume that

$$G_1 = (t - t_0 + (\sqrt{\lambda+2\mu})^{-1}\, r)^{1+n},\ \ G_2 = -(t - t_0 - (\sqrt{\lambda+2\mu})^{-1}\, r)^{1+n}, \tag{19}$$

where $n \in R, \quad n > 1$.
Then

$$\begin{aligned}
\gamma^1 &= -\frac{x-x_0}{r^3}((t-t_0+(\sqrt{\lambda+2\mu})^{-1}r)^{1+n} - (t-t_0-(\sqrt{\lambda+2\mu})^{-1}r)^{1+n})- \\
&-\frac{(1+n)(x-x_0)(\sqrt{\lambda+2\mu})^{-1}}{r^2}((t-t_0+(\sqrt{\lambda+2\mu})^{-1}r)^{n} + (t-t_0-(\sqrt{\lambda+2\mu})^{-1}r)^{n}), \\
\delta^1 &= -\frac{y-y_0}{r^3}((t-t_0+(\sqrt{\lambda+2\mu})^{-1}r)^{1+n} - (t-t_0-(\sqrt{\lambda+2\mu})^{-1}r)^{1+n})- \\
&-\frac{(1+n)(y-y_0)(\sqrt{\lambda+2\mu})^{-1}}{r^2}((t-t_0+(\sqrt{\lambda+2\mu})^{-1}r)^{n} + (t-t_0-(\sqrt{\lambda+2\mu})^{-1}r)^{n}), \\
\beta^1 &= \frac{(1+n)}{r(\lambda+2\mu)}((t-t_0+(\sqrt{\lambda+2\mu})^{-1}r)^{n} + (t-t_0-(\sqrt{\lambda+2\mu})^{-1}r)^{n}), \quad \alpha^1 = 0.
\end{aligned} \tag{20}$$

Now let us find the solution of equations (1) in the form

$$\gamma^1 = \Psi_y, \quad \delta^1 = -\Psi_x. \tag{21}$$

Then from (12) it follows

$$\beta^1_t = 0, \quad \alpha^1_t = \Psi_{tt}/\mu.$$

Further on, it is assumed that

$$\beta^1 = 0, \quad \alpha^1 = \Psi_t/\mu. \tag{22}$$

Let us find the solution of equation (16) in the form of Kirchhoff

$$\Phi = \frac{1}{r}(G_3(t-t_0+(\sqrt{\mu})^{-1}\, r) + G_4(t-t_0-(\sqrt{\mu}\,)^{-1}r).$$

Let us assume that

$$G_3 = (t-t_0+(\sqrt{\mu})^{-1}\, r)^{1+m}, \;\; G_4 = -(t-t_0-(\sqrt{\mu}\,)^{-1}r)^{1+m}, \tag{23}$$

where $m \in R$.
Then

$$\begin{aligned}
\gamma^1 &= -\frac{x-x_0}{r^3}((t-t_0+(\sqrt{\mu})^{-1}r)^{1+m} - (t-t_0-(\sqrt{\mu})^{-1}r)^{1+m})- \\
&-\frac{(1+m)(x-x_0)(\sqrt{\mu})^{-1}}{r^2}((t-t_0+(\sqrt{\mu})^{-1}r)^{m} + (t-t_0-(\sqrt{\mu})^{-1}r)^{m}), \\
\delta^1 &= -\frac{y-y_0}{r^3}((t-t_0+(\sqrt{\mu})^{-1}r)^{1+m} - (t-t_0-(\sqrt{\mu})^{-1}r)^{1+m})- \\
&-\frac{(1+m)(y-y_0)\sqrt{\mu}}{r^2}((t-t_0+\sqrt{\mu}r)^{m} + (t-t_0-\sqrt{\mu}r)^{m}), \\
\alpha^1 &= \frac{(1+m)}{r\mu}((t-t_0+(\sqrt{\mu})^{-1}r)^{m} + (t-t_0-(\sqrt{\mu})^{-1}r)^{m}), \quad \beta^1 = 0.
\end{aligned} \tag{24}$$

4. Solving Cauchy problem for resolving system of equations

Characteristic cones with the origin at the point (t_0, x_0, y_0) are shown in Fig. 1. The lateral surface of the outer cone is given by the equation

$$S_1: \; (\lambda+2\mu)(t-t_o)^2 - (x-x_0)^2 - (y-y_0)^2 = 0, \tag{25}$$

and the lateral surface of the inner cone is given by the equation

$$S_2: \ \mu(t-t_o)^2 - (x-x_0)^2 - (y-y_0)^2 = 0. \tag{26}$$

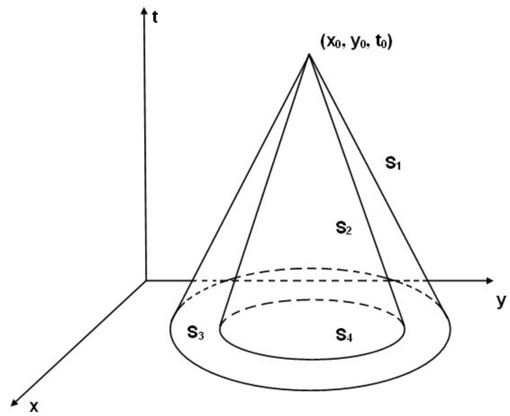


Fig. 1. Characteristic cones

Intersections of cones (25) and (26) with the plane $t = 0$ are circles S_3, S_4. Initial conditions on functions u, v, θ, ω are given inside these circles.
Let us consider domain V_1 bounded by surface S_1 and by plane $t = 0$. Then it follows from (9) that

$$\iiint\limits_{V_1} (A_t + B_x + C_y)dxdydt = 0. \tag{27}$$

Let us consider cylinder T_ε of radius $(x-x_0)^2 + (y-y_0)^2 = \varepsilon^2$ inside the outer cone as shown in Fig. 2.

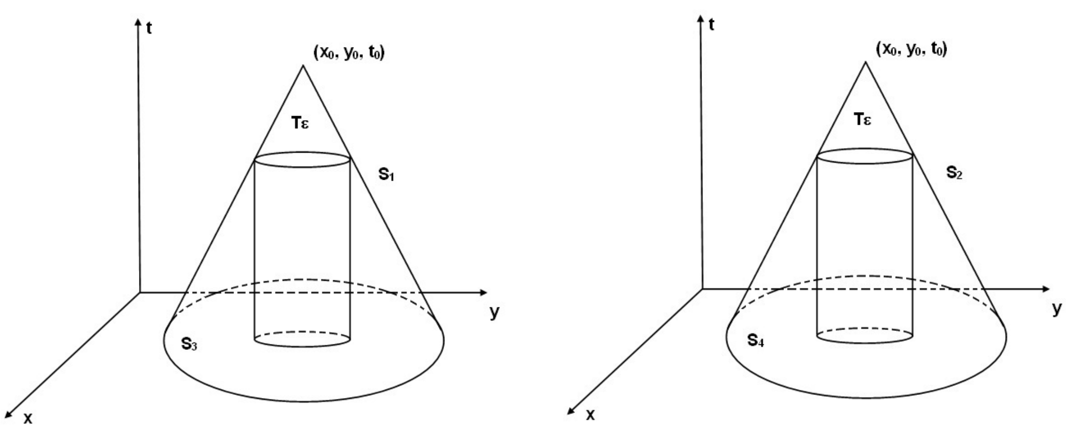


Fig. 2. Solving the Cauchy problem to find $\theta(x_0, y_0, t_0)$

Functions $\alpha^1, \beta^1, \gamma^1, \delta^1$ have no peculiarities inside the domain bounded by surface S_1, by cylindrical surface T_ε and by plane $t = 0$. Using the Gauss–Ostrogradskiy formula, one can obtain from (27) that

$$\begin{aligned}&\iiint\limits_{V_1\setminus T_\varepsilon}(A_t+B_x+C_y)dxdydt=\iint\limits_{S_1}Adxdy+Bdydt+Cdtdx+\\&+\iint\limits_{T_\varepsilon}Adxdy+Bdydt+Cdtdx+\iint\limits_{S_3}Adxdy+Bdydt+Cdtdx=0.\end{aligned}\tag{28}$$

By virtue of choosing function Φ the integral $\iint\limits_{S_1}Adxdy+Bdydt+Cdtdx=0$. It is not difficult to see that the integral $\iint\limits_{S_3}Adxdy+Bdydt+Cdtdx$ has no peculiarities. That is why, it is necessary to calculate only the integral

$$\iint\limits_{T_\varepsilon}Bdydt+Cdtdx\tag{29}$$

on the assumption that ε is small. Assume that $x-x_0=\varepsilon\cos\phi,\ y-y_0=\varepsilon\sin\phi$. Let us substitute these expressions into (29) and obtain

$$\begin{aligned}&\iint\limits_{T_\varepsilon}Bdydt+Cdtdx=\\&=\int\limits_0^{t_0}\varepsilon dt\int\limits_0^{2\pi}\Big((-\gamma^1u-\delta^1v-(\lambda+2\mu)\alpha^1\theta-\mu\beta^1\omega)\cos\phi-(\delta^1u-\gamma^1v-(\lambda+2\mu)\beta^1\theta+\mu\alpha^1\omega)\sin\phi\Big)d\phi.\end{aligned}$$

Since

$$\begin{aligned}\gamma^1&=-\frac{2\cos\phi}{\varepsilon\sqrt{\lambda+2\mu}}(2n+1)(t-t_0)^n+o(\varepsilon),\\\delta^1&=-\frac{2\sin\phi}{\varepsilon\sqrt{\lambda+2\mu}}(2n+1)(t-t_0)^n+o(\varepsilon),\\\alpha^1&=\frac{2}{\sqrt{\lambda+2\mu}}(n+1)(t-t_0)^n+o(\varepsilon),\beta^1=0,\end{aligned}$$

it follows that

$$\begin{aligned}&\iint\limits_{T_\varepsilon}Bdydt+Cdtdx=\\&=-(\lambda+2\mu)\int_0^{t_0}\left(\int_0^{2\pi}\theta(\alpha^1\cos\phi+\beta^1\sin\phi)d\phi-\mu\int_0^{2\pi}\omega(\beta^1\cos\phi-\alpha^1\sin\phi)d\phi\right)dt=\\&=2\pi\sqrt{\lambda+2\mu}(2n+1)\int_0^{t_0}(t-t_0)^n\theta(x_0,y_0,t)dt.\end{aligned}$$

The last expression is obtained with $\varepsilon\to0$.
Finally, it follows from (28) and (29) that

$$2\pi\sqrt{\lambda+2\mu}(2n+1)\int_0^{t_0}(t-t_0)^n\theta(x_0,y_0,t)dt=\iint\limits_{S_3}Adxdy.$$

Differentiating the last expression with respect to t_0, one can obtain that

$$\theta(x_0, y_0, t_0) = \frac{1}{2\pi(n+1)\sqrt{\lambda+2\mu}} \frac{\partial}{\partial t_0} \iint\limits_{S_3} Adxdy, \tag{30}$$

where $A = \alpha^1 u + \beta^1 v + \gamma^1 \theta + \delta^1 \omega$,

$$\gamma^1 = -\frac{x-x_0}{r^3}\left(\left(\frac{r}{\sqrt{\lambda+2\mu}} - t_0\right)^{1+n} - \left(-t_0 - \frac{r}{\sqrt{\lambda+2\mu}}\right)^{1+n}\right) -$$
$$-\frac{(x-x_0)(1+n)}{r^2\sqrt{\lambda+2\mu}}\left(\left(\frac{r}{\sqrt{\lambda+2\mu}} - t_0\right)^{n} + \left(-t_0 - \frac{r}{\sqrt{\lambda+2\mu}}\right)^{n}\right),$$

$$\delta^1 = -\frac{y-y_0}{r^3}\left(\left(\frac{r}{\sqrt{\lambda+2\mu}} - t_0\right)^{1+n} - \left(-t_0 - \frac{r}{\sqrt{\lambda+2\mu}}\right)^{1+n}\right) -$$
$$-\frac{(y-y_0)(1+n)}{r^2\sqrt{\lambda+2\mu}}\left(\left(\frac{r}{\sqrt{\lambda+2\mu}} - t_0\right)^{n} + \left(-t_0 - \frac{r}{\sqrt{\lambda+2\mu}}\right)^{n}\right),$$

$$\beta^1 = \frac{(1+n)}{r(\gamma+2\mu)}\left(\left(\frac{r}{\sqrt{\lambda+2\mu}} - t_0\right)^{n} - \left(-t_0 - \frac{r}{\sqrt{\lambda+2\mu}}\right)^{n}\right), \quad \alpha^1 = 0.$$

Now let us perform the same procedure for the inner cone but for solutions (20), (21) and obtain

$$\omega(x_0, y_0, t_0) = \frac{1}{2\pi(n+1)\sqrt{\mu}} \frac{\partial}{\partial t_0} \iint\limits_{S_3} Adxdy, \tag{31}$$

where $A = \alpha^1 u + \beta^1 v + \gamma^1 \theta + \delta^1 \omega$,

$$\alpha^1 = -\frac{y-y_0}{r^3}\left(\left(\frac{r}{\sqrt{\mu}} - t_0\right)^{1+m} - \left(-t_0 - \frac{r}{\sqrt{\mu}}\right)^{1+m}\right) -$$
$$-\frac{(y-y_0)(1+m)}{r^2\sqrt{\mu}}\left(\left(\frac{r}{\sqrt{\mu}} - t_0\right)^{m} + \left(-t_0 - \frac{r}{\sqrt{\mu}}\right)^{m}),$$

$$\beta^1 = -\frac{x-x_0}{r^3}\left(\left(\frac{r}{\sqrt{\mu}} - t_0\right)^{1+m} - \left(-t_0 - \frac{r}{\sqrt{\mu}}\right)^{1+m}\right) -$$
$$-\frac{(y-y_0)(1+m)}{r^2\sqrt{\mu}}\left(\left(\frac{r}{\sqrt{\mu}} - t_0\right)^{m} + \left(-t_0 - \frac{r}{\sqrt{\mu}}\right)^{m}\right),$$

$$\gamma^1 = \frac{(1+m)}{r}\left(\left(\frac{r}{\sqrt{\mu}} - t_0\right)^{m} - \left(-t_0 - \frac{r}{\sqrt{\mu}}\right)^{m}\right), \quad \delta^1 = 0.$$

Now, taking into account (30)–(31) and initial conditions (2)

$$u_t = (\lambda+2\mu)\theta_x - \mu\omega_y, \;\; v_t = (\lambda+2\mu)\theta_y + \mu\omega_x$$

, one can obtain from (6) that

$$w_t^1 = u = \int_0^t ((\gamma+2\mu)\theta_x - \mu\omega_y)dt + g^1(x,y), \;\; w_t^2 = v = \int_0^t ((\gamma+2\mu)\theta_y + \mu\omega_x)dt + g^2(x,y).$$

Taking into account (5) and initial conditions (2), one can finally find that

$$\begin{aligned} w^1 &= \int_0^t udt = \int_o^t \left(\int_0^t ((\gamma+2\mu)\theta_x - \mu\omega_y)dt\right)dt + g^1(x,y)t + f^1(x,y), \\ w^2 &= \int_0^t vdt = \int_o^t \left(\int_0^t ((\lambda+2\mu)\theta_y + \mu\omega_x)dt\right)dt + g^2(x,y)t + f^2(x,y). \end{aligned} \tag{32}$$

Relations (32) provide the solution of Cauchy problem for system of equations (1).

Note. The method of solving Cauchy problem stated in this paper can be used with some modifications to solve three-dimensional dynamic problems for equations of elasticity. This will be performed in the following works.

This paper was carried out by the team of the scientific laboratory "Smart Materials and Structures" within the state assignment of the Ministry of Science and Higher Education of the Russian Federation for the implementation of the project "Development of multifunctional smart materials and structures based on modified polymer composite materials capable to function in extreme conditions" (Project no. FEFE-2020-0015).

References

[1] V.Novatsky, Theory of elasticity, Novosibirsk, Nauka, 1983 (in Russian).

[2] N.I.Ostrosablin, Symmetry operators and general solutions of equations of linear theory of elasticity, *Applied Mechanics and Technical Physics*, **36**(1995), no. 5, 98–104.

[3] N.I Ostrosablin, General solutions and reduction of systems of equations of linear theory of elasticity to a diagonal form, *Applied Mechanics and Technical Physics*, **34**(1993), no. 5, 112–122.

[4] S.L.Sobolev, Equations of mathematical physics, Moscow: State Publishing House of Technical and Theoretical Literature, 1956 (in Russian).

[5] L.V.Ovsyannikov, Group analysis of differential equations, Moscow, Nauka, 1978 (in Russian).

[6] B.D.Annin, V.O.Bytev, S.I.Senashov, Group properties of elasticity and plasticity equations, Novosibirsk, Nauka, 1983 (in Russian).

[7] S.I.Senashov, I.L.Savostyanova, On elastic torsion around three axes, *Siberian Journal of Industrial Mathematics*, **24**(2021), no. 1, 120–125 (in Russian).
DOI: 10.33048/sibjim.2021.24.109

[8] B.D.Annin, V.D.Bondar, S.I.Senashov, Determination of elastic and plastic deformation regions in the problem of uniaxial tension of a plate weakened by holes, *Siberian Journal of Industrial Mathematics*, **23**(2020), no. 1, 11–16 (in Russian).

[9] O.V.Gomonova, S.I.Senashov, Group analysis and exact solutions of the equations of plane deformation of an incompressible nonlinear elastic body, *Applied Mechanics and Technical Physics*, **62**(2021), no. 1, 179–186 (in Russian).

[10] O.V.Gomonova, S.I.Senashov, O.N.Cherepanova, Group analysis of ideal plasticity equations, *Applied Mechanics and Technical Physics*, **62**(2021), no. 5, 208–216 (in Russian).

[11] V.I.Smirnov, Course of Higher Mathematics, Moscow, Nauka, 1981 (in Russian).

[12] V.Yu.Prudnikov, Yu.A.Chirkunov, Group bundle of Lame equations, *Solid State Mechanics*, **22**(2009), no. 3, 471–477 (in Russian).

Решение задачи Коши для уравнений упругости в плоском динамическом случае

Сергей И. Сенашов
Ирина Л. Савостьянова
Сибирский государственный университет науки и технологий имени академика М. Ф. Решетнева
Красноярск, Российская Федерация
Ольга Н. Черепанова
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. Рассмотрены уравнения упругости в плоском динамическом случае. Эта система заменена равносильной системой дифференциальных уравнений первого порядка. Равносильная система есть групповое расслоение исходной системы уравнений, она является объединением разрешающей и автоморфных систем. Для разрешающей системы уравнений найдены специальные классы законов сохранения, которые позволили найти решение исходных уравнений в виде поверхностных интегралов по границе упругого тела.

Ключевые слова: уравнения упругости в плоском динамическом случае, задача Коши, законы сохранения, точные решения

EDN: MLYWGB
УДК 621.396; 669.8

Model of Receiving Channels of an Adaptive Antenna Array to Assess the Impact of Differences in their Characteristics on the Efficiency of Interference Suppression

Valery N. Tyapkin[*]
Dmitry D. Dmitriev[†]
Siberian Federal University
Krasnoyarsk, Russian Federation
Pavel V. Shtro[‡]
Pavel V. Luferchik[§]
JSC Scientific-Production Enterprise Radiosvyaz
Krasnoyarsk, Russian Federation
Igor V. Tyapkin[¶]
Evgenii D. Mikhov[‖]
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 10.08.2024, received in revised form 15.09.2024, accepted 24.10.2024

Abstract. The article proposes a mathematical model that allows us to evaluate the impact of differences in the characteristics of reception channels on the quality of noise suppression in radio devices equipped with antenna arrays.The influence of differences in the bandwidth of receiving channels, settings of their central frequencies and electrical lengths was studied. The dependences of the losses of the average interference suppression coefficient on the dispersion of parameters of differences in the characteristics of receiving channels are presented.It is advisable to use the proposed model when justifying the requirements for the permissible difference in the characteristics of reception channels of designed radio devices.

Keywords: antenna array, interference suppression, difference in characteristics of receiving channels, interference suppression coefficient.

Citation: V.N. Tyapkin, D.D. Dmitriev, P.V. Shtro, P.V. Luferchik, I.V. Tyapkin, E.D. Mikhov, Model of Receiving Channels of an Adaptive Antenna Array to Assess the Impact of Differences in their Characteristics on the Efficiency of Interference Suppression, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 81–90. EDN: MLYWGB.

Spatial selection methods are currently considered the most effective methods of dealing with interference [1–4]. In this case, the maximum immunity to radio interference in the useful signal band is determined by the dynamic range of the radio path and the analog-to-digital converter.These elements must maintain linear operation at the maximum permissible radio

[*]vtyapkin@sfu-kras.ru
[†]ddmitriev@sfu-kras.ru
[‡]faust256@mail.ru
[§]luferchikp@gmail.com
[¶]ityapkin@sfu-kras.ru
[‖]emikhov@sfu-kras.ru

interference power. Only in this case will the signal and interference be successfully converted into digital form and it will be possible to use methods for optimal filtering of useful signals and spatial selection of interference.If the power of the interfering signal is too high, the radio path cannot operate in linear mode and the receiver is practically blocked. In this case, optimal filtering or frequency division of signals does not help in suppressing interference. Expanding the dynamic range of the radio path and ADC is one of the main conditions for creating noise-resistant radio equipment [5, 6]. The effectiveness of interference suppression by spatial selection methods is largely determined by the degree of interference correlation between the receiving channels of the adaptive antenna array. This explains the high requirements for the identity of frequency and phase characteristics of receiving channels, nonlinearity parameters of paths, accuracy of calculation of weight coefficients and other decorrelating factors. In this case, the interference suppression coefficient in the adaptive antenna array depends on the modulus of the interchannel interference correlation coefficient. The closer the correlation coefficient is to unity, the higher the interference suppression coefficient [7]. Many methods for improving the noise immunity of radio devices are aimed at equalizing the characteristics of receiving channels. This is the equalization of time delays between antenna elements, taking into account the geometry of the location of antenna elements and the wave front of received interference oscillations, and correction of the frequency characteristics of receiving channels [8, 9].In addition, to form the required shape of the radiation pattern in the antenna array, it is necessary to take into account all the delays that arise in the receiving paths, starting from the feeds to the beamforming device, with an accuracy of several degrees in the phase of the carrier frequency [10, 11]. In a number of practical cases, it is quite difficult to ensure the fulfillment of the listed conditions, which inevitably entails a decrease in the efficiency of the adaptive antenna array. It is required to evaluate the impact of differences in the characteristics of receiving channels on the efficiency of interference suppression. This assessment will make it possible to justify the requirements for the permissible difference in the characteristics of the reception channels of the designed radio devices.

1. Mathematical description of the model of receiving channels of an adaptive antenna array

The block diagram of the model of the receiving channels of the adaptive antenna array and the assessment of the differences in their characteristics is shown in Fig. 1

Here it is assumed that the N-dimensional vector of complex amplitudes of a mixture of interference and internal noise $\mathbf{y}(t) = \{y_i(t)\}_{i=1}^{N}$, processed during spatial filtering, is the result of transforming the components of the interference vector from the output of the adaptive antenna array $\mathbf{y}_{AAA}(t) = \{y_m^{(AAA)}(t)\}_{m=1}^{N}$ in N linear filters having different impulse characteristics $\nu_m(t), m \in 1, N$. This difference in impulse characteristics decorrelates the interference in the receiving channels, as a result of which the possible level of their compensation is reduced. Assessing the impact of differences in the impulse characteristics of linear filters on the achievable level of noise compensation is the goal of further analysis.

Vectors $\mathbf{y}(t)$ and $\mathbf{y}_{AAA}(t)$ are related to each other by equalities

$$y(t) = \int_{-\infty}^{\infty} \mathbf{D}(\tau)\mathbf{y}_{AAA}(t-\tau)dt, \tag{1}$$

where $\mathbf{D}(t) = diag\ \{\ \mathbf{v}_m(t)\}_{m=1}^{N}$ is the diagonal matrix of impulse characteristics of receiving

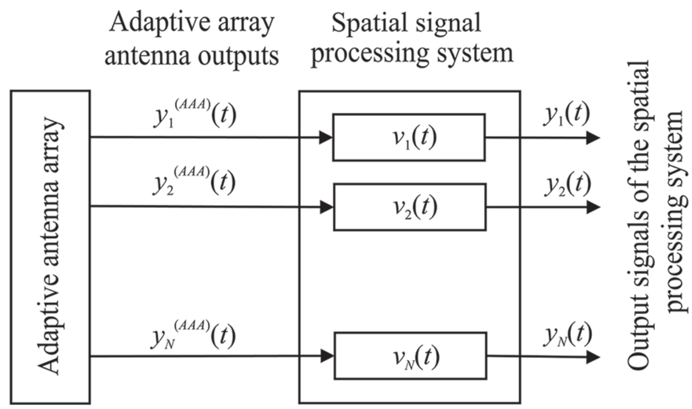


Fig. 1. Block diagram of the model of the receiving channels of the adaptive antenna array and assessment of the differences in their characteristics

channels. By integral of a vector we mean a vector of integrals of its elements. The correlation matrix of the vector $\mathbf{y}$(t), which determines the achievable level of interference compensation, in accordance with (1) is equal to:

$$\mathbf{\Phi} = \{\varphi_{pq}\}_{p,q=1}^{N} = \overline{\mathbf{y}(t)\mathbf{y}^*(t)} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \mathbf{D}(\tau)\mathbf{\Phi}_{AAA}(\tau,s)\mathbf{D}^*(s)d\tau ds \tag{2}$$

where

$$\mathbf{\Phi}_{AAA}(\tau,s) = \overline{\mathbf{y}_{AAA}(t-\tau)\ (\mathbf{y}_{AAA}(t-s))^*} - \tag{3}$$

correlation matrix of vector $\mathbf{y}_{AAA}(t)$ output signals of adaptive antenna array modules.

In the case under consideration, this vector corresponds to a mixture of Gaussian noise and stationary noise with a zero average value and a correlation matrix

$$\mathbf{\Phi}_{AAA}(\tau,s) = \mathbf{\Phi}_{AAA}\ \delta(\tau - s) \tag{4}$$

where $\delta(x)$ is the delta function. Under these conditions, the correlation matrix (2) has the form

$$\mathbf{\Phi} = \{\varphi_{pq}\}_{p,q=1}^{N} = \int_{-\infty}^{\infty} \mathbf{D}(s)\,\mathbf{\Phi}_{AAA}\,\mathbf{D}^*(s)\,ds, \tag{5}$$

$$\varphi_{pq} = \int_{-\infty}^{\infty} \nu_p(s)\,\varphi_{pq}^{(AAA)}\,\nu_q^*(s)\ ds = \varphi_{pq}^{(AAA)}\,a_{pq}, p, q \in 1, N, \tag{6}$$

$$\mathbf{A} = \{a_{pq}\}_{p,q=1}^{N} = \int_{-\infty}^{\infty} \mathbf{v}(t)\,\mathbf{v}^*(t)dt, a_{pq} = \int_{-\infty}^{\infty} \nu_p(t)\nu_q^*(t)dt. \tag{7}$$

It follows that each element of the correlation matrix is equal to the product of the corresponding elements of the correlation matrix (2) of the vector $\mathbf{y}_{AAA}(t)$ and the correlation matrix (6) of the vector $\nu(t) = \{\nu_m(t)\}_{m=1}^{N}$ of the impulse characteristics of the receiving channels. Therefore, the matrix $\mathbf{\Phi}$ is the Schur–Hadamard product of the matrices $\mathbf{\Phi}_{AAA}$ and $\mathbf{A}$, which is usually denoted as

$$\mathbf{\Phi} = \mathbf{\Phi}_{AAA} \otimes \mathbf{A}. \tag{8}$$

In the particular case of identical impulse characteristics $\nu_m(t) = \nu_0(t)$, $m \in \overline{1, N}$, when

$$\begin{aligned} &\nu(t) = \nu_0(t)\ \mathbf{e}, \\ &\mathbf{e}^* = [\ 1\ ,\ 1\ ,\ \ldots\ ,\ 1\ ], \\ &\mathbf{A} = c \cdot\ \mathbf{e}\mathbf{e}^*, \\ &c = \int_{-\infty}^{\infty} |\nu_0(t)|^2 dt, \end{aligned} \tag{9}$$

matrix (4) is proportional to matrix (3), due to which the achievable interference suppression coefficient remains the same as when using directly the vector of output signals $\mathbf{y}_{AAA}(t)$ of the adaptive antenna array modules. However, in real conditions, the impulse characteristics of receiving channels are not identical, and the loss of the interference suppression coefficient is determined by matrix (7). This matrix depends on the magnitude of the differences in the impulse characteristics of the receiving channels of the antenna array.

2. Estimation of the dependence of the interference suppression coefficient on the magnitude of the differences between the characteristics of the receiving channels of the antenna array

Quantitative estimates of the influence of differences in impulse characteristics on the noise suppression coefficient were carried out for the case of Gaussian impulse characteristics of the form

$$\nu_m(t) = \exp(-\pi \cdot F_m^2 \cdot (t - \tau_m)^2) \cdot \exp(j\ \cdot 2\pi \cdot\ (f_0 + \delta f_m) \cdot (t - \tau_m)),\ \ m \in \overline{1,\ N} \tag{10}$$

whose parameters are:

- $F_m = 1/T_m$ — the width of the frequency response (bandwidth) of the m-th filter at level $exp(-\pi/4) \approx 0.456$ from the maximum, inverse to the time length T_m of its impulse characteristics at the same level;

- τ_m — delay associated with the "electrical length" of the m-th reception path;

- δf_m — shift of the center frequency of the m-th filter from the value f_0.

An additional parameter of the m-th filter in the general case is also its gain c_m. However, it does not affect the desired level of achievable interference suppression coefficient K_{IS}, which, when protecting the first (main) channel by a system of $N_k = N - 1$ auxiliary (compensation) channels, is equal to

$$K_{IS} = \varphi_{11}\omega_{11}, \tag{11}$$

where ω_{11} is the first diagonal element of the matrix inverse to the correlation matrix (2), (8)

$$\mathbf{\Psi} = \{\omega_{pq}\}_{p,q=1}^{N} = \mathbf{\Phi}^{-1}. \tag{12}$$

Indeed, let the impulse characteristics of the m-th filter be equal to $\tilde{\nu}_m(t) = c_m \cdot \nu_m(t)$, then the corresponding impulse characteristics vector is equal to $\tilde{\nu}(t) = \mathbf{C} \cdot \nu(t)$, where $\mathbf{C} = diag\{c_m\}_{m=1}^{N}$ is the real diagonal gain matrix. In this case, the matrix $\tilde{\mathbf{A}}$ a is equal to $\tilde{\mathbf{A}} = \int\limits_{-\infty}^{\infty} \tilde{\nu}(t)\tilde{\nu}^*(t)dt =$

$= \mathbf{C}\cdot\mathbf{A}\cdot\mathbf{C}$, so the matrices $\tilde{\mathbf{\Phi}}$ and $\tilde{\mathbf{\Psi}}$ are respectively equal to $\tilde{\mathbf{\Phi}} = \{\tilde{\varphi}_{pq}\}_{p,q}^{N} = \mathbf{\Phi}_{AAA}\otimes\tilde{\mathbf{A}} = \mathbf{C}\cdot\mathbf{\Phi}\cdot\mathbf{C}$ and $\tilde{\mathbf{\Psi}} = \{\tilde{\omega}_{pq}\} = \tilde{\mathbf{\Phi}}^{-1} = \mathbf{C}^{-1}\cdot\mathbf{\Psi}\cdot\mathbf{C}^{-1}$.

The corresponding value of the achievable interference suppression coefficient (11) in this case is equal to $\tilde{K}_{IS} = \tilde{\varphi}_{11}\cdot\tilde{\omega}_{11} = c_1\cdot\tilde{\varphi}_{11}\cdot c_1\cdot c_1^{-1}\cdot\tilde{\omega}_{11}\cdot c_1^{-1} = \tilde{\varphi}_{11}\cdot\tilde{\omega}_{11} = K_{IS}$ and, therefore, coincides with the value obtained without taking into account different gain factors. Therefore, in what follows, impulse characteristics of the form (10) are used without unimportant additional amplification parameters.

Under these conditions, the elements of matrix (7) are equal

$$\begin{aligned}
&a_{pq} = \int_{-\infty}^{\infty} g_{pq}(t)\ dt,\\
&g_{pq}(t) = \nu_p(t)\cdot\nu_q^*(t) = \exp(-s_{pq}(t))\cdot\exp(j\cdot 2\cdot\pi\cdot\varphi_{pq}(t)),\\
&s_{pq}(t) = \pi\cdot(F_p^2\cdot(t-\tau_p)^2 + F_q^2\cdot(t-\tau_q)^2) =\\
&\qquad = \pi\cdot((F_p^2+F_q^2)\cdot(t-b)^2 + \frac{F_p^2\cdot F_q^2}{F_p^2+F_q^2}\cdot(\tau_p-\tau_q)^2),\\
&b = \frac{F_q^2\cdot\tau_q + F_p^2\cdot\tau_p}{F_q^2+F_p^2},\\
&\varphi_{pq}(t) = (\delta f_p - \delta f_q)\cdot t + \delta f_q\cdot\tau_p - \delta f_p\cdot\tau_q.
\end{aligned}\tag{13}$$

Using the well-known integral

$$\int_{-\infty}^{\infty}\exp(-a\cdot x^2)\exp(-j\cdot\beta\cdot x)dx = \int_{-\infty}^{\infty}\exp(-a\cdot x^2)\cdot\cos(\beta\cdot x)dx = \sqrt{\pi/a}\cdot\exp\left(-\frac{\beta^2}{4a}\right)$$

the elements of matrix (7) can be written in the form

$$\begin{aligned}
a_{pq} = &\frac{c}{\sqrt{F_p^2+F_q^2}}\cdot\exp\left(-\pi\cdot\frac{\nu_q^2\cdot\nu_p^2(\chi_p-\chi_q)^2 + (\mu_p-\mu_q)^2}{\nu_p^2+\nu_q^2}\right)\times\\
&\times\exp\left(-j\cdot 2\pi\cdot\frac{(\nu_p^2\cdot\mu_q+\nu_q^2\cdot\mu_p)\cdot(\chi_p-\chi_q)}{\nu_p^2+\nu_q^2}\right),\\
&p,\ q\in 1,\ N,
\end{aligned}\tag{14}$$

where $\nu_p = F_p/F_0 = 1+e_p$, $\mu_p = \delta f_p/F_0$, $\chi_p = \tau_p/T_0$, $p\in 1,\ N$ are the relative values of the corresponding filter parameters, c is a constant that does not affect the level of noise suppression. It is convenient to choose it so that, with the same filter parameters of all channels with nominal parameters, when $F_q^2 = F_p^2 = F_0^2$, $\nu_p = 1$, $\mu_p = \mu_q = 0$, $\tau_p = \tau_q$, $p, q\in 1, N$ the value of a = 1. This is done at the value $c = \sqrt{2\cdot F_0}$, at which

$$\begin{aligned}
a_{pq} = &\frac{\sqrt{2}}{\sqrt{\nu_p^2+\nu_q^2}}\cdot\exp\left(-\pi\cdot\frac{\nu_q^2\cdot\nu_p^2(\chi_p-\chi_q)^2 + (\mu_p-\mu_q)^2}{\nu_p^2+\nu_q^2}\right)\times\\
&\times\exp\left(-j\cdot 2\pi\cdot\frac{(\nu_p^2\cdot\mu_q+\nu_q^2\cdot\mu_p)\cdot(\chi_p-\chi_q)}{\nu_p^2+\nu_q^2}\right),\\
&p,\ q\in 1,\ N
\end{aligned}\tag{15}$$

The last formula, together with (8), (11), (12), allows us to obtain quantitative values of the interference suppression coefficient for arbitrary values of the parameters of the impulse characteristics of the filters (Fig. 1) of the receiving channels of the antenna array. In the general case, these parameters are random, so the values of the corresponding suppression coefficients (11) obtained on their basis are also random. What is practically important is its average value $\overline{K_{IS}} = \overline{\varphi_{11} \cdot \omega_{11}}$ over the set of filter parameters, which depends on their distribution laws. Below are the results of its assessment, obtained under the assumption that these parameters are mutually independent and have normal (Gaussian) distributions with zero means and variances σ_ε^2, σ_μ^2, σ_χ^2 respectively.

Fig. 2 and 3 show the dependence of the magnitude of the decrease in the average interference suppression coefficient $\overline{K_{IS}}$ from n = 2 to 5 active jammers on the dispersion $\sigma_\varepsilon^2 = \sigma_\mu^2 = \sigma_\chi^2 = \sigma^2$ of the parameters of the differences in the characteristics of the receiving paths:

$$\delta = \frac{\overline{K_{IS}(k_{\max}, \ell_{\max})}}{K_{ISa\nu}}, \quad K_{IS}(k_{\max}, \ell_{\max}) = (k_{\max}, \ell_{\max})^{-1} \cdot \sum_{k=1}^{k_{\max}} \sum_{\ell=1}^{\ell_{\max}} K_{ISk,\ell} \tag{16}$$

The terms of the sum in (16) are the values of the interference suppression coefficient for the k-th ($k \in 1$, $k_{max} = 500$) implementation of a random set of parameters for differences in the characteristics of receiving paths with a given dispersion in the ℓ-th ($\ell \in 1$, $\ell_{max} = 1000$) version of the random location of interference sources in space.The denominator (16) $K_{ISa\nu}$ corresponds to the average value of the interference suppression coefficient over the ℓ_{max} positions of active jammers under hypothetical conditions of complete coincidence of the characteristics of all receiving channels and the nominal value of their parameters ν_p, μ_p, χ_p. The ratio of the total interference power to the internal noise power in the main reception channel is $\eta = 20dB$ (Fig. 2) and $\eta = 30dB$ (Fig. 3).

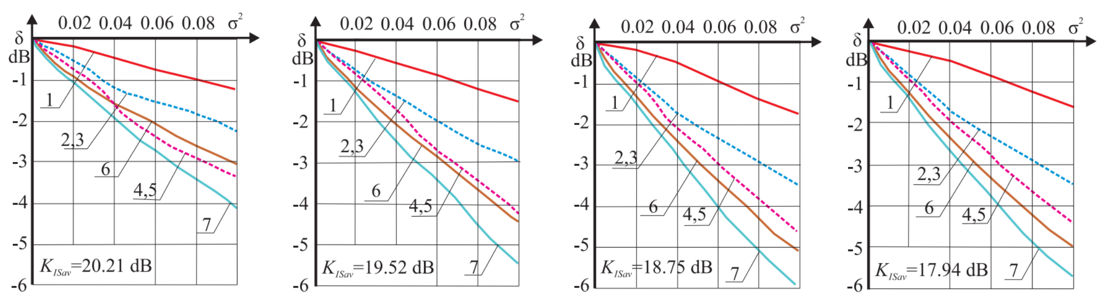


Fig. 2. Dependence of the magnitude of the reduction in the interference suppression coefficient on the dispersion of differences in the characteristics of receiving channels $(\eta{=}20dB) : a - n{=}2;$ $b - n = 3; c - n = 4; d - n = 5$

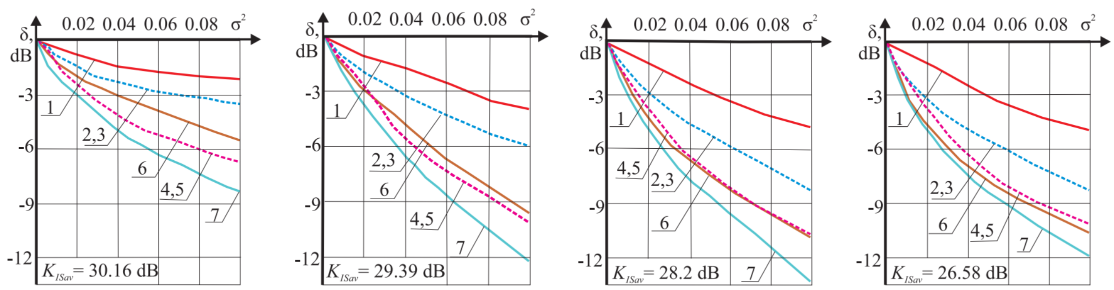
The dependence curves in these figures have the following meaning:

- dependence curve 1. The electrical lengths of the receiving paths are the same, there is no shift in their central frequencies, and only the widths of their passbands differ, i.e., $\nu_p^2 \neq \nu_q^2$, $\mu_p = \mu_q = 0$, $\chi_p = \chi_q$, and in accordance with (15)

$$a_{pq} = \sqrt{2}/\sqrt{\nu_p^2 + \nu_q^2}, \quad p, q \in 1, \ldots, N \tag{17}$$

- dependence curve 2. The electrical lengths and bandwidths of the receiving paths are the same, but the settings of the central frequencies differ, i.e. $\nu_p^2 = \nu_q^2 = 1$, $\mu_p \neq \mu_q$, $\chi_p = \chi_q$,

$$a_{pq} = \exp(-\frac{\pi}{2}(\mu_p - \mu_q)^2), \quad p, q \in 1, \ldots, N; \tag{18}$$



(a) (b) (c) (d)

Fig. 3. Dependence of the magnitude of the reduction in the interference suppression coefficient on the dispersion of differences in the characteristics of receiving channels ($\eta{=}30dB$) : $a-n{=}2$; $b-n=3; c-n=4; d-n=5$

- dependence curve 3. The settings of the central frequencies and bandwidth of the receiving paths are the same, but their electrical lengths differ, i.e. $\nu_p^2=\nu_q^2=1$, $\mu_p=\mu_q=0$, $\chi_p\neq\chi_q$,

$$a_{pq}=\exp(-\frac{\pi}{2}(\chi_p-\chi_q)^2),\quad p,q\in 1,\ldots,N; \tag{19}$$

- dependence curve 4. The electrical lengths of the receiving paths are the same, but the settings of their central frequencies and bandwidths differ, i.e. $\nu_p^2\neq\nu_q^2$, $\mu_p\neq\mu_q$, $\chi_p=\chi_q$,

$$a_{pq}=\frac{\sqrt{2}}{\sqrt{\nu_p^2+\nu_q^2}}\cdot\exp(-\pi\cdot\frac{(\mu_p-\mu_q)^2}{\nu_p^2+\nu_q^2}),\quad p,q\in 1,\ldots,N \tag{20}$$

- dependence curve 5. The setting of the central frequencies is the same, but the passbands and electrical lengths of the receiving channels differ, i.e. $\nu_p^2\neq\nu_q^2$, $\mu_p=\mu_q=0$, $\chi_p\neq\chi_q$,

$$a_{pq}=\frac{\sqrt{2}}{\sqrt{\nu_p^2+\nu_q^2}}\cdot\exp(-\pi\cdot\frac{\nu_p^2\cdot\nu_q^2(\chi_p-\chi_q)^2}{\nu_p^2+\nu_q^2}),\quad p,q\in 1,\ldots,N \tag{21}$$

- dependence curve 6. The bandwidths of the receiving paths are the same, but the settings of their central frequencies and electrical lengths differ, i.e. $\nu_p^2=\nu_q^2=1$, $\mu_p\neq\mu_q$, $\chi_p\neq\chi_q$,

$$a_{pq}=\exp\left(-\pi\cdot\frac{(\chi_p-\chi_q)^2+(\mu_p-\mu_q)^2}{2}\right)\cdot\exp(-j\cdot\pi\cdot(\mu_p+\mu_q)\cdot(\chi_p-\chi_q)),$$
$$p,q\in 1,\ldots,N; \tag{22}$$

- dependence curve 7. All characteristics of receiving paths differ — bandwidths, center frequency settings and electrical lengths. The elements a_{pq} are calculated using (15).

In Fig. 4 and 5 show the empirical distribution functions of the reduction in the interference suppression coefficient (16) over a set of $L=1000$ locations of two ($n=2$) (a, b) and four ($n=4$) (c, d) active jammers with values of dispersion parameters of the differences in the characteristics of receiving paths of $\sigma^2=0.02(a,c)$ and $\sigma^2=0.1(b,d)$. The ratio of the total interference power to the internal noise power in the main reception channel is $\eta=20dB$ (Fig. 4) and $\eta=30dB$ (Fig. 5). They provide more complete information about the statistical properties of losses, allowing one to estimate their confidence intervals in the analyzed situations.

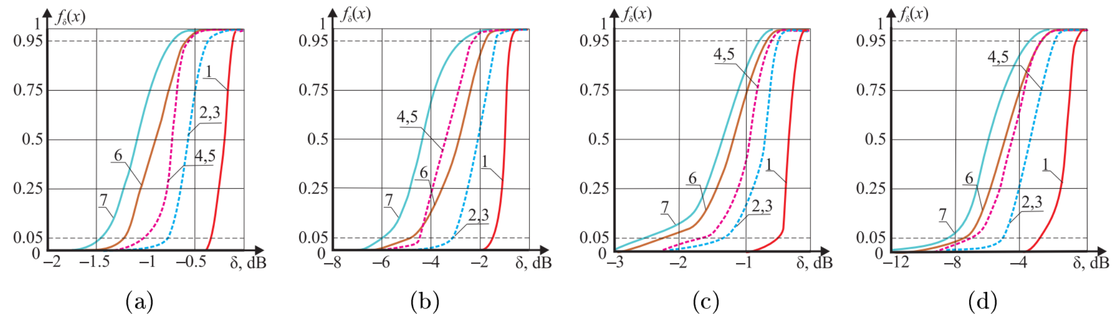


Fig. 4. Empirical distribution functions for the reduction in interference suppression coefficient due to differences in the characteristics of receiving channels ($\eta = 20dB$) : $a - n = 2$, $\sigma^2 = 0.02$; $b - n = 2$, $\sigma^2 = 0.1$; $c - n = 4\, \sigma^2 = 0.02$; $d - n = 4$, $\sigma^2 = 0.1$

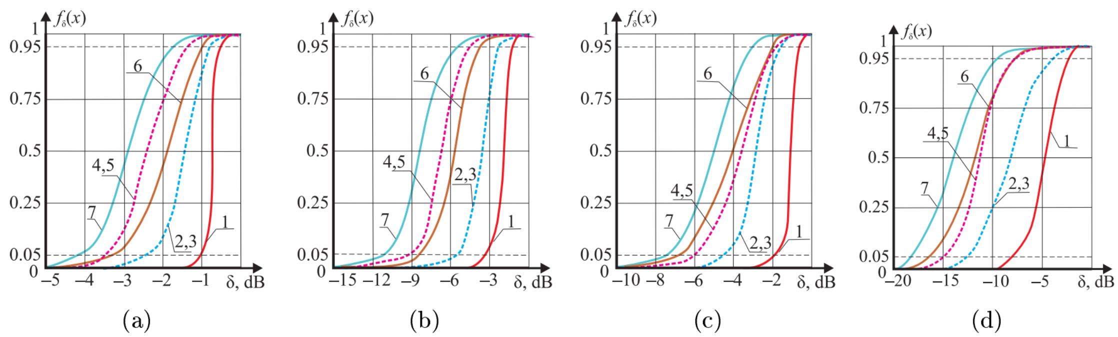


Fig. 5. Empirical distribution functions for the reduction in interference suppression coefficient due to differences in the characteristics of receiving channels ($\eta = 30dB$) : $a - n = 2, \sigma^2 = 0.02$; $b - n = 2$, $\sigma^2 = 0.1$; $c - n = 4$, $\sigma^2 = 0.02$; $d - n = 4$, $\sigma^2 = 0.1$

3. Analysis of the calculation results for reducing the level of interference suppression coefficient caused by differences in the characteristics of receiving channels

Analysis of the results of calculations performed to reduce the value of the interference suppression coefficient caused by differences in the characteristics of receiving channels allows us to draw the following conclusions:

1. The average (over multiple positions of active jammers) reduction in the achievable level of interference compensation due to differences in the characteristics of receiving channels depends on: - the nature and extent of differences; - number and intensity of interference sources.

2. The difference in the bandwidths of receiving channels has the least influence (dependence curve 1). With dispersion $\sigma_\varepsilon^2 = 0.01$ of random relative bands $\nu_p = F_p/F_0 = 1 + \varepsilon_p$, $p, q \in 1, N$, the average loss of the interference suppression coefficient K_{IS} when changing the number of active jammers from 2 to 5 is from 1 to 1.7 dB with a ratio of interference power to internal noise power of $\eta = 20dB$ (Fig. 2) and from 2.2 up to 5 dB with $\eta = 30dB$ (Fig. 3).

3. Differences in the setting of central frequencies and electrical lengths of receiving paths with equal dispersions $\sigma_\mu^2 = \sigma_\chi^2$ of random delays $\chi_p = \tau_p/T_0$ and relative shifts of the central frequency $\mu_p = \delta f_p/F_0$, $p \in 1$, N have almost the same effect on the amount of losses (dependence

curves 2 and 3). The reason for this is the coincidence in this case of the elements $a_{pq}(18)$ of the "decorrelation matrix" A(7). These elements are on average smaller than in the previous case, which is why the negative impact of the factors caused by them is greater.

4. Under practically important conditions of "small" dispersions $\sigma_\varepsilon^2 = 0.01$, the elements of the a_{pq} (20) and (21) do not have significant differences. Because of this, the influence of differences in passbands simultaneously with a shift in the center frequency or with a difference in the electrical lengths of the receiving paths (dependence curves 4 and 5) is approximately the same and has greater weight than the influence of the previous factors. At the same time, the combined effect of differences in the central frequencies and electrical lengths of the receiving paths with the same passbands (dependence curve 6) can reduce the value of K_{IS} both more and less than in the previous case.

5. The average reduction in the interference suppression coefficient K_{IS} under the isolated and combined action of the factors under consideration increases with increasing intensity and number n of interference sources. As follows from the analysis of Fig. 4 and 5, the confidence intervals are maximum under the combined action of the factors under consideration and increase with increasing dispersion σ^2 of the parameters of differences in the characteristics of receiving paths, the number and relative intensity of interference.

It is advisable to use the proposed model and the program that implements it when justifying the requirements for the permissible value of differences in the characteristics of the receiving channels of the designed radio devices.

References

[1] V.N.Tyapkin, V.N.Ratushnyak, D.D.Dmitriev, V.G.Konnov, Space-time processing of signals in angle measurement navigation receivers, 2016 International Siberian Conference on Control and Communications, SIBCON 2016 – Proceedings, Moscow, 12–14 may 2016, Moscow, 2016, 7491671. DOI: 10.1109/SIBCON.2016.7491671

[2] S.N.Karutin, V.N.Kharisov, V.S.Pavlov, Optimal Space-Time Processing Algorithms for High-Precision Applications, *Radioengineering*, (2018), no. 9, 131–138. DOI: 10.18127/j00338486-201809-23.

[3] Radio-Electronic Systems: Fundamentals of Construction and Theory. Directory, Ed. 2nd, revised and additional, Edited by Ya. D. Shirman, Moscow, Radiotekhnika, 2007.

[4] V.N.Tyapkin, I.N.Kartsan, D.D.Dmitriev, S.V.Efremova, Algorithms for adaptive processing of signals in a flat phased antenna array, International Siberian Conference on Control and Communications, SIBCON 2017 – Proceedings, Astana, 29–30 june 2017, Astana, 2017, 7998452. DOI: 10.1109/SIBCON.2017.7998452.

[5] V.N.Kharisov, A.V.Peltin, The Spatial-Temporal Algorithm for Processing of Multipath Signal for Receivers with an Antenna Array, *Radioengineering*, (2017), no. 11, 32–38.

[6] V.N.Tyapkin, D.D.Dmitriev, T.G.Moshkina, Potential Interference Immunity of Navigation Equipment of Customers of Satellite Radio Navigational Systems, *Vestnik SibSAU*, **43**(2012), no. 3, 113–119 (in Russian).

[7] V.N.Tyapkin, D.D.Dmitriev, Yu.L.Fateev, N.S.Kremez, The Synthesis Algorithm for Spatial Filtering to Maintain a Constant Level of the Useful Signal, *J. Sib. Fed. Univ. Math. Phys.*, **9**(2016), no. 2, 258–268. DOI: 10.17516/1997-1397-2016-9-2-258-268

[8] V.N.Tyapkin, I.N.Kartsan, D.D.Dmitriev, A.E.Goncharov, Correcting Non-Indentityin Receiving Channels in Interference-Immune Systems for Glonass and GPS, International Siberian Conference on Control and Communications, SIBCON 2015 – Proceedings, Omsk, 21–23 may 2015, Omsk, 2015, 7147246. DOI: 10.1109/SIBCON.2015.7147246.

[9] V.N.Tyapkin, E.N.Garin, V.N. Ratushniak, et al., The Spatial Noise Suppression In Various Configurations Goniometric Of Navigation Equipment, *Science Intensive Technologies*, **17**(2016), no. 8, 52–56 (in Russian).

[10] V.S.Efimenko, V.N.Kharisov, Adaptive Space-Time Filtering for Multi-Channel Reception, *Soviet Journal of Communications Technology and Electronics*, **32**(1987), no. 9, 1893–1901 (in Russian).

[11] I.N.Kartsan, A.E.Goncharov, P.V.Zelenkov, et al., Synthesis of an Algorithm for Interference Immunity, IOP Conference Series: Materials Science and Engineering, Krasnoyarsk, 11–15 april 2016, Vol. 155, Krasnoyarsk: Institute of Physics Publishing, 2016, 012019. DOI: 10.1088/1757-899X/155/1/012019.

Модель приемных каналов адаптивной антенной решетки для оценки влияния различия их характеристик на эффективность подавления помех

Валерий Н. Тяпкин
Дмитрий Д. Дмитриев
Сибирский федеральный университет
Красноярск, Российская Федерация
Павел В. Штро
Павел В. Луферчик
АО "НПП "Радиосвязь"
Красноярск, Российская Федерация
Игорь В. Тяпкин
Евгений Д. Михов
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. В статье предложена математическая модель, позволяющая оценить влияние различий в характеристиках приемных каналов на качество подавления шумовых помех в радиотехнических устройствах, оснащенных антенными решетками. Исследовано влияние различий в ширине полосы пропускания приемных каналов, настройки их центральных частот и электрических длин.Приведены зависимости потерь величины среднего коэффициента подавления помех от дисперсии параметров различий в характеристиках приемных каналов. Предложенную модель целесообразно использовать при обосновании требований к допустимой величине различий характеристик приемных каналов проектируемых радиотехнических устройств

Ключевые слова: антенная решетка, подавление помех, различие характеристик приемных каналов, коэффициент подавления помех.

EDN: RURNCB
УДК 517.95

Weighted Analogue of LBB Conditions for Solving the Stokes Problem with Model Boundary Conditions in a Domain with Singularity

Alexey V. Rukavishnikov*
Institute of Applied Mathematics
Far Eastern Branch of the Russian Academy of Sciences
Khabarovsk, Russian Federation

Received 10.08.2024, received in revised form 15.09.2024, accepted 24.10.2024

Abstract. The concept of a R_ν-generalized solution for the Stokes problem with model boundary conditions in a domain with a corner singularity is defined in the paper. Weighted analogue of the Ladyzhenskaya–Babuska–Brezzi conditions in a domain with a re-entrant corner is proven.

Keywords: corner singularity, Stokes problem with model boundary conditions, R_ν-generalized solution.

Citation: A.V. Rukavishnikov, Weighted Analogue of LBB Conditions for Solving the Stokes Problem with Model Boundary Conditions in a Domain with Singularity, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 91–99. EDN: RURNCB.

Introduction

The Stokes system is considered in the paper. Solution of such system is the main problem in computational fluid dynamics. The system with homogeneous Dirichlet boundary conditions for the velocity field has been studied from both theoretical and practical points of view. A detailed analysis of the problem is presented in [1, 2]. Using Schur complement operator (see, for example, [3]), the problem is reduced to separately find velocity field $\mathbf{u}$ and pressure p. Moreover, in order to find the pressure function, it is not necessary to know its values on the boundary of a domain and require additional smoothness of the solution (see, for example, [3]). In the presented paper, fundamentally different boundary conditions are considered, namely, $\mathbf{u}\cdot\mathbf{n} = 0$ and $\operatorname{curl} \mathbf{u} = 0$, where $\mathbf{u}\cdot\mathbf{n} = u_1 n_1 + u_2 n_2, \mathbf{n}$ is the outer unit normal vector to the boundary and $\operatorname{curl} \mathbf{u} = \dfrac{\partial u_2}{\partial x_1} - \dfrac{\partial u_1}{\partial x_2}$. Such boundary conditions will be called model conditions. They are of particular interest from a practical point of view associated with the Schur complement operator. More details on this can be found in [4]. On the other hand, it is fundamental to consider the Stokes system in a polygonal non-convex domain Ω with a re-entrant corner ω on the boundary, i.e., a corner greater than π. In this case, a problem with a corner singularity is considered. Moreover, it is known (see, for example, [5]) that generalized solution of such problem in the velocity-pressure variables $(\mathbf{u}, p)$ does not belong to the Sobolev spaces $\mathbf{W}_2^2(\Omega)$ and $W_2^1(\Omega)$, respectively. Therefore, using any classical approximate approach (see [6]), by the principle of consistent estimates its approximate solution converges to an exact one at the rate

*78321a@mail.ru https://orcid.org/0000-0002-7585-4559

no faster than $\mathcal{O}(h^{\alpha})$, where h is the space step, and α is significantly less than one and decreases with increasing the value of corner ω. At the same time, an appropriate convergence rate is of the order $\mathcal{O}(h)$ as is in the case of a convex domain Ω. The solution of the Stokes problem with model boundary conditions is defined in the paper as R_ν-generalized solution in sets of weighted spaces. In this case, solution is sought in sets of more general spaces than weighted Sobolev spaces $W^k_{2,\beta}(\Omega)$, $\beta > 0$. Note that resulting variational formulation for determining generalized solution of the problem is not symmetric unlike the classical one [4]. This further adds difficulties to the proof of existence and uniqueness of R_ν-generalized solution of the Stokes problem with the proposed model boundary conditions. For the first time R_ν-generalized solution was defined for elliptic problems with Dirichlet boundary conditions [7]. Approximate R_ν-generalized solution obtained with the weighted finite element method converges to an exact solution with the rate $\mathcal{O}(h)$ for various differential problems with Dirichlet boundary conditions [8–12]. Convergence rate of such solutions does not depend on the value of a re entrant corner ω. Moreover, the result is achieved without refinement of the mesh in the vicinity of the singularity point. In the presented paper function properties in sets of weighted spaces are studied. A weighted analogue of the Ladyzhenskaya-Babushka-Brezzi conditions for the Stokes problem with the considered model boundary conditions is established.

1. Formulation of the Stokes problem with model boundary conditions. Definition of an R_ν-generalized solution

Let domain Ω be a bounded non-convex polygon with the boundary $\partial\Omega$. It contains a re-entrant corner $\omega, \omega \in (\pi, 2\pi)$ at the origin $\mathcal{O} = (0,0)$, $\bar{\Omega} = \Omega \cup \partial\Omega$.

Let $\mathbf{x} = (x_1, x_2)$ be an element of R^2, $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$, $d\mathbf{x} = dx_1 dx_2$. The Stokes problem is formulated as follows. For given functions $\mathbf{f} = (f_1, f_2)$ and g in Ω find the velocity field $\mathbf{u} = (u_1, u_2)$ and pressure p, which satisfy the system of differential equations and boundary conditions of the form

$$-\triangle \mathbf{u} + \nabla p = \mathbf{f}, \qquad \text{div } \mathbf{u} = g \qquad \text{in} \qquad \Omega, \tag{1}$$

$$\mathbf{u} \cdot \mathbf{n} = 0 \qquad \text{on} \qquad \partial\Omega, \tag{2}$$

$$\text{curl } \mathbf{u} = 0 \qquad \text{on} \qquad \partial\Omega, \tag{3}$$

where $\mathbf{u}\cdot\mathbf{n} = u_1 n_1 + u_2 n_2$, $\text{curl } \mathbf{u} = \dfrac{\partial u_2}{\partial x_1} - \dfrac{\partial u_1}{\partial x_2}$ and $\mathbf{n} = (n_1, n_2)$ is the outer unit normal vector to $\partial\Omega$.

Let us define necessary spaces and sets of weight functions. The weighted space of functions $v(\mathbf{x})$ with limited norm is denoted by $L_{2,\alpha}(\Omega)$:

$$\|v\|_{L_{2,\alpha}(\Omega)} = \left(\int_\Omega \rho^{2\alpha}(\mathbf{x})\, v^2(\mathbf{x}) d\mathbf{x} \right)^{1/2}, \qquad \alpha > 0.$$

Spaces and sets of functions $\mathbf{v} = (v_1, v_2)$ are marked in bold. Here $\mathbf{v} \in \mathbf{L}_{2,\alpha}(\Omega)$ if the quantity $\|\mathbf{v}\|_{\mathbf{L}_{2,\alpha}(\Omega)} = \left(\|v_1\|^2_{L_{2,\alpha}(\Omega)} + \|v_2\|^2_{L_{2,\alpha}(\Omega)} \right)^{1/2}$ is limited.

Let $\mathbf{H}_\alpha(\text{curl})(\Omega)$ be the space of functions $\mathbf{v}(\mathbf{x})$ such that $\mathbf{v} \in \mathbf{L}_{2,\alpha}(\Omega)$ and $\text{curl } \mathbf{v} \in L_{2,\alpha}(\Omega)$ with bounded norm

$$\|\mathbf{v}\|_{\mathbf{H}_\alpha(\text{curl})(\Omega)} = \left(\|\mathbf{v}\|^2_{\mathbf{L}_{2,\alpha}(\Omega)} + \|\text{curl } \mathbf{v}\|^2_{L_{2,\alpha}(\Omega)} \right)^{1/2}.$$

The space of functions $\mathbf{v}(\mathbf{x})$ is denoted by $\mathbf{H}_\alpha(\mathrm{div})(\Omega)$ such that $\mathbf{v} \in \mathbf{L}_{2,\alpha}(\Omega)$ and $\mathrm{div}\ \mathbf{v} \in L_{2,\alpha}(\Omega)$ with limited norm

$$\|\mathbf{v}\|_{\mathbf{H}_\alpha(\mathrm{div})(\Omega)} = \Big(\|\mathbf{v}\|^2_{\mathbf{L}_{2,\alpha}(\Omega)} + \|\mathrm{div}\ \mathbf{v}\|^2_{L_{2,\alpha}(\Omega)}\Big)^{1/2}.$$

Let $\overset{\circ}{\mathbf{H}}_\alpha(\mathrm{div})(\Omega)$ be the subspace of $\mathbf{H}_\alpha(\mathrm{div})(\Omega)$ such that $\{\mathbf{v} \in \mathbf{H}_\alpha(\mathrm{div})(\Omega) : \mathbf{v}\cdot\mathbf{n} = 0$ на $\partial\Omega\}$ with bounded norm of space $\mathbf{H}_\alpha(\mathrm{div})(\Omega)$. Next, the intersection of spaces $\overset{\circ}{\mathbf{H}}_\alpha(\mathrm{div})(\Omega)$ and $\mathbf{H}_\alpha(\mathrm{curl})(\Omega)$ of functions $\mathbf{v}(\mathbf{x})$ is denoted by $\mathbf{U}_\alpha(\Omega)$ with limited norm

$$\|\mathbf{v}\|_{\mathbf{U}_\alpha(\Omega)} = \Big(\|\mathbf{v}\|^2_{\mathbf{L}_{2,\alpha}(\Omega)} + \|\mathrm{div}\ \mathbf{v}\|^2_{L_{2,\alpha}(\Omega)} + \|\mathrm{curl}\ \mathbf{v}\|^2_{L_{2,\alpha}(\Omega)}\Big)^{1/2}.$$

Let $W^1_{2,\alpha}(\Omega)$ be the weighted space of functions $v(\mathbf{x})$ with bounded norm

$$\|v\|_{W^1_{2,\alpha}(\Omega)} = \Big(\|v\|^2_{L_{2,\alpha}(\Omega)} + \sum_{|l|=1}\int_\Omega \rho^{2\alpha}(\mathbf{x})|D^l v(\mathbf{x})|^2 d\mathbf{x}\Big)^{1/2},$$

where $D^l v(\mathbf{x}) = \dfrac{\partial^{|l|}v(\mathbf{x})}{\partial x_1^{l_1}\partial x_2^{l_2}}$, $l = (l_1, l_2)$, $|l| = l_1 + l_2$, l_i are non-negative integers $i \in \{1, 2\}$. The subspace of functions $v(\mathbf{x})$ from $W^1_{2,\alpha}(\Omega)$ is denoted by $\overset{\circ}{W}{}^1_{2,\alpha}(\Omega)$ such that $v = 0$ on $\partial\Omega$ with limited norm $W^1_{2,\alpha}(\Omega)$. Similarly, spaces $\mathbf{W}^1_{2,\alpha}(\Omega)$ and $\overset{\circ}{\mathbf{W}}{}^1_{2,\alpha}(\Omega)$ of functions $\mathbf{v} = (v_1, v_2)$ are introduced such that $v_i \in W^1_{2,\alpha}(\Omega,\delta)$ and $v_i \in \overset{\circ}{W}{}^1_{2,\alpha}(\Omega,\delta)$, respectively, with bounded norm $\|\mathbf{v}\|_{\mathbf{W}^1_{2,\alpha}(\Omega)} = \Big(\|v_1\|^2_{W^1_{2,\alpha}(\Omega)} + \|v_2\|^2_{W^1_{2,\alpha}(\Omega)}\Big)^{1/2}$. The intersection of the circle with radius δ centred at the origin $\mathcal{O}$ with $\bar\Omega$ is denoted by $\Omega_\delta = \{\mathbf{x} \in \bar\Omega : \|\mathbf{x}\| \leqslant \delta \ll 1, \delta > 0\}$. Let us introduce the weight function $\rho(\mathbf{x})$ in $\bar\Omega$ as follows $\rho(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|, & \text{if } \mathbf{x} \in \Omega_\delta, \\ \delta\ , & \text{if } \mathbf{x} \in \bar\Omega \setminus \Omega_\delta. \end{cases}$

Let us define the following conditions for function $v(\mathbf{x})$

$$\|v\|_{L_{2,\alpha}(\Omega\setminus\Omega_\delta)} \geqslant C_1 > 0, \tag{4}$$

$$|v(\mathbf{x})| \leqslant C_2\delta^{\alpha-\tau}\rho^{\tau-\alpha}(\mathbf{x}), \quad \mathbf{x} \in \Omega_\delta, \tag{5}$$

where C_2 is a positive constant, τ is a small positive parameter independent of δ, α and $v(\mathbf{x})$. A set of functions $v(\mathbf{x})$ from space $L_{2,\alpha}(\Omega)$ satisfying conditions (4) and (5) with limited norm $L_{2,\alpha}(\Omega)$ is denoted by $L_{2,\alpha}(\Omega,\delta)$. Let $L^0_{2,\alpha}(\Omega,\delta)$ be a subset of functions $v(\mathbf{x})$ from $L_{2,\alpha}(\Omega,\delta)$ such that $\int_\Omega \rho^\alpha(\mathbf{x})v(\mathbf{x})d\mathbf{x} = 0$ with bounded norm $L_{2,\alpha}(\Omega)$.

Next, we define sets $\mathbf{H}_\alpha(\mathrm{curl})(\Omega,\delta)$, $\mathbf{H}_\alpha(\mathrm{div})(\Omega,\delta)$, $\overset{\circ}{\mathbf{H}}_\alpha(\mathrm{div})(\Omega,\delta)$ and $\mathbf{U}_\alpha(\Omega,\delta)$ of functions $\mathbf{v} = (v_1, v_2)$ from spaces $\mathbf{H}_\alpha(\mathrm{curl})(\Omega), \mathbf{H}_\alpha(\mathrm{div})(\Omega), \overset{\circ}{\mathbf{H}}_\alpha(\mathrm{div})(\Omega)$ and $\mathbf{U}_\alpha(\Omega)$, respectively, which components satisfy conditions (4) and (5) with limited norms of relevant spaces. Let $W^1_{2,\alpha}(\Omega,\delta)$ and $\overset{\circ}{W}{}^1_{2,\alpha}(\Omega,\delta)$ are sets of functions from spaces $W^1_{2,\alpha}(\Omega)$ and $\overset{\circ}{W}{}^1_{2,\alpha}(\Omega)$ respectively, satisfying conditions (4), (5) and $|D^1 v(\mathbf{x})| \leqslant C_2\delta^{\alpha-\tau}\rho^{\tau-\alpha-1}(\mathbf{x})$, $\mathbf{x} \in \Omega_\delta$, with bounded norm of space $W^1_{2,\alpha}(\Omega)$. Sets of functions $\mathbf{v} = (v_1, v_2)$ are denoted by $\mathbf{W}^1_{2,\alpha}(\Omega,\delta)$ and $\overset{\circ}{\mathbf{W}}{}^1_{2,\alpha}(\Omega,\delta)$ such that $v_i \in W^1_{2,\alpha}(\Omega,\delta)$ and $v_i \in \overset{\circ}{W}{}^1_{2,\alpha}(\Omega,\delta)$, respectively.

Let us prove the following assertion.

Lemma 1. *Let function* $\mathbf{u} \in \mathbf{U}_\nu(\Omega, \delta)$ *and* $\mathit{curl}\,\mathbf{u} = 0$ *on* $\partial\Omega$*. Then for arbitrary function* $\mathbf{v} \in \mathbf{U}_\nu(\Omega, \delta)$ *the identity*

$$\int_\Omega \nabla\mathbf{u} : \nabla(\rho^{2\nu}\mathbf{v})d\mathbf{x} + I(\mathbf{u},\mathbf{v}) = \int_\Omega \mathit{curl}\,\mathbf{u}\ \mathit{curl}\,(\rho^{2\nu}\mathbf{v})d\mathbf{x} + \int_\Omega \mathit{div}\,\mathbf{u}\ \mathit{div}\,(\rho^{2\nu}\mathbf{v})d\mathbf{x} \tag{6}$$

holds, where

$$I(\mathbf{u},\mathbf{v}) := -\Bigg[\int\limits_{\partial\Omega} \rho^{2\nu}\frac{\partial u_1}{\partial x_1}n_1 v_1 ds + \int\limits_{\partial\Omega} \rho^{2\nu}\frac{\partial u_1}{\partial x_2}n_2 v_1 ds + \int\limits_{\partial\Omega} \rho^{2\nu}\frac{\partial u_2}{\partial x_1}n_1 v_2 ds + \int\limits_{\partial\Omega} \rho^{2\nu}\frac{\partial u_2}{\partial x_2}n_2 v_2 ds\Bigg]. \tag{7}$$

Proof. By definition

$$\int_\Omega \nabla\mathbf{u} : \nabla(\rho^{2\nu}\mathbf{v})d\mathbf{x} = \int_\Omega \Big[\frac{\partial u_1}{\partial x_1}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_1} + \frac{\partial u_2}{\partial x_2}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_2}\Big]d\mathbf{x}, \tag{8}$$

$$\begin{aligned}\int_\Omega \mathrm{curl}\,\mathbf{u}\ \mathrm{curl}\,(\rho^{2\nu}\mathbf{v})d\mathbf{x} = \int_\Omega \Big[&\frac{\partial u_2}{\partial x_1}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_1} - \frac{\partial u_2}{\partial x_1}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_2} - \\ &- \frac{\partial u_1}{\partial x_2}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_2}\Big]d\mathbf{x},\end{aligned} \tag{9}$$

$$\begin{aligned}\int_\Omega \mathrm{div}\,\mathbf{u}\ \mathrm{div}\,(\rho^{2\nu}\mathbf{v})d\mathbf{x} = \int_\Omega \Big[&\frac{\partial u_1}{\partial x_1}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_1} + \frac{\partial u_1}{\partial x_1}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_2} + \\ &+ \frac{\partial u_2}{\partial x_2}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_1} + \frac{\partial u_2}{\partial x_2}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_2}\Big]d\mathbf{x}.\end{aligned} \tag{10}$$

The following equalities

$$-\int_\Omega \frac{\partial u_2}{\partial x_1}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_2}d\mathbf{x} = -\int_\Omega \frac{\partial u_2}{\partial x_2}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_1}d\mathbf{x} + \int\limits_{\partial\Omega} \rho^{2\nu}\frac{\partial u_2}{\partial x_2}n_1 v_1 ds - \int\limits_{\partial\Omega} \rho^{2\nu}\frac{\partial u_2}{\partial x_1}n_2 v_1 ds, \tag{11}$$

$$-\int_\Omega \frac{\partial u_1}{\partial x_2}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_1}d\mathbf{x} = -\int_\Omega \frac{\partial u_1}{\partial x_1}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_2}d\mathbf{x} + \int\limits_{\partial\Omega} \rho^{2\nu}\frac{\partial u_1}{\partial x_1}n_2 v_2 ds - \int\limits_{\partial\Omega} \rho^{2\nu}\frac{\partial u_1}{\partial x_2}n_1 v_2 ds \tag{12}$$

are valid. Using expansions (8)-(10), one can obtain

$$\int_\Omega \nabla\mathbf{u} : \nabla(\rho^{2\nu}\mathbf{v})d\mathbf{x} = \int_\Omega \mathrm{curl}\,\mathbf{u}\ \mathrm{curl}\,(\rho^{2\nu}\mathbf{v})d\mathbf{x} + \int_\Omega \mathrm{div}\,\mathbf{u}\ \mathrm{div}\,(\rho^{2\nu}\mathbf{v})d\mathbf{x} - E(\mathbf{u},\mathbf{v}), \tag{13}$$

where

$$E(\mathbf{u},\mathbf{v}) := \int_\Omega \Big[\frac{\partial u_1}{\partial x_1}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_2} + \frac{\partial u_2}{\partial x_2}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_1} - \frac{\partial u_2}{\partial x_1}\frac{\partial(\rho^{2\nu}v_1)}{\partial x_2} - \frac{\partial u_1}{\partial x_2}\frac{\partial(\rho^{2\nu}v_2)}{\partial x_1}\Big]d\mathbf{x}. \tag{14}$$

Applying equalities (11) and (12) to (14), one can conclude

$$E(\mathbf{u},\mathbf{v}) = \int\limits_{\partial\Omega} \rho^{2\nu}\Big[\frac{\partial u_2}{\partial x_2}n_1 v_1 + \frac{\partial u_1}{\partial x_1}n_2 v_2 - \frac{\partial u_2}{\partial x_1}n_2 v_1 - \frac{\partial u_1}{\partial x_2}n_1 v_2\Big]ds. \tag{15}$$

Using together (7), (15), conditions $\mathbf{v}\cdot\mathbf{n} = 0$ and $\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2} = 0$ on $\partial\Omega$, and equality $\int\limits_{\partial\Omega} \rho^{2\nu}\Big(\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2}\Big)n_2 v_1 ds = \int\limits_{\partial\Omega} \rho^{2\nu}\Big(\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2}\Big)n_1 v_2 ds$, one can obtain $I(\mathbf{u},\mathbf{v}) = E(\mathbf{u},\mathbf{v})$. Lemma 1 is proven. □

Let us introduce bilinear and linear forms

$$a(\mathbf{u},\mathbf{v}) = \int_\Omega \text{curl }\mathbf{u}\text{ curl }(\rho^{2\nu}\mathbf{v})d\mathbf{x} + \int_\Omega \text{div }\mathbf{u}\text{ div }(\rho^{2\nu}\mathbf{v})d\mathbf{x}, \quad b_1(\mathbf{v},s) = -\int_\Omega s\text{ div }(\rho^{2\nu}\mathbf{v})d\mathbf{x},$$

$$b_2(\mathbf{u},q) = -\int_\Omega (\rho^{2\nu}\, q)\text{ div }\mathbf{u}\, d\mathbf{x}, \quad l(\mathbf{v}) = \int_\Omega \mathbf{f}\cdot(\rho^{2\nu}\mathbf{v})d\mathbf{x}, \quad c(q) = \int_\Omega \rho^{2\nu} g\, q d\mathbf{x}.$$

Let us define the concept of an R_ν-generalized solution of the Stokes problem (1)–(3) with model boundary conditions in weighted sets.

Definition 1. *A pair of functions $(\mathbf{u}_\nu,p_\nu) \in \mathbf{U}_\nu(\Omega,\delta)\times L^0_{2,\nu}(\Omega,\delta)$ is called R_ν-generalized solution of the Stokes problem (1)—(3) with function $\mathbf{u}_\nu$ satisfies boundary conditions (2) and (3) if for all pairs of functions $(\mathbf{v},q)\in \mathbf{U}_\nu(\Omega,\delta)\times L^0_{2,\nu}(\Omega,\delta)$ the integral identities*

$$a(\mathbf{u}_\nu,\mathbf{v}) + b_1(\mathbf{v},p_\nu) = l(\mathbf{v}), \tag{16}$$

$$b_2(\mathbf{u}_\nu,q) = c(q) \tag{17}$$

hold, where $\mathbf{f}\in \mathbf{L}_{2,\gamma}(\Omega)$, $g\in L_{2,\beta}(\Omega)$, $0\leqslant \gamma,\beta\leqslant \nu$ and $\mathbf{u}_\nu = (u_{1,\nu},\, u_{2,\nu})$.

Remark 1. *Since bilinear form $b_2(\cdot,\cdot)$ does not coincide with bilinear form $b_1(\cdot,\cdot)$ the variational formulation for R_ν-generalized solution of the problem is not symmetric, in contrast to the standard variational formulation for a generalized solution of the problem [4].*

Remark 2. *Bilinear form $a(\cdot,\cdot)$ is not a symmetric one.*

2. Auxiliary statements

Let us formulate and prove necessary statements.

Lemma 2 ([13])**.** *Let $\nu>0$. For an arbitrary function $z\in L_{2,\nu}(\Omega)$ that satisfies conditions (4), (5) the following estimate*

$$\int_{\Omega_\delta} \rho^{2\nu-2} z^2 d\mathbf{x} \leqslant C_3^2\delta^{2\nu}\|z\|^2_{L_{2,\nu}(\Omega)} \tag{18}$$

is valid, where $C_3 = \dfrac{C_2}{C_1}\sqrt{\dfrac{\varphi_1-\varphi_0}{2\tau}}$, $(\varphi_1-\varphi_0)$ is the magnitude of the change of a re-entrant corner.

Corollary 1. *Let conditions of Lemma 2 be satisfied then*

$$\int_\Omega \Big[\sum_{i=1}^2\Big(\frac{\partial\rho^\nu}{\partial x_i}\Big)^2\Big] z^2 d\mathbf{x} \leqslant \nu^2 C_3^2\delta^{2\nu}\|z\|^2_{L_{2,\nu}(\Omega)}. \tag{19}$$

Proof. Estimate (19) of Corollary 1 directly follows from the fact that $\sum\limits_{i=1}^2\Big(\dfrac{\partial\rho^\nu}{\partial x_i}\Big)^2 =$ $= \begin{cases}\nu^2\rho^{2\nu-2}, \mathbf{x}\in\Omega_\delta,\\ 0, \mathbf{x}\in\bar\Omega\setminus\Omega_\delta\end{cases}$ inequality (18).

Let us connect norms of functions $\mathbf{z}$ and $\rho^\nu\mathbf{z}$ from sets $\mathbf{U}_\nu(\Omega,\delta)$ and $\mathbf{U}_0(\Omega,\delta)$, respectively.

Lemma 3. *Function* $\mathbf{z} \in \mathbf{U}_\nu(\Omega, \delta)$ *if and only if* $\rho^\nu \mathbf{z} \in \mathbf{U}_0(\Omega, \delta)$ *and*

$$\|\rho^\nu \mathbf{z}\|_{\mathbf{U}_0(\Omega)} \leqslant C_4 \|\mathbf{z}\|_{\mathbf{U}_\nu(\Omega)}, \tag{20}$$

$$\|\mathbf{z}\|_{\mathbf{U}_\nu(\Omega)} \leqslant C_4 \|\rho^\nu \mathbf{z}\|_{\mathbf{U}_0(\Omega)}, \tag{21}$$

where $C_4 = \max\{\sqrt{2}, \sqrt{1 + 4\nu^2 C_3^2 \delta^{2\nu}}\}$.

Proof. 1. Let function $\mathbf{z} \in \mathbf{U}_\nu(\Omega, \delta)$. Let us show that $\rho^\nu \mathbf{z} \in \mathbf{U}_0(\Omega, \delta)$ and inequality (20) holds. Consider the following decompositions

$$\mathrm{curl}(\rho^\nu \mathbf{z}) = \rho^\nu \mathrm{curl}\ \mathbf{z} + \Big[z_2 \frac{\partial \rho^\nu}{\partial x_1} - z_1 \frac{\partial \rho^\nu}{\partial x_2}\Big], \tag{22}$$

$$\mathrm{div}(\rho^\nu \mathbf{z}) = \rho^\nu \mathrm{div}\ \mathbf{z} + \Big[z_1 \frac{\partial \rho^\nu}{\partial x_1} + z_2 \frac{\partial \rho^\nu}{\partial x_2}\Big]. \tag{23}$$

Using expansions (22), (23), one can conclude that

$$\|\mathrm{curl}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} \leqslant 2\|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + 4\int_\Omega \Big(\frac{\partial \rho^\nu}{\partial x_1}\Big)^2 z_2^2 d\mathbf{x} + 4\int_\Omega \Big(\frac{\partial \rho^\nu}{\partial x_2}\Big)^2 z_1^2 d\mathbf{x}, \tag{24}$$

$$\|\mathrm{div}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} \leqslant 2\|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + 4\int_\Omega \Big(\frac{\partial \rho^\nu}{\partial x_1}\Big)^2 z_1^2 d\mathbf{x} + 4\int_\Omega \Big(\frac{\partial \rho^\nu}{\partial x_2}\Big)^2 z_2^2 d\mathbf{x}. \tag{25}$$

Since $\|\rho^\nu \mathbf{z}\|_{\mathbf{L}_{2,0}(\Omega)} = \|\mathbf{z}\|_{\mathbf{L}_{2,\nu}(\Omega)}$ then applying relations (24), (25) and estimate (19), one can obtaib the chain of inequalities

$$\|\rho^\nu \mathbf{z}\|^2_{\mathbf{U}_0(\Omega)} = \|\rho^\nu \mathbf{z}\|^2_{\mathbf{L}_{2,0}(\Omega)} + \|\mathrm{curl}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} + \|\mathrm{div}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} \leqslant \|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} +$$

$$+2\|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + 2\|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + 4\sum_{j=1}^{2} \int_\Omega \Big[\sum_{i=1}^{2} \Big(\frac{\partial \rho^\nu}{\partial x_i}\Big)^2\Big] z_j^2 d\mathbf{x} \leqslant \Big(1 + 4\nu^2 C_3^2 \delta^{2\nu}\Big) \|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} +$$

$$+2\|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + 2\|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} \leqslant \max\{2, 1 + 4\nu^2 C_3^2 \delta^{2\nu}\} \|\mathbf{z}\|^2_{\mathbf{U}_\nu(\Omega)}.$$

Estimate (20) is proven and $\rho^\nu \mathbf{z} \in \mathbf{U}_0(\Omega, \delta)$.

2. Let function $\rho^\nu \mathbf{z} \in \mathbf{U}_0(\Omega, \delta)$. Let us show that $\mathbf{z} \in \mathbf{U}_\nu(\Omega, \delta)$ and inequality (21) holds. Let us express in (22) and (23) terms $\big(\rho^\nu \mathrm{curl}\ \mathbf{z}\big)$ and $\big(\rho^\nu \mathrm{div}\ \mathbf{z}\big)$, respectively. Then

$$\|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} \leqslant 2\|\mathrm{curl}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} + 4\int_\Omega \Big(\frac{\partial \rho^\nu}{\partial x_1}\Big)^2 z_2^2 d\mathbf{x} + 4\int_\Omega \Big(\frac{\partial \rho^\nu}{\partial x_2}\Big)^2 z_1^2 d\mathbf{x}, \tag{26}$$

$$\|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} \leqslant 2\|\mathrm{div}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} + 4\int_\Omega \Big(\frac{\partial \rho^\nu}{\partial x_1}\Big)^2 z_1^2 d\mathbf{x} + 4\int_\Omega \Big(\frac{\partial \rho^\nu}{\partial x_2}\Big)^2 z_2^2 d\mathbf{x}. \tag{27}$$

Since $\|\mathbf{z}\|_{\mathbf{L}_{2,\nu}(\Omega)} = \|\rho^\nu \mathbf{z}\|_{\mathbf{L}_{2,0}(\Omega)}$ then applying inequalities (26), (27) and estimate (19), one can obtaib the chain of inequalities

$$\|\mathbf{z}\|^2_{\mathbf{U}_\nu(\Omega)} = \|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} + \|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + \|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} \leqslant \|\rho^\nu \mathbf{z}\|^2_{\mathbf{L}_{2,0}(\Omega)} + 2\|\mathrm{curl}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} +$$

$$+2\|\mathrm{div}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} + 4\sum_{j=1}^{2} \int_\Omega \Big[\sum_{i=1}^{2} \Big(\frac{\partial \rho^\nu}{\partial x_i}\Big)^2\Big] z_j^2 d\mathbf{x} \leqslant \max\{2, 1 + 4\nu^2 C_3^2 \delta^{2\nu}\} \|\rho^\nu \mathbf{z}\|^2_{\mathbf{U}_0(\Omega)}.$$

Estimate (21) is obtained and $\mathbf{z} \in \mathbf{U}_\nu(\Omega, \delta)$. Lemma 3 is proven. □

Lemma 4. *Let $\nu > 0$. Then there exists a value $\delta_0 = \delta_0(\nu) > 0$, such that for any $\delta \in (0, \delta_0]$ and arbitrary function $\mathbf{z} \in \mathbf{U}_\nu(\Omega, \delta)$ inequality*

$$\|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} \leqslant 8C^2_\Omega\big(\|curl\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + \|div\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)}\big) \tag{28}$$

holds.

Proof. Using Lemma 3 and Lemma 3.6 [6] and if $z \in \mathbf{U}_\nu(\Omega, \delta)$ then $\rho^\nu \mathbf{z} \in \mathbf{U}_0(\Omega, \delta)$, one can obtain

$$\|\rho^\nu \mathbf{z}\|_{\mathbf{L}_{2,0}(\Omega)} \leqslant C_\Omega\big(\|\mathrm{curl}(\rho^\nu \mathbf{z})\|_{L_{2,0}(\Omega)} + \|\mathrm{div}(\rho^\nu \mathbf{z})\|_{L_{2,0}(\Omega)}\big),$$

i. e.,

$$\|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} \leqslant 2C^2_\Omega\big(\|\mathrm{curl}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)} + \|\mathrm{div}(\rho^\nu \mathbf{z})\|^2_{L_{2,0}(\Omega)}\big). \tag{29}$$

Applying estimates (24) and (25) for the first and second terms of the right-hand side of (29), respectively and then inequality (19), one can find

$$\|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} \leqslant 4C^2_\Omega\big(\|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + \|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)}\big) + 8\nu^2 C_3^2 C^2_\Omega \delta^{2\nu} \|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)}$$

and

$$\Big(1 - 8\nu^2 C_3^2 C^2_\Omega \delta^{2\nu}\Big)\|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} \leqslant 4C^2_\Omega\big(\|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + \|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)}\big). \tag{30}$$

For $\nu > 0$, there exists such a value $\delta_0 = \delta_0(\nu) > 0 : \nu^2 C^2_\Omega C_3^2 \delta_0^{2\nu} = \frac{1}{16}$ that for each $\delta \in (0, \delta_0]$ the following chain of relations

$$\frac{1}{2}\|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} \leqslant \Big(1 - 8\nu^2 C_3^2 C^2_\Omega \delta^{2\nu}\Big)\|\mathbf{z}\|^2_{\mathbf{L}_{2,\nu}(\Omega)} \leqslant 4C^2_\Omega\big(\|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + \|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)}\big)$$

is valid, according to (30). Lemma 4 is proven.

By the definition of norm in space $\mathbf{U}_\nu(\Omega)$ the following statement is derived directly from Lemma 4.

Corollary 2. *Let conditions of Lemma 4 be satisfied then*

$$\|\mathbf{z}\|^2_{\mathbf{U}_\nu(\Omega)} \leqslant \big(1 + 8C^2_\Omega\big)\big(\|\mathrm{curl}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)} + \|\mathrm{div}\ \mathbf{z}\|^2_{L_{2,\nu}(\Omega)}\big). \tag{31}$$

3. Weighted analogue of LBB-conditions of forms $b_i(\mathbf{v}, s)$

Let us consider a weighted analogue of LBB-conditions of forms $b_i(\mathbf{v}, s)$ on sets of functions $\mathbf{v} \in \mathbf{U}_\nu(\Omega, \delta)$ and $s \in L^0_{2,\nu}(\Omega, \delta)$.

Theorem 1. *For each $\nu > 0$ there exists a value $\delta_1 = \delta_1(\nu) > 0$ ($\delta_1 \leqslant \delta_0, \delta_0$ from Lemma 4) such that for all $\delta \in (0, \delta_1]$ and arbitrary function $s \in L^0_{2,\nu}(\Omega, \delta)$ the following inequalities*

$$0 < \beta_i \|s\|_{L_{2,\nu}(\Omega)} \leqslant \sup_{\mathbf{v} \in \mathbf{U}_\nu(\Omega,\delta)} \frac{b_i(\mathbf{v}, s)}{\|\mathbf{v}\|_{\mathbf{U}_\nu(\Omega)}}$$

hold, where $\beta_i = \dfrac{\gamma_i}{2\sqrt{1 + 8C^2_\Omega}}$, $i = 1, 2$.

Proof. It was proved [14] that there exists a value $\delta_2 = \delta_2(\nu) > 0$ such that for all $\delta \in (0, \delta_2]$ and arbitrary function $s \in L^0_{2,\nu}(\Omega, \delta)$ the following inequalities

$$0 < \gamma_i \|s\|_{L_{2,\nu}(\Omega)} \leqslant \sup_{\mathbf{v} \in \overset{\circ}{\mathbf{W}}{}^1_{2,\nu}(\Omega,\delta)} \frac{b_i(\mathbf{v}, s)}{\|\mathbf{v}\|_{\mathbf{W}^1_{2,\nu}(\Omega)}}, \quad \gamma_i > 0 \tag{32}$$

hold. If inequality (31) is used and estimate for arbitrary function $\mathbf{v} \in \overset{\circ}{\mathbf{W}}{}^{1}_{2,\nu}(\Omega,\delta)$ is

$$\|\text{div } \mathbf{v}\|_{L_{2,\nu}(\Omega)} + \|\text{curl } \mathbf{v}\|_{L_{2,\nu}(\Omega)} \leqslant 2\|\mathbf{v}\|_{\mathbf{W}^1_{2,\nu}(\Omega)}$$

then, due to the fact that $\overset{\circ}{\mathbf{W}}{}^{1}_{2,\nu}(\Omega,\delta) \subset \mathbf{U}_\nu(\Omega,\delta)$ and for all $\delta \in (0,\delta_1]$, where $\delta_1 = \min\{\delta_0,\delta_2\}$, one can obtain from (32) the chain of inequalities

$$\gamma_i\|s\|_{L_{2,\nu}(\Omega)} \leqslant \sup_{\mathbf{v}\in\overset{\circ}{\mathbf{W}}{}^{1}_{2,\nu}(\Omega,\delta)} \frac{b_i(\mathbf{v},s)}{\|\mathbf{v}\|_{\mathbf{W}^1_{2,\nu}(\Omega)}} \leqslant 2 \sup_{\mathbf{v}\in\overset{\circ}{\mathbf{W}}{}^{1}_{2,\nu}(\Omega,\delta)} \frac{b_i(\mathbf{v},s)}{\|\text{div } \mathbf{v}\|_{L_{2,\nu}(\Omega)} + \|\text{curl } \mathbf{v}\|_{L_{2,\nu}(\Omega)}} \leqslant$$

$$\leqslant 2 \sup_{\mathbf{v}\in\mathbf{U}_\nu(\Omega,\delta)} \frac{b_i(\mathbf{v},s)}{\|\text{div } \mathbf{v}\|_{L_{2,\nu}(\Omega)} + \|\text{curl } \mathbf{v}\|_{L_{2,\nu}(\Omega)}} \leqslant 2\sqrt{1+8C_\Omega^2} \sup_{\mathbf{v}\in\mathbf{U}_\nu(\Omega,\delta)} \frac{b_i(\mathbf{v},s)}{\|\mathbf{v}\|_{\mathbf{U}_\nu(\Omega)}}.$$

The estimate of Theorem 1 is obtained. □

Conclusions

The concept of R_ν-generalized solution for the Stokes problem with model boundary conditions in a polygonal non-convex domain with a re-entrant corner on the boundary in weighted sets is defined in the paper. In this case, the variational formulation of the problem is not symmetric. Weighted analogue of the Ladyzhenskaya–Babushka–Brezzi conditions in special norms of weighted spaces is established.

The research was carried out within the state assignment for IAM FEB RAS (no. 075-00459-24-00).

References

[1] D.Boffi, F.Brezzi, M.Fortin, Mixed Finite Element Methods and Applications, Springer, Berlin/Heidelberg, 2013.

[2] R.Temam, Navier-Stokes equations. Theory and numerical analysis, North-Holland, Amsterdam, 1984.

[3] P.M.Gresho, R.L.Sani, On pressure boundary conditions for incompressible Navier-Stokes equations, *Internat J. Numer. Methods Fluids*, **7**(1987), 1111–1145. DOI: 10.1002/fld.1650071008

[4] M.A.Ol'shanskii, On the Stokes problem with model boundary conditions, *Sb. Math.*, **188**(1997), 4, 603–620. DOI: 10.1070/SM1997v188n04ABEH000220

[5] M.Dauge, Stationary Stokes and Navier–Stokes system on two- or three-dimensional domains with corners. I. Linearized equations, *SIAM J. Math. Anal.*, **20**(1989), 74–97. DOI: 10.1137/0520006

[6] V.Girault, P.A.Raviart, Finite element methods for Navier-Stokes equations. Theory and algorithms, Berlin-Heidelberg-New, York-Tokyo: Springer-Verlag, 1986.

[7] V.A.Rukavishnikov, Differential properties of an R_ν-generalized solution of the Dirichlet problem, *Soviet Mathematics Doklady*, **40**(1990), 653–655.

[8] V.A.Rukavishnikov, A.O.Mosolapov, E.I.Rukavishnikova, Weighted finite element method for elasticity problem with a crack, *Comput. Struct.*, **243**(2021), 106400. DOI: 10.1016/j.compstruc.2020.106400

[9] V.A.Rukavishnikov, E.I.Rukavishnikova, Weighted finite element method and body of optimal parameters for elasticity problem with singularity, *Comput. Math. Appl.*, **151**(2023), 408–417. DOI: 10.1016/j.camwa.2023.10.021

[10] V.A.Rukavishnikov, A.V.Rukavishnikov, Theoretical analysis and construction of numerical method for solving the Navier-Stokes equations in rotation form with corner singularity, *J. Comput. Appl. Math.*, **429**(2023), 115218. DOI: 10.1016/j.cam.2023.115218

[11] V.A.Rukavishnikov, A.V.Rukavishnikov, Weighted finite element method for the Stokes problem with corner singularity, *J. Comput. Appl. Math.*, **341**(2018), 144–156. DOI: 10.1016/j.cam.2018.04.014

[12] V.A.Rukavishnikov, A.V.Rukavishnikov, The method of numerical solution of the one stationary hydrodynamics problem in convective form in L-shaped domain, *Comput. Res. Model.*, **12**(2020), 1291–1306 (in Russian). DOI: 10.20537/2076-7633-2020-12-6-1291-1306

[13] V.A.Rukavishnikov, A.V.Rukavishnikov, On the properties of operators of the Stokes problem with corner singularity in nonsymmetric variational formulation, *Mathematics*, **10**(2022), 6, 889. DOI: 10.3390/math10060889

[14] A.V.Rukavishnikov, V.A.Rukavishnikov, New numerical approach for the steady-state Navier–Stokes equations with corner singularity, *Int. J. Comput. Methods*, **19**(2022), 9, 2250012. DOI:10.1142/S0219876222500128

Весовой аналог LBB-условий для решения задачи Стокса с модельными граничными условиями в области с сингулярностью

Алексей В. Рукавишников
Институт прикладной математики
Дальневосточное отделение Российской академии наук
Хабаровск, Российская Федерация

Аннотация. В работе определено понятие R_ν-обобщённого решения задачи Стокса с модельными граничными условиями в области с угловой сингулярностью. Доказан весовой аналог условий Ладыженской–Бабушки–Брецци в области с входящим углом.

Ключевые слова: угловая особенность, задача Стокса с модельными граничными условиями, R_ν-обобщённое решение.

EDN: PJYJQI
УДК 537.9

Diffusion and Electromigration of Ions – Products of the Proton Exchange Reaction in a Benzoic Acid Melt

Vitaly A. Demin*
Maxim I. Petukhov†
Perm State University
Perm, Russian Federation

Received 10.08.2024, received in revised form 15.09.2024, accepted 24.10.2024

Abstract. The work is devoted to a numerical study of the transport of proton exchange reaction products in a benzoic acid melt after the interaction of its molecules with a lithium niobate crystal. Due to dissociative adsorption from the surface of the substrate, positive lithium ions and negative benzoate ions diffuse into the acid. The transfer of these reaction products is described using equations in the continuos media approximation. The mathematical model takes into account the diffusion and electromigration mechanisms of transport, as well as the recombination of ions. As a result of the solution, stationary distributions of ion concentrations are obtained. Due to the large difference in the kinetics of the reaction products, benzoate ions are grouped predominantly near the substrate, while lithium ions tend to move away from it to a much greater distance. The work shows that the size of the computational domain approaches the size of the reactor working space when the ions of both types form boundary layers.

Keywords: proton exchange, boundary layer, numerical simulation.

Citation: V.A. Demin, M.I. Petukhov, Diffusion and Electromigration of Ions – Products of the Proton Exchange Reaction in a Benzoic Acid Melt, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 100–108. EDN: PJYJQI.

Introduction

The technology of enriching lithium niobate or lithium tantalate crystals with protons has been actively used in the manufacturing of waveguides over the past decades. This technology was first presented in the works [1, 2]. Its essence lies in the fact that a crystal sample is placed in a benzoic acid melt at characteristic temperatures ~ 500 K. As a result a complex chemical reaction occurs on the substrate surface. Dissociation of acid molecules after their interaction with the substrate leads to origin of negative benzoate ions and positive hydrogen ions on the crystal surface. Protons tend to penetrate the crystal and replace lithium ions, which diffuse from the near-surface region of the crystal into the acid. After some time, lithium ions recombine with the benzoate ions remaining in the acid. As a result the lithium benzoate molecules are formed. Thus, the protons penetrate into the crystal because of dissociative adsorption. As experiments show, protonation with duration $\sim 1-2$ h leads to the formation of a proton-enriched layer in the substrate with a thickness of $\sim 5-6$ μm.

Active implementation of this technology in the production of channel and planar waveguides has led to the emergence of a large number of its modifications [3]. Thus, to regulate the

*demin@psu.ru
†geniusmaxp@yandex.ru

protonation rate, impurities are used, added both to benzoic acid [4, 5] and to the crystal [6] before protonation. The presence of protons in the crystal changes its refractive index. Numerous experiments show that the refractive index profiles, as well as the proton concentration profiles in the formed near-surface layer of the crystal, have a characteristic stepped shape [1, 7]. These profiles can be smoothed by additional annealing after proton exchange [8]. Many works are devoted to the study of the complex phase structure of the crystal after its enrichment with protons [9] and the effect of the conditions of proton exchange on this structure. For example, the surfaces of samples are subjected to preliminary plasma treatment [10]. At present, special attention is paid to the protonation of thin-film lithium niobate [11, 12].

Theoretical investigation of the protonation process began with the work of Vohra [7], devoted to the study of hydrogen ion transport inside the crystal. The results of this work demonstrated a strong influence of the ion flux nonlinearity in the transport equation during description the process of lithium ion substitution by protons. In turn, the annealing process is described using the classical diffusion equation.

According to other works, the diffusion of lithium ions and benzoate ions into the benzoic acid melt during protonation leads to the formation of a stationary boundary layer by benzoate ions [13, 14]. This investigation shows that the transport and subsequent recombination of the reaction products can be described using equations in the continuous media approximation. Most of the lithium benzoate is produced in thin near-surface layer in the melt. It is known from experiments that the presence of this impurity reduces the intensity of proton exchange even at a small content of lithium benzoate in the melt [5]. Clarification of the characteristics of the resulting boundary layer allows us to understand possible methods for controlling the proton exchange process.

In the mentioned work [14] the size of the computational domain was comparable with the thickness of the photolithographic mask ($\sim 2\ \mu$m), which, on the one hand, is sufficient to qualitatively show the presence of the boundary layer, but, on the other hand, can greatly underestimate its size. Lithium ions, having a mobility an order of magnitude greater than benzoate ions, are capable of moving away from the protonated surface. Then lithium ions reach the opposite impenetrable boundary of the computational domain and, due to more intensive recombination, prevent the diffusion of benzoate ions deep into the melt. This circumstance is the reason why the transport of reaction products must be considered in regions which size is much larger than the characteristic thickness of the photolithographic mask.

In the work [15] an asymptotic solution is obtained, which allows to estimate the thickness of the boundary layer in the case when the size of the region under consideration are comparable with the size of the reactor in which the proton exchange takes place. It should be noted that, despite more realistic problem geometry, in this work a strong assumption is made. It consist in neglecting the influence of the electric field formed in the region occupied by ions on their concentration profiles. Because of this, according to the obtained analytical solution, the thickness of the boundary layer grows indefinitely with the size of the reactor according to a power law.

In this paper, a more complete mathematical formulation is considered in order to obtain realistic characteristics of the boundary layer formed by ions.

1. Formulation of the problem

Let us consider a benzoic acid melt in contact with a homogeneous surface of a lithium niobate crystal. The temperature of this melt is assumed to be constant and equal to 500 K, which is sufficient for proton exchange. Thus, lithium ions and benzoate ions should diffuse the melt from the crystal side. Diffusion, electromigration, and recombination of ions are described using the following system of equations in the continuous media approximation [16–18]:

$$\mathrm{div}(\varepsilon_0\varepsilon\boldsymbol{E}) = e(n_+ - n_-), \qquad \boldsymbol{E} = -\nabla\varphi, \tag{1}$$

$$\frac{\partial n_\pm}{\partial t} = D_\pm \Delta n_\pm \mp \nabla(k_\pm n_\pm \boldsymbol{E}) - k_R n_+ n_-, \tag{2}$$

where $\boldsymbol{E}$, φ, $n_\pm$ are the fields of electric field strength, electric potential, and concentrations of positive and negative ions. Concentration determines the number of ions per unit volume and has a unit of measurement of m^{-3}. Parameters e, ε, ε_0 are the electron charge, the permittivity of benzoic acid, and the electric constant. The diffusion coefficients, the mobility of lithium and benzoate ions, and the recombination coefficient are denoted, correspondingly, by $D_\pm$, $k_\pm$, and k_R.

The coefficient of ion mobility relates the speed of their drift in the melt to the intensity of the external electric field [19]. The estimation of diffusion and ion mobility coefficients for the conditions of the present problem, as well as the comparison of these values with known experimental data, were carried out in accordance with [13, 19, 20].

According to the results obtained in [14], in the formulation under consideration there are no conditions for convective mass transfer [21], which allows us to describe the ion transport in a one-dimensional formulation. After eliminating the electric field strength, the system of equations (1)–(2) will have the form:

$$\varepsilon_0\varepsilon\varphi'' = e(n_- - n_+), \tag{3}$$

$$\frac{\partial n_\pm}{\partial t} = D_\pm n''_\pm \pm k_\pm(n'_\pm\varphi' + n_\pm\varphi'') - k_R n_+ n_-. \tag{4}$$

The prime denotes the derivative with respect to the x coordinate. The coordinate axis is directed along the normal to the crystal toward the melt. The interphase boundary is being considered as the origin of coordinates. The boundary conditions on it relate the derivative of the concentration to the ion flux density J, and also determine the reference point for the electric potential:

$$x = 0: \qquad n'_\pm = -\frac{J}{D_\pm}, \qquad \varphi = 0. \tag{5}$$

At a distance h from the crystal, an impenetrable reactor wall is modeled, on which, in addition, there is no electric field:

$$x = h: \qquad n'_\pm = 0, \qquad \varphi' = 0. \tag{6}$$

The combination of boundary conditions (5)–(6) ensures the electrical neutrality of the system:

$$\int_0^h n_+ dx = \int_0^h n_- dx. \tag{7}$$

The boundary value problem (3)–(7) was solved in dimensionless variables. The units of length, time, concentration and electric potential were taken to be h, h^2/D_+, Jh/D_+ and $eJh^3/\varepsilon_0\varepsilon D_+$, correspondingly. In these variables, the system of equations and boundary conditions are written in the following form:

$$\varphi'' = n_- - n_+, \tag{8}$$

$$\frac{\partial n_\pm}{\partial t} = A_\pm n''_\pm \pm B_\pm (n'_\pm \varphi' + n_\pm \varphi'') - C n_+ n_-, \tag{9}$$

$$x = 0: \qquad n'_\pm = -\frac{D_+}{D_\pm}, \qquad \varphi = 0, \tag{10}$$

$$x = 1: \qquad n'_\pm = 0, \qquad \varphi' = 0, \tag{11}$$

where

$$A_\pm = \frac{D_\pm}{D_+}, \qquad B_\pm = \frac{k_\pm e J h^3}{D_+^2 \varepsilon_0 \varepsilon}, \qquad C = \frac{k_R J h^3}{D_+^2}. \tag{12}$$

When varying the dielectric constant, it is taken into account that the recombination coefficient also changes with it [17]:

$$k_R = \frac{e(k_+ + k_-)}{\varepsilon_0 \varepsilon}. \tag{13}$$

2. Solution technique

To solve the boundary value problem (8)–(11), an explicit finite-difference scheme was used, implemented using a program written in the C++ programming language. The Poisson equation (8) was solved using the Liebman [22] scheme. Zero concentration and electric potential fields were used as initial conditions: $n_\pm(t = 0, x) = \varphi(t = 0, x) = 0$. The following values of dimensional parameters were used in the calculations: $k_+ = 1.5 \cdot 10^{-7}$ m^2/s·V, $k_- = 2 \cdot 10^{-8}$ m^2/s·V, $D_+ = 10^{-8}$ m^2/s, $D_- = 10^{-9}$ m^2/s, $J = 10^{18}$ s^{-1}m^{-2}. The value of the permittivity ε varied in the range from 1 to 20 [14]. In turn, the sizes of the computational domain h were taken in the range from $2 \cdot 10^{-6}$ to 10^{-4} m.

Based on the results obtained earlier in [15], equations (4) have an analytical solution in the case when $k_\pm = 0$. Using the multiple scales method, we can determine the boundary layer thickness δ, which is represented in the following form:

$$\delta = \frac{5h^{\frac{1}{4}}}{\left(\frac{k_R J}{D_-}\left(\frac{1}{D_-} - \frac{1}{D_+}\right)\right)^{\frac{1}{4}}}. \tag{14}$$

It follows from this formula that for the values of h used in the present calculations the size of the boundary layer will be 1–2 orders of magnitude smaller than the size of the computational domain. In order to minimize the numerical errors that potentially arise in the thin region occupied by the boundary layer, and at the same time to save computer time, it is necessary to thicken the computational grid near the origin.

The grid thickening was carried out according to the many-stage scheme. After determine the size of the computational domain h, a uniform grid was formed, the nodes of which were located at points with coordinates x_i. Then, the variable step h_i was calculated for a non-uniform grid with the same number of nodes:

$$h_i = a + b x_i^c. \tag{15}$$

Thus, the computational domain was divided into 201 computational nodes, with a step of h_i. The values of the coefficients a, b and index c were selected in such a way that there were at least 100 nodes per ion boundary layer. The specificity of distribution (15) allows the grid nodes to be thickened in such a way that the grid step in the region occupied by the boundary layer is constant. An example of grid step distribution is shown in Fig. 1a.

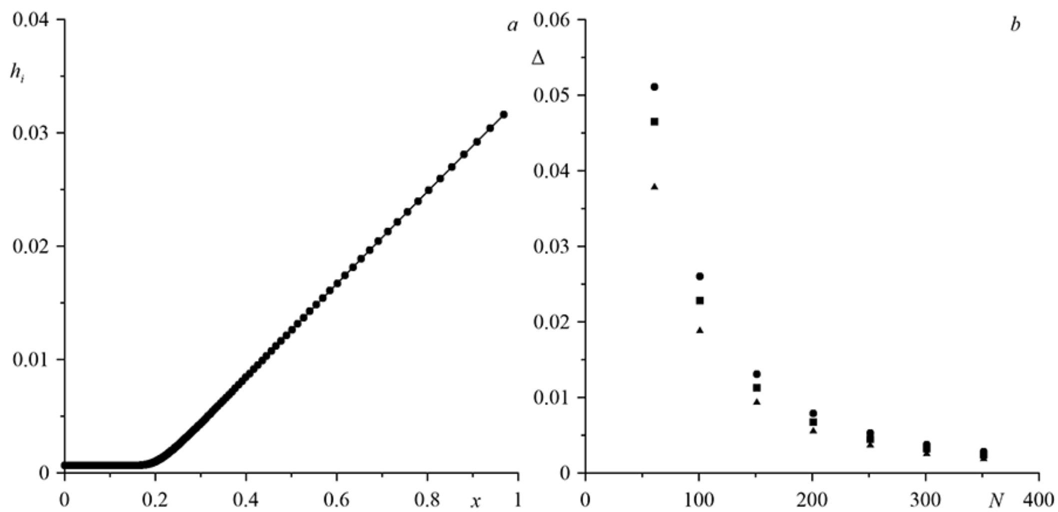


Fig. 1. Example of grid step distribution depending on coordinate for $a = 0.1$, $b = 5$, $c = 16$ (a); relative error Δ of the boundary layer thickness calculation (triangles) and the concentration of positive and negative ions in the center of the computational domain (circles and squares, correspondingly) depending on the number of nodes N (b)

When representing the system of equations (8)–(9) in finite difference form, the partial derivatives with respect to the coordinate x were calculated as follows:

$$\frac{\partial u}{\partial x} = \frac{u_{i+1} - u_{i-1}}{h_i + h_{i+1}}, \tag{16}$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{2}{h_i + h_{i+1}} \left(\frac{u_{i+1} - u_i}{h_{i+1}} - \frac{u_i - u_{i-1}}{h_i} \right). \tag{17}$$

When calculating the spatial derivatives in the boundary conditions of the second kind (10)–(11), as well as the time derivative in equations (9), a one-sided difference was used:

$$\frac{\partial u}{\partial \tilde{x}} = \frac{u_{i+1} - u_i}{\tilde{h}}, \tag{18}$$

where $\tilde{x}$ is a generalized designation of the differentiation variable. In turn, $\tilde{h}$ is a generalized designation of the step along the coordinate at the side points of the computational domain, as well as the time step. An example of the convergence of the implemented scheme is shown in Fig. 1b. Expressions (16)–(18) are first-order formulas.

3. Results of numerical simulation

Solution of the boundary problem (8)–(11) gives stationary profiles of concentration and electric potential in the benzoic acid melt. Their characteristic form is shown in Fig. 2. The time necessary to originate stationary profiles in the model case, when the only ion transport mechanism is diffusion, is ~ 0.02 s. In turn, the reverse effect of the electric field on the concentration profiles additionally stabilizes the system and reduces this time up to ~ 0.01 s. It should be noted that in this case the process of stationary profiles origination is longer compared to one discussed in [14], since the size of the computational domain is significantly larger.

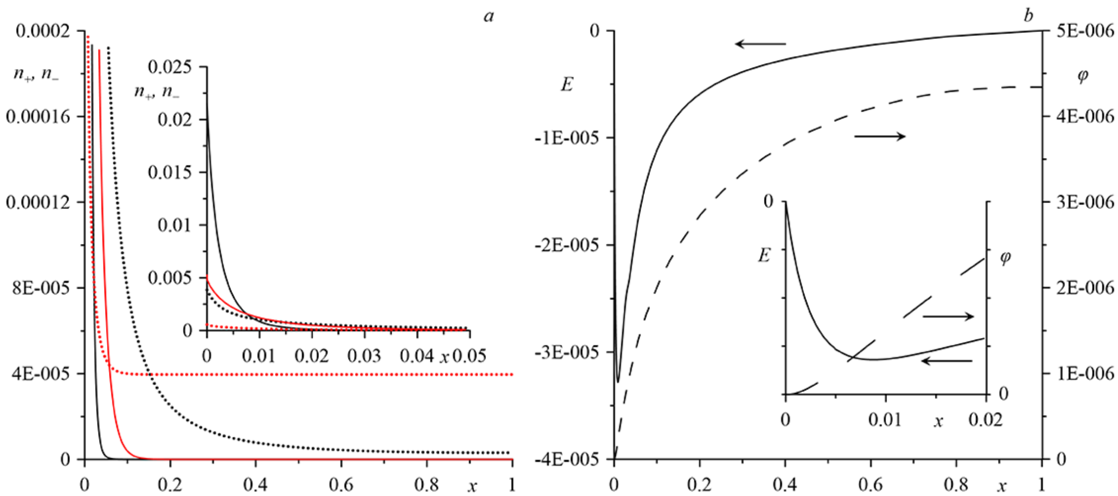


Fig. 2. Profiles of concentration (a), field strength and electric potential (b) at $h = 8 \cdot 10^{-5}$ m and $\varepsilon = 10$. Black and red curves correspond to the cases with and without electromigration, correspondingly. a) Solid lines — concentration of benzoate ions, dots — concentration of lithium ions. b) Solid lines — electric field strength, dashed lines — electric potential

Depending on whether electromigration is taken into account or not, a fundamental difference in the profiles is noticeable from Fig. 2a. In the purely diffusion case ($k_\pm = 0$), lithium ions reach the opposite boundary one way or another by the time of stationary profiles origination. These ions are present everywhere in the calculation region albeit in small concentrations. Including electromigration in the model allows the electric field formed, the strength of which is directed toward the crystal, to affect the ions. Thus, the lithium ion profile is slightly "pressed" to the interphase boundary. As calculations have shown, in the case of large h, the concentration of positive ions reaching the reactor wall tends to zero. Thus, concentration profile of these ions is qualitatively no different from the concentration profile of benzoate ions.

Having significantly lower mobility than lithium ions, benzoate ions are not able to reach the reactor wall in any case. These ions always form a boundary layer, the thickness of which varies depending on how far the lithium ions can move away from the interface. In other words, taking into account electromigration in the case where the size of the computational domain is large enough leads to the formation of a coupled boundary layer.

Fig. 2b shows the profiles of the electric potential and electric field strength. Comparing these profiles with the concentration profiles, it is easy to see that the electric field is most intense in the region of the benzoate ion boundary layer. The strength locally reaches $3 \cdot 10^4$ V/m. The electric field strength on the crystal surface is considered to be zero. This should be true in the case where the X-cut of lithium niobate is protonated. Despite the fact that the condition on the potential (5) is set at the crystal-melt boundary (but not on its derivative), the zero value of the strength is obtained in the solution process.

The characteristic thickness of the benzoate ion boundary layer depending on the size of the domain is shown in Fig. 3. It is evident that the results of solving the diffusion-recombination problem qualitatively repeat the analytical solution [15] and the layer thickness grows indefinitely with increasing h. In turn, electromigration does not allow lithium ions to move away from the crystal at an arbitrarily large distance, therefore, at a certain value of $h^* \approx 2 \cdot 10^{-5}$ m, they form a boundary layer. At $h > h^*$, the characteristics of the boundary layers remain the same as in the case of $h = h^*$. According to this result one can say that h^* corresponds to Debye length of the investigated system.

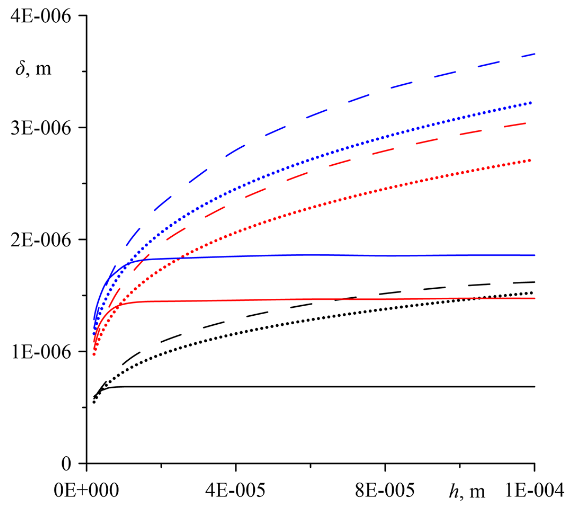


Fig. 3. Dimensional values of boundary layer thickness for the case of different permittivities. Dots and dashed lines are the results of analytical [15] and numerical solutions for $k_{\pm} = 0$; solid lines are the numerical solution with account of electromigration. Black, red and blue curves correspond to the case of $\varepsilon = 1$, 10 and 20, correspondingly

It is interesting to note that the effect of the electric field on the concentration profiles not only "presses" the positive lithium ion profiles toward the crystal, but also reduces the size of the region in which negative benzoate ions are present. The primary cause of this effect is a more intense recombination associated with an increase in the concentration of lithium ions affected by the electric field in the region near the crystal occupied by benzoate ions. Since the flux density of positive and negative ions from the crystal surface is the same, the total number of ions present in the melt should decrease, causing the boundary layer to shift toward the crystal.

Conclusion

As calculations have shown, after diffusion of lithium ions and benzoate ions into the benzoic acid melt, ion transport complicated by electromigration and recombination takes place. Due to the large difference in the kinetics of negative and positive ions participating in the proton exchange, the concentration profiles differ considerably from each other and an electric field is formed near the interphase. Taking into account the reverse effect of this field on ions greatly changes the concentration profiles in the case when the size of the computational domain exceed $2 \cdot 10^{-5}$ m. Analysis of the behavior of ions in a cavity with sufficiently large sizes allows us to establish that not only benzoate ions, but also lithium ions form a stationary boundary layer. The sizes of these structures remain unchanged with further growth of the computational domain and are $2 \cdot 10^{-6}$ m for benzoate ions and $2 \cdot 10^{-5}$ m for lithium ions, correspondingly.

The work was supported by the Russian Science Foundation (Grant no. 24-29-20277) and Perm Region.

References

[1] J.L.Jackel, C.E.Rice, J.J.Veselka, Proton exchange for high-index waveguides in $LiNbO_3$, *Appl. Phys. Lett.*, **41**(1982), 607–608.

[2] J.L.Jackel, Proton exchange: past, present, and future, *Proc. SPIE*, **1583**(1991), 54–63.

[3] M.Kuneva, Optical waveguides obtained via proton exchange technology in $LiNbO_3$ and $LiTaO_3$ – a short review, *International Journal of Scientific Research in Science and Technology*, **2**(2016), 40–50.

[4] S.S.Mushinsky, A.M.Minkin, I.V.Petukhov et al., Water Effect on Proton Exchange of X-cut Lithium Niobate in the Melt of Benzoic Acid, *Ferroelectrics*, **476**(2015), 84–93. DOI: 10.1080/00150193.2015.998530

[5] V.I.Kichigin, I.V.Petukhov et al., Structure and properties of proton exchange waveguides on Z cut of lithium niobate crystal fabricated in molten benzoic acid with the addition of lithium benzoate, International Conference and Seminar of Young Specialists on Micro/Nanotechnologies and Electron Devices, 2012, 238–241.

[6] A.V.Sosunov, S.S.Mushinsky et al., Evalua-tion of applicability of lithium niobate crystals Z-cut with predetermined impurity distribution for manufacturing of proton-exchanged waveguides, *Bulletin of Perm University. Physics*, **2**(2017), 69–73. DOI: 10.17072/1994-3598-2017-2-69-74

[7] S.T.Vohra, A.R.Mickelson, S.E.Asher, Diffusion characteristics and waveguiding properties of proton-exchanged and annealed $LiNbO_3$ channel waveguides, *J. Appl. Phys.*, **66**(1989), 5161–5174. DOI: 10.1063/1.343751

[8] M.De Micheli, J.Botineau, S.Neveu, P.Sibillot, D.B.Ostrowsky, Independent control of index and profiles in proton-exchanged lithium niobate guides, *Optics Lett.*, **8**(1983), 114–115. DOI: 10.1364/ol.8.000114

[9] Yu.N.Korkishko, V.A.Fedorov, Structural phase diagram of $H_xLi_{1-x}NbO_3$ waveguides: the correlation between optical and structural properties, *IEEE J. Sel. Top. Quantum Electron*, **2**(1996), 187–196. DOI: 10.1109/2944.577359

[10] I.V.Petukhov, V.I.Kichigin, S.S.Mushinskii, D.I.Sidorov, O.R.Semenova, The influence of plasma treatment of lithium niobate crystal surface on the proton exchange process in molten benzoic acid, *Bulletin of Perm University. Chemistry*, **9**(2019), 371-379. DOI: 10.17072/2223-1838-2019-4-371-379

[11] A.A.Kozlov, U.O.Salgaeva, V.A.Zhuravlev, A.B.Volyntsev, The study of the kinetics of thin-film lithium niobate reactive ion etching in a fluorine-containing plasma, *Bulletin of Perm University. Physics*, **1**(2024), 56–71.

[12] Y.Li, T.Lan, D.Yang, Z.Wang, Fabrication of ridge optical waveguide in thin film lithium niobate by proton exchange and wet etching, *Optical Materials*, **120**(2021), 111433. DOI: 10.1016/j.optmat.2021.111433

[13] V.A.Demin, M.I.Petukhov, R.S.Ponomarev, An ionic boundary layer near the lithium niobate surface in the proton exchange process, *Surface Engineering and Applied Electrochemistry*, **59**(2023), 321–328. DOI: 10.3103/S1068375523030055

[14] V.A.Demin, M.I.Petukhov, R.S.Ponomarev, M.Kuneva, Effect of Permittivity on the Ionic Boundary Layer upon Protonation of Lithium Niobate, *Journal of Siberian Federal University. Mathematics and Physics*, **16(5)**(2023), 611–619. EDN: GVVEKJ

[15] V.A.Demin, M.I.Petukhov, Application of multiple scales method to the problem about characteristics of the ionic layer near the surface of lithium niobate crystal in a benzoic acid melt, *Microgravity Science and Technology*, **36, 33**(2024). DOI: 10.1007/s12217-024-10113-z

[16] L.D. Landau, E.M. Lifshitz, Fluid Mechanics, Vol 6, Butterworth-Heinemann, 1987.

[17] F.Pontiga, A.Castellanos, Physical mechanisms of instability in a liquid layer subjected to an electric field and a thermal gradient, *Phys. Fluids*, **6**(1994), 1684.

[18] E.A.Demekhin, N.V.Nikitin, V.S.Shelistov, Direct numerical simulation of electrokinetic instability and transition to chaotic motion, *Phys. Fluids*, **25**(2013), 122001. DOI: 10.1063/1.4843095

[19] G.I.Skanavi, Fizika dielektrikov. Oblast' slabyh polej, *Moskva, Gosudarstvennoe izdatel'stvo tekhniko-teoreticheskoj literatury*, 1949 (in Russian).

[20] I.K.Kikoin, Tablicy fizicheskih velichin, Moskva, Atomizdat, 1976) (in Russian).

[21] G.Z.Gershuni, E.M.Zhukhovitskii, Convective stability of incompressible fluids, Jerusalem, Keter Publishing House, 1976.

[22] P.Roache, Computational fluid dynamics, Albuquerque, Hermosa publishers, 1976.

Диффузия и электромиграция ионов – продуктов реакции протонного обмена в расплаве бензойной кислоты

Виталий А. Демин
Максим И. Петухов
Пермский государственный университет
Пермь, Российская Федерация

Аннотация. Работа посвящена численному исследованию транспорта продуктов реакции протонного обмена в расплаве бензойной кислоты после взаимодействия ее молекул с кристаллом ниобата лития. Вследствие диссоциативной адсорбции с поверхности подложки в кислоту проникают положительные ионы лития и отрицательные бензоат-ионы. Перенос данных продуктов рекции описывается при помощи уравнений в приближении сплошной среды. В математической модели учитываются диффузионный и электромиграционный механизмы транспорта, а также рекомбинация ионов. В результате решения получаются стационарные распределения концентрации ионов. Из-за большой разницы в кинетике продуктов реакции бензоат-ионы группируются преимущественно вблизи подложки, в то время как ионы лития стремятся отдалиться от нее на гораздо большее расстояние. В работе показано, что при устремлении размеров расчетной области к размерам рабочего пространства реактора, ионы обоих типов формируют пограничные слои.

Ключевые слова: протонный обмен, пограничный слой, численное моделирование.

Journal of Siberian Federal University. Mathematics & Physics 2025, 18(1), 109–118

EDN: UVZXKX
УДК 517.10

On Weakly Contractions Via w-distances

Hossein Lakzian[*]
Department of Mathematics
Payame Noor University
Tehran, Iran

Sedigheh Barootkoob[†]
Faculty of Basic Sciences
University of Bojnord
Bojnord, Iran

Nicola Fabiano[‡]
"Vinča" Institute of Nuclear Sciences – National Institute of the Republic of Serbia
University of Belgrade
Belgrade, Serbia

Stojan Radenović[§]
Faculty of Mechanical Engineering
University of Belgrade
Belgrad, Serbia

Received 10.08.2024, received in revised form 27.09.2024, accepted 29.11.2024

Abstract. In this article, we will check whether the known results remain valid if the metric d is replaced by the w-distance p. we show that in some contractive conditions where w-distance p participates instead of metric d, symmetry of w-distance p can be assumed and the proofs can be shorter. We are talking about results such as Banach's contraction principle, Kannan's theorem, Boyd–Wong, Meir–Keeler, Chatterje's, Reich's, Hardy–Rogers', Karapinars' and Wardowskis' theorems and many others.

By doing so, we would obtain generalizations of the above results.

Keywords: fixed point, w-distance, p-interpolative Kannan type contraction, p-Hardy–Rogers contraction, $(\mathcal{F}, p)$-contraction.

Citation: H. Lakzian, S. Barootkoob, N. Fabiano, S. Radenović, On Weakly Contractions Via w-distances, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 109–118. EDN: UVZXKX.

1. Introduction and preliminaries

One of the generalizations of the well-known Banach theorem from 1922 is the introduction after 75 years of the so-called w-distance p in the given metric space (X, d). Thus we obtained an ordered triple (X, d, p) where (X, d) is the given metric space and p is a function from $X \times X$ in $[0, +\infty)$ that satisfies the following three axioms:

p1) *$p(x, z) \leqslant p(x, y) + p(y, z)$ for all $x, y, z \in X$;*

p2) *For any $x \in X$, the function $p(x, \cdot) : X \to [0; +\infty)$ is $d-$lower semi-continuous;*

p3) *For any $\varepsilon > 0$, there exists $\delta > 0$ such that $p(z, x) \leqslant \delta$ and $p(z, y) \leqslant \delta$ imply $d(x, y) \leqslant \varepsilon$. Then, p is called a w-distance on X.*

[*]h-lakzian@pnu.ac.ir
[†]s.barutkub@ub.ac.ir
[‡]nicola.fabiano@gmail.com
[§]sradenovic@mas.bg.ac.rs

The last two axioms are new compared to those known for metric space. The second represents the lower semi-continuity in the second variable and the third connects the metric d and the w-distance p.

A typical example of a w-distance is the metric d itself defined on a nonempty set X. Actually, **p1**) is fulfilled as a triangle relation. Since the metric d is a continuous function with 2 variables, it is also semi-continuous from below in the second variable. Indeed, if y_n is a sequence in X that converges to y by metric d then $p(x,y) = d(x,y) = \lim_{n\to+\infty} d(x,y_n) = \liminf_{n\to+\infty} d(x,y_n)$ and **p2**) is fulfilled. Assuming that $\delta = \frac{\varepsilon}{2}$, we get that **p3**) is fulfilled, because $p(x,y) = d(x,y) \leqslant d(x,z) + d(y,z) \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$.

Now we list several typical examples of w-distances, some of which were also mentioned in the first paper on w-distances.

1. Let (X,d) be a metric space. Then a function $p : X \times X \to [0,+\infty)$ defined by $p(x,y) = c$ for every $x,y \in X$ is a w-distance on X, where c is a positive real number.

2. Let X be a normed linear space with norm $\|\cdot\|$. Then a function $p : X \times X \to [0,+\infty)$ defined by $p(x,y) = \|x\| + \|y\|$ for every $x,y \in X$ is a w-distance on X.

3. The similar as example 3. only the function $p : X \times X \to [0,+\infty)$ is defined by $p(x,y) = \|x\|$ for all $x,y \in X$.

For several examples of w-distances see [2], pages 382, 383, 384.

An important note about **p1)** and **p3)**. Since Example 1.3. from [3] (see also Example 4 from [3]) shows that the w-distance in the general case is not symmetric, i.e., it is not $p(x,y) = p(y,x)$ for every $x,y \in X$, then the triangle relation as well as the axiom **p3)** should be understood as the introduced order x,z; x,y; y,z and, z,x; z,y and finally x,y.

The following Lemma is one of the most important that is used in the study of w-distance metric spaces. It relates metric convergence to w-distance convergence. It is also important because it gives us the information (a sufficient condition) when the sequence x_n is Cauchy in the metric space (X,d). In the sequel, we denote by $\mathbb{R}^+$, $\mathbb{R}$ and $\mathbb{N}$, the sets of positive real numbers, real numbers and natural numbers, respectively.

Lemma 1.1 ([2], Lemma 1.)**.** *Let X be a metric space with metric d and let p be a w-distance on X. Let $\{x_n\}$ and $\{y_n\}$ be sequences in X, let $\{\alpha_n\}$ and $\{\beta_n\}$ be sequences in $[0,+\infty)$ converging to 0, and let $x,y,z \in X$. Then the following hold:*

(i) *If $p(x_n,y) \leqslant \alpha_n$ and $p(x_n,z) \leqslant \beta_n$ for any $n \in \mathbb{N}$, then $y = z$. In particular, if $p(x,y) = 0$ and $p(x,z) = 0$, then $y = z$;*

(ii) *if $p(x_n,y_n) \leqslant \alpha_n$ and $p(x_n,z) \leqslant \beta_n$ for any $n \in \mathbb{N}$, then $\{y_n\}$ converges to z;*

(iii) *if $p(x_n,x_m) \leqslant \alpha_n$ for any $n,m \in \mathbb{N}$ with $m > n$, then $\{x_n\}$ is a Cauchy sequence;*

(iv) *if $p(y,x_n) \leqslant \alpha_n$ for any $n \in \mathbb{N}$, then $\{x_n\}$ is a Cauchy sequence.*

The similar as in the context of metric spaces ([1,6]) we recall the following two lemmas that we will use in the proofs of our results. These both lemmas are important and are used to prove the Cauchyness of the sequence $x_n = fx_{n-1}, n \in \mathbb{N}$.

Lemma 1.2. *Let $\{u_n\}$ be a Picard sequence in metric space (X,d) with the w-distance p such that*

$$p(u_{n+1},u_n) < p(u_n,u_{n-1}) \tag{1}$$

or

$$p(u_n,u_{n+1}) < p(u_{n-1},u_n) \tag{2}$$

in both cases for all $n \in \mathbb{N}$. Then $u_n \neq u_m$ whenever $n \neq m$.

Proof. Consider the case (1). Suppose on the contrary that $u_n = u_m$ for some $n < m$. Then, $u_{n+1} = fu_n = fu_m = u_{m+1}$, hence

$$p(u_{n+1},u_n) = p(u_{m+1},u_m) < p(u_m,u_{m-1}) < \cdots < p(u_{n+1},u_n),$$

we obtain a contradiction. For the case (2) the proof is the same. $\square$

Lemma 1.3. *Let (X, d) be a metric space with w-distance p and let $\{u_n\}$ be a sequence in X such that both $p(u_{n+1}, u_n)$ and $p(u_n, u_{n+1})$ tend to 0 as $n \to +\infty$. If $\{u_n\}$ is not a Cauchy sequence in metric space (X, d), then there exist $\varepsilon > 0$ and two sequences $\{m(k)\}$ and $\{n(k)\}$ of positive integers such that $n(k) > m(k) > k$ and the following sequences tend to ε^+ when $k \to +\infty$:*

$$\begin{gathered}\left\{p\left(u_{m(k)}, u_{n(k)}\right)\right\}, \left\{p\left(u_{m(k)}, u_{n(k)-1}\right)\right\}, \left\{p\left(u_{m(k)+1}, u_{n(k)}\right)\right\}, \\ \left\{p\left(u_{m(k)-1}, u_{n(k)+1}\right)\right\}, \left\{p\left(u_{m(k)+1}, u_{n(k)+1}\right)\right\}, \ldots \end{gathered} \tag{3}$$

or

$$\begin{gathered}\left\{p\left(u_{n(k)}, u_{m(k)}\right)\right\}, \left\{p\left(u_{n(k)-1}, u_{m(k)}\right)\right\}, \left\{p\left(u_{n(k)}, u_{m(k)+1}\right)\right\}, \\ \left\{p\left(u_{n(k)+1}, u_{m(k)-1}\right)\right\}, \left\{p\left(u_{n(k)+1}, u_{m(k)+1}\right)\right\}, \ldots \end{gathered} \tag{4}$$

Proof. Since $\{u_n\}$ is not a d-Cauchy sequence, from Lemma 1 (iii) of [2], it follows that $p(u_n, u_m)$ does not tend to 0 as $n, m \to +\infty$. This means that there exist $\varepsilon > 0$ and subsequences $\{n(k)\}, \{m(k)\}$ such that $m(k) > n(k) > k$ and

$$p\left(u_{n(k)}, u_{m(k)}\right) \geqslant \varepsilon \quad \text{and} \quad p\left(u_{n(k)-1}, u_{m(k)}\right) < \varepsilon.$$

Then, using the axiom **(p1)** and the fact that both $p(u_{n+1}, u_n)$ and $p(u_n, u_{n+1})$ tend to 0 as $n \to +\infty$ it follows, in the same way as in metric spaces (see for instance [6]) that the given sequences tend to ε^+. $\square$

The w-distance p is symmetric if $p(x, y) = p(y, x)$ for all $x, y \in X$. For such a w-distances, if $a \neq b$ then $p(a, b) > 0$, i.e., from $p(a, b) = 0$ follows $a = b$. In many contractive conditions, symmetry of the w-distance can be assumed. This is achieved by introducing a new function q from $X \times X$ to $[0, +\infty)$ defined by $q(x, y) = \max\{p(x, y), p(y, x)\}$. It is easy to show that q is a w-distance. Often when proving fixed point results in metric spaces with w-distance one finds that the condition $x \neq y$ implies that $p(x, y) > 0$. But many examples show that if $p(x, y)$ is not a symmetric w-distance that this need not be true. Such an example is: $X = [0, +\infty)$, $p(x, y) = y$. Indeed, $1 \neq 0$ while $p(1, 0) = 0$. So, for many contractive conditions, symmetry of the w-distance can be assumed. For this purpose, we introduce the following function: $q(x, y) = \max\{p(x, y), p(y, x)\}$. It is easily shown that it is a symmetric w-distance. The proof uses the fact that $p(x, y)$ is semi-continuous from below in the second variable if and only if $p(y, x)$ it is semi-continuous from below in the first variable. And then for the function q we have:

$$\begin{aligned} q(x, y_n) &= \max\{p(x, y_n), p(y_n, x)\} \geqslant \\ &\geqslant \max\left\{\lim \inf_{n \to +\infty} p(x, y_n), \lim \inf_{n \to +\infty} p(y_n, x)\right\} \geqslant \\ &\geqslant \max\{p(x, y), p(y, x)\} = \\ &= q(x, y), \end{aligned}$$

that is., $q(x, y) \leqslant \lim\inf_{n \to +\infty} q(x, y_n)$, i.e., $q(x, y)$ is semi-continuous from below in second variable.

In some contractive conditions where w-distance p participates instead of metric d, symmetry of w-distance p can be assumed. Namely, using the property that from $0 \leqslant a \leqslant b$ and $0 \leqslant c \leqslant d$ follows $\max\{a, c\} \leqslant \max\{b, d\}$, from which yields that the w-distance p, one moves to the above symmetric w-distance q.

The following results are natural in the framework of complete metric spaces with w-distance, and for mapping $T : X \to X$, the next will be assumed either T is continuous or the infimum of

the number set
$\{p(x,y)+p(x,Tx) : x \in X\}$ where y is not a fixed point of T, is positive.

$p-$Banach contraction theorem: $p(Tx,Ty) \leqslant k \cdot p(x,y), k \in (0,1)$

$p-$Kannan contraction: $p(Tx,Ty) \leqslant l \cdot [p(x,Tx)+p(y,Ty)], l \in \left(0,\frac{1}{2}\right)$

$p-$Chatterjea contraction: $p(Tx,Ty) \leqslant m \cdot [p(x,Ty)+p(y,Tx)], m \in \left(0,\frac{1}{2}\right)$

$p-$Reich contraction: $p(Tx,Ty) \leqslant A \cdot p(x,y) + B \cdot p(x,Tx) + C \cdot p(y,Ty)$, $A,B,C \geqslant 0$ and $A+B+C<1$

$p-$Hardy–Rogers contraction: $p(Tx,Ty) \leqslant A \cdot p(x,y) + B \cdot p(x,Tx) + C \cdot p(y,Ty) + D \cdot p(x,Ty) + E \cdot p(y,Tx)$, $A,B,C,D,E \geqslant 0$ and $A+B+C+D+E<1$

$p-$*Ćirić* (I) $p(Tx,Ty) \leqslant \lambda_1 \cdot \max\left\{p(x,y), \frac{p(x,Tx)+p(y,Ty)}{2}, \frac{p(x,Ty)+p(y,Tx)}{2}\right\}$, $\lambda_1 \in (0,1)$

$p-$*Ćirić* (II) $p(Tx,Ty) \leqslant \lambda_2 \cdot \max\left\{p(x,y), p(x,Tx), p(y,Ty), \frac{p(x,Ty)+p(y,Tx)}{2}\right\}$, $\lambda_2 \in (0,1)$

$p-$*Ćirić* (III) $p(Tx,Ty) \leqslant \lambda_3 \cdot \max\left\{p(x,y), \frac{p(x,Tx)+p(y,Ty)}{2}, p(x,Ty), p(y,Tx)\right\}$, $\lambda_3 \in (0,1)$

$p-$*Ćirić* (IV) $p(Tx,Ty) \leqslant \lambda_4 \cdot \max\{p(x,y), p(x,Tx), p(y,Ty), p(x,Ty), p(y,Tx)\}$, $\lambda_4 \in (0,1)$

$p-$Bryant contraction: $p(T^n x, T^n y) \leqslant r \cdot p(x,y), r \in (0,1), n \in \mathbb{N}$

If (X,d) is compact metric space and if

$p-$Nemytzki contraction: $p(Tx,Ty) < p(x,y)$ whenever $x \neq y$.

$p-$Browder contraction: $p(Tx,Ty) \leqslant \phi(p(x,y)), \phi$ nondecreasing and continuous from the right function from $(0,+\infty)$ into$(0,+\infty)$ such that $\phi(t) < t$

$p-$Boyd–Wong contraction: $p(Tx,Ty) \leqslant \phi(p(x,y)), \phi$ is a real function, upper semi-continuous from the right, satisfying $\phi(t) < t$ for $t > 0$.

$p-$Meir–Keeler contraction: For all $\varepsilon > 0$ there exists $\delta > 0$ such that

$\varepsilon \leqslant p(x,y) < \varepsilon + \delta$ implies $p(Tx,Ty) < \varepsilon$

Further there are Jungck, Fisher and many other contractions.

2. p-Hardy-Rogers contraction

In section 5 of [2] the authors proved one theorem and three corollaries. In all four results, they assume that the contraction coefficient k belongs to the set $[0,1)$ or $[0,1/2)$. It is easy to see that this is imprecise and that the assumption must be that k belongs to $(0,1)$ or $(0,1/2)$. This is the difference obtained when, under the known contractive conditions of metric spaces, the metric d is replaced by the w-distance p. For example, in Theorem 4 from [2], if $k=0$ is set, $p(Tx,T^2x)=0$ is obtained. Whence it does not have to follow as with metric spaces that then $Tx=T^2x$ because the equality $p(a,b)=0$ does not necessarily follow $a=b$.

In this article, we will check, among other things, whether the known theorems (results) remain valid if the metric d is replaced by the w-distance p. We are talking about results such as Banach's contraction principle, Kanan's theorem, Chatterje's, Reich's and Hardy–Rogers' theorems. Then the Boyd–Wong and Meir–Keeler theorems and many others.

By doing so, we would obtain generalizations of the above results because each metric is a w-distance. One of the following sufficient conditions may be used for the existence of a fixed point:

(i) The mapping of T from X to X is continuous;

(ii) The number set $\{p(x,y)+p(x,Tx) : x \in X,\ y \in X$ but such that y is different from $Ty\}$ has a positive infimum.

First, we will formulate and prove the Hardy–Rogers theorem within metric spaces with w-distance p.

Theorem 2.1. *Let (X,d,p) be a complete metric space with w-distance p, T a mapping from X to X and let there exist non-negative constants A,B,C,D and E such that $A+B+C+D+E<1$ and* *that for every x, y from X it holds* $p(Tx, Ty) \leqslant Ap(x, y) + Bp(x, Tx) + Cp(y, Ty) + Dp(x, Ty) + Ep(y, Tx)$.

Then T has a unique fixed point say $z \in X$ such that $p(z, z) = 0$ if at least one of the above conditions (i) or (ii) holds.

Proof. Let us first assume that z is a fixed point of the mapping T and show that then $p(z, z) = 0$. Then by putting in the contractive condition $x = y = Tx = Ty$ we get: $p(x, x) \leqslant (A + B + C + D + E)p(x, x) < p(x, x)$ which is not possible if $p(x, x) > 0$. In order to show the uniqueness of a possible fixed point of the mapping T, let us assume that there are two different fixed points of it u and v. Using the fact that according to the already shown $p(u, u) = p(v, v) = 0$, then based on that and putting $x = u$, $y = v$ in the contractive condition we get: $p(u, v) \leqslant Ap(u, v) + Bp(u, u) + Cp(v, v) + Dp(u, v) + Ep(v, u) = Ap(u, v) + Dp(u, v) + Ep(v, u)$ i.e., $(1 - A - D)p(u, v) \leqslant Ep(v, u)$. Similarly, we get that $(1 - A - D)p(v, u) \leqslant Ep(u, v)$. By taking the maximum of the left and right sides, we get $(1 - A - D)q(u, v) \leqslant Eq(u, v)$ where $q(a, b) = \max\{p(a, b), p(b, a)]$. If it is assumed that $q(u, v) > 0$, we get a contradiction with $A + B + C + D + E < 1$. Otherwise, from $q(u, v) = 0$ and since q is a symmetric w-distance, we conclude $u = v$.

The rest of the proof is very simialar to the one for the metric spaces. □

3. p-interpolative Kannan type contraction

In 1969, Kannan [4] proved the following fixed point theorem.

Theorem 3.1. *Let $f : X \to X$ be a Kannan contraction mapping, i.e., $d(fx; fy) \leqslant k(d(x; fx) + d(y; fy))$ for all $x; y \in X$ and some $0 \leqslant k < \frac{1}{2}$, of a f-orbitally complete metric space. Then f has a unique fixed point.*

Afterwards, T. Suzuki published a nice paper [7] in which generalized Kannan's result in two new ones. He introduced the concept of weakly Kannan contraction mappings and non-weakly Kannan contraction mappings. For more details, see Section 4 and 5 in that paper.

Recently, the concept of interpolative Kannan type contraction mappings was introduced by E. Karapinar in [5]; and he proved the following theorem:

Theorem 3.2. *Let (X, d) be a complete metric space. Suppose that $f : X \to X$ is a interpolative Kannan type contraction self-map; i.e. if there exist a constant $\alpha \in (0; 1)$ and $k \in [0; 1)$ such that either of the followings hold:*

$$d(fx; fy) \leqslant k[d(x; fx)]^{\alpha}[d(y; fy)]^{1-\alpha}; \tag{5}$$

for all $x, y \in X$ with $x \neq fx$. Then f has a unique fixed point in X.

In this section, we introduced the concept of weakly interpolative Kannan type contraction mappings and we will prove and generalize Karapinar's theorem in the setting of w-distances.

We know that a w-distance p is not symmetric; i.e. $p(x; y)$ is not equal to $p(y; x)$ in general. So, we can define weakly interpolative Kannan type contractions as follows.

Definition 3.1. *Let $(X; d)$ be a metric space. The mapping $f : X \to X$ is said to be a weakly interpolative Kannan type contraction or p-interpolative Kannan type contraction, if there exist a constant $\alpha \in (0; 1)$ and $k \in [0; 1)$ such that either of the followings hold for all $x; y \in X$:*

$$p(fx, fy) \leqslant k[p(fx, x)]^{\alpha}[p(fy, y)]^{1-\alpha}, \tag{6}$$

or

$$p(fx, fy) \leqslant k[p(fx, x)]^{\alpha}[p(y, fy)]^{1-\alpha}, \tag{7}$$

or

$$p\,(fx, fy) \leqslant k[p\,(x, fx)]^{\,\alpha}[p\,(y, fy)]^{1-\,\alpha}, \tag{8}$$

or

$$p\,(fx, fy) \leqslant k[p\,(x, fx)]^{\,\alpha}[p\,(fy, y)]^{1-\,\alpha}. \tag{9}$$

If $p = d$ then f is called interpolative Kannan type contraction [5].

The following example shows that the class of weakly interpolative Kannan type contraction is more than interpolative Kannan type contraction.

Example 3.1. *Let $X = \{x, y, z\}$. Consider the metric d and the w-distance p on X, as follows.*

$$d(x, x) = d(y, y) = d(z, z) = 0, \quad d(x, y) = d(y, x) = 3,$$

$$d(x, z) = d(z, x) = 1, \quad d(y, z) = d(z, y) = 2.$$

$$p(x, x) = p(z, z) = 1, \quad p(y, y) = 0, \quad p(z, y) = \frac{3}{2}, \quad p(y, z) = 3;$$

$$p(y, x) = p(x, y) = 2, \quad p(x, z) = p(z, x) = 4.$$

Also, define $f : X \to X$ by $f(x) = f(y) = y$ and $f(z) = x$. Then for $\alpha = \dfrac{1}{2}$ and for each $k \in \left(\frac{1}{\sqrt{2}}, 1\right)$ the contractions (6)–(9) are true. while these contractions are not true for d, for each $k \in (0, 1)$ and each $\alpha \in (0, 1)$. For example,

$$d(fx, fz) = d(y, x) = 3 > 3^{\alpha} > k3^{\alpha} = kd(y, x)^{\alpha}d(x, z)^{1-\alpha} = kd(fx, x)^{\alpha}d(fz, z)^{1-\alpha}.$$

Obviousely y is the fixed point of f. Now define $q(x, y) = \max\{p(x, y), p(y, x)\}$ which is a symmetric w-distance. Then

$$q(x, x) = q(y, y) = 1, \quad q(y, y) = 0,$$

$$q(x, y) = 2, \quad q(x, z) = 4, \quad q(y, z) = 3.$$

and for $\alpha = \dfrac{1}{2}$ and for each $k \in \left(\frac{1}{\sqrt{2}}, 1\right)$, the contractions (6) and (7) are true for q.

In the latter theorem we conclude that for proving the existence of fixed point of a self map it sufficies to consider the symmetric w-distances. We apply the following remark for proving this main theorem.

Remark 1. *Note that if p is a symmetric w-distance, then for each $x \neq y$ we have $p(x, y) > 0$. Since if $p(x, y) = 0$, then*

$$p(x, x) \leqslant p(x, y) + p(y, x) = 2p(x, y) = 0.$$

Therefore $p(x, x) = 0 = p(x, y)$. So Lemma 1.1 implies $x = y$, a contradiction.

Therefore we have the following theorem.

Theorem 3.3. *Let p be a w-distance on a complete metric space (X, d). Suppose that $f : X \to X$ is a weakly interpolative Kannan type contraction self-map. Then f has a fixed point x in X such that $p(x, x) = 0$. In addition, if one of the equations (6)–(9) holds for all $x, y \in X$, then x is unique.*

Proof. Let $x_0 \in X$. Define by induction $x_n = f(x_{n-1})$. If for some n, $x_{n+1} = x_n$, then x_n is a fixed point of f. Otherwise if $x_{n+1} \neq x_n$, for each n, without loss of generality, we may assume p is symmetric. Since otherwise we can define $q(x, y) = \max\{p(x, y), p(y, x)\}$ which is a symmetric w-distance. Then if each of the equations (6)–(9) holds, then

$$p(fx, fy) \leqslant k\,[q(fx, x)]^{\alpha}\,[q(fy, y)]^{1-\alpha}\,,$$

and similarly

$$p(fy, fx) \leqslant k\,[q(fy, y)]^{\alpha}\,[q(fx, x)]^{1-\alpha}\,.$$

Therefore for each x, y with $fx \neq x$ and $fy \neq y$ we have

$$q(fx, fy) \leqslant k\,[q(fx, x)]^{\alpha}\,[q(fy, y)]^{1-\alpha}\,,$$

or

$$q(fx, fy) \leqslant k\,[q(fy, y)]^{\alpha}\,[q(fx, x)]^{1-\alpha}\,,$$

Note that since q is symmetric and $x_n \neq x_{n+1}$, we have $q(x_{n+1}, x_n) \neq 0$. So for each n, we have

$$q(x_{n+1}, x_n) \leqslant k\,[q(x_{n+1}, x_n)]^{\alpha}\,[q(x_n, x_{n-1})]^{1-\alpha}\,;$$

or

$$q(x_{n+1}, x_n) \leqslant k\,[q(x_{n+1}, x_n)]^{\beta}\,[q(x_n, x_{n-1})]^{1-\beta}\,;$$

where $0 < \beta = 1 - \alpha < 1$ and so,

$$[q(x_{n+1}, x_n)]^{1-\alpha} \leqslant k\,[q(x_n, x_{n-1})]^{1-\alpha}\,;$$

or

$$[q(x_{n+1}, x_n)]^{1-\beta} \leqslant k\,[q(x_n, x_{n-1})]^{1-\beta}\,.$$

Now we will have

$$[q(x_{n+1}, x_n)]^{1-\alpha} \leqslant k\,[q(x_n, x_{n-1})]^{1-\alpha} \leqslant k^2\,[q(x_{n-1}, x_{n-2})]^{1-\alpha}\,;$$

or

$$\begin{aligned} [q(x_{n+1}, x_n)]^{1-\alpha} &\leqslant k\,[q(x_n, x_{n-1})]^{1-\alpha} = \\ &= k\left([q(x_n, x_{n-1})]^{1-\beta}\right)^{\frac{1-\alpha}{1-\beta}} \leqslant \\ &\leqslant k^2\left([q(x_{n-1}, x_{n-2})]^{1-\beta}\right)^{\frac{1-\alpha}{1-\beta}} = \\ &= k^2\,[q(x_{n-1}, x_{n-2})]^{1-\alpha}\,. \end{aligned}$$

Therefore by an inductive method we conclude that $[q(x_{n+1}, x_n)]^{1-\alpha} \leqslant \ldots \leqslant k^n\,[q(x_1, x_0)]^{1-\alpha}$. This implies that $\lim_n q(x_{n+1}, x_n) = 0$ (since $0 < k < 1$). (Note that if p is symmetric, then we can replace q with p and also, all of the statements after "or" can be omitted and the proof is shorter.)
In the sequel, since applying "or" and working with β is very similar to working with α, we omitted them and we assume p is symmetric.
Now for each m, n we have

$$p(x_n, x_m) \leqslant k\,[p(x_n, x_{n-1})]^{\alpha}\,[p(x_m, x_{m-1})]^{1-\alpha} \to 0.$$

Therefore $\lim_{n,m} p(x_n, x_m) = 0$ and so by Lemma 1.1, $\{x_n\}$ is a Cauchy sequence. Now since (X, d) is complete, $\{x_n\}$ is convergent. Hence, there is $x \in X$ such that $\lim_n x_n = x$. In the

sequel we show that x is the fixed point of f.
By contrary if x is not the fixed point of f, then by Remark 1 $p(fx,x) \neq 0$ and

$$p(fx,x) \leqslant \liminf_{n} p(fx,x_{n+1}) = \liminf_{n} p(fx,fx_n) \leqslant k\,[p(fx,x)]^{\alpha}\,[p(x_{n+1},x_n)]^{1-\alpha}.$$

and so,

$$[p(fx,x)]^{1-\alpha} \leqslant k\,[p(x_{n+1},x_n)]^{1-\alpha} \to 0.$$

Therefore for each $\epsilon > 0$, $p(fx,x) < \epsilon$, a contradiction. Therefore $fx = x$ and $p(x,x) = = p(fx,x) = 0$. For uniqueness, let x, y are fixed points of f. Then

$$p(x,y) = p(fx,fy) \leqslant k\,[p(fx,x)]^{\alpha}\,[p(fy,y)]^{1-\alpha} = 0.$$

That is, $p(x,y) = 0$ and so by Remark 1 $x = y$.

□

4. $(\mathcal{F}, p)$-contraction

Suppose that $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}$ is a mapping satisfying the following properties.
($\mathcal{F}_1$) The mapping $\mathcal{F}$ is strictly increasing;
($\mathcal{F}_2$) For every sequence $\{t_n\} \subseteq \mathbb{R}^+$, $\lim_{n\to+\infty} t_n = 0$ if and only if $\lim_{n\to+\infty} \mathcal{F}(t_n) = -\infty$.
($\mathcal{F}_3$) $\lim_{t\to 0^+} t^k \mathcal{F}(t) = 0$ for some $k \in (0,1)$.

Definition 4.1. *Let (X,d) be a metric space with a w-distance p. A mapping $f : X \to X$ is said to be a weakly $\mathcal{F}$-contraction or $(\mathcal{F}, p)$-contraction, if there exist a constant $\alpha > 0$ such that*

$$p(fx,fy) > 0 \text{ implies } \alpha + \mathcal{F}\big(p(fx,fy)\big) \leqslant \mathcal{F}\big(p(x,y)\big), \tag{1}$$

for all $x, y \in X$.
If $p = d$ then f is called $\mathcal{F}$-contraction; [8].

The following example shows that the category of $(\mathcal{F}, p)$-contractions are bigger than its for $\mathcal{F}$-contractions:

Example 4.1. *Consider $X = \{x,y,z\}$ with the metric d which is defined by*

$$d(x,x) = d(y,y) = d(z,z) = 0;$$

$$d(x,y) = d(y,x) = d(x,z) = d(z,x) = d(y,z) = d(z,y) = 2.$$

Define $f : x \to X$ by $fx = fy = x$ and $fz = y$ and assume $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}$ satisfies in at least ($\mathcal{F}_1$), Then f is not an $\mathcal{F}$-contraction. Indeed for each $\alpha > 0$ we have

$$\alpha + \mathcal{F}(d(fx,fz)) = \alpha + \mathcal{F}(2) > \mathcal{F}(2) = \mathcal{F}(d(x,y)).$$

Now define the w-distance p with

$$p(x,x) = p(y,y) = 0; \quad p(z,z) = 1;$$

$$p(x,y) = 1; \quad p(y,x) = \frac{1}{2}; \quad p(y,z) = p(z,y) = p(x,z) = p(z,x) = 2.$$

then for $\mathcal{F}(r) = Ln(r)$ and for $\alpha = Ln2$ and each of the positive cases $p(fx,fz)$, $p(fy,fz)$, $p(fz,fx)$, $p(fz,fy)$ the contraction (1) hold. Therefore f is $(\mathcal{F}, p)$-contraction.

Wardowski in [8] proved the following theorem.

Theorem 4.1. *Each $\digamma$-contraction f on a complete metric space (X,d) has a unique fixed point. Moreover, for each $x_0 \in X$, the corresponding Picard sequence $\{f^n x_0\}$ converges to that fixed point.*

Then in [1] it is shown that the proof of the above theorem needs only the condition $(\digamma_1)$. Indeed $(\digamma_1)$ implies that $\digamma$ is almost every where continous and moreover the left and right limits exist in each $a \in (0,+\infty)$ and $\lim_{r\to a^+} \digamma(r) = \digamma(a^+)$. Then two Lemmas similar as Lemmas 1.2 and 1.3 for the metric d applied for the proof.

In the sequel we prove this theorem for $(\digamma, p)$-contraction.

Theorem 4.2. *Let $f : X \to X$ be a $(\digamma, p)$-contraction mapping on a complete metric space (X,d) with a w-distance p. If f is continuous, or for every $w \in X$ with $w \neq fw$, we have $\inf\{p(x,w) + p(x,fx) : x \in X\} > 0$, then f has a unique fixed point $u \in X$; and every sequence $\{f^n x_0\}_{n\in\mathbb{N}}$ is convergent to u, for every $x_0 \in X$.*

Proof. As we see in Theorem 4.2, we may consider p symmetric. Then the proof is similar to [1, Theorem 2.3] in the case where f is continuos. If f is not continous then similar to [1, Theorem 2.3] we can show that $f^n x_0 \to x$ and if x is not the fixed point of f, then

$$0 < \inf\{p(y,x) + p(y,fy) : y \in X\} \leqslant \inf\{p(x_n,x) + p(x_n,x_{n+1}) : n \in \mathbb{N}\} = 0.$$

Which is a contradiction. So x must be a fixed point. For uniqueness let x, y be the distinct fixed points of f, then since we consider p as a symmetric w-distance and $x \neq y$, we have $p(fx,fy) = p(x,y) > 0$ and so

$$\alpha + \digamma\big(p(x,y)\big) = \alpha + \digamma\big(p(fx,fy)\big) \leqslant \digamma\big(p(x,y).$$

Which means that $\alpha \leqslant 0$, a contradiction. So x is the unique fixed point of f. □

References

[1] N.Fabiano, Z.Kadelburg, N.Mirkov, V. Šešum Čavić, S.Radenović, On F-contractions: A Survey, *Contamporary Mathematics*, **3**(2022), no. 3, 327. DOI: 10.37256/cm.3320221517

[2] O.Kada, T.Suzuki, W.Takahashi, Nonconvex minimization theorems and fixed point theorems in complete metric spaces, *Math. Japonica*, **44**(1996), no. 2, 381–391

[3] Z.Kadelburg, S.Radenović, Some new observations on w-distance and F-contractions, submitted to Matematički Vesnik.

[4] R.Kannan, Some results on fixed points II., *Am. Math. Mon.*, **76**(1969), 405–408.

[5] E.Karapinar, Revisiting the Kannan type contractions via interpolation, *Advances in the Theory of Nonlinear Analysis and its Applications*, **2**(2018), no. 2, 85–87. DOI: 10.31197/atnaa.431135

[6] S.Radenović, Z.Kadelburg, D.Jandrlić, A.Jandrlić, Some results on weakly contractive maps, *Bulletin of the Iranian Mathematical Society*, **38**(2012), no. 3, 625–645.

[7] T.Suzuki, Several fixed point theorems in complete metric spaces, *Yokohama Mathematical Journa*, **44**(1997), 61–72

[8] D.Wardowski, Fixed points of a new type of contractive mappings in complete metric spaces, *Fixed Point Theory Appl.*, (2012), 2012:94. DOI: 10.1186/1687-1812-2012-94

О слабых сокращениях через w-расстояния

Хоссейн Лакзиан
Факультет математики
Университет Пайаме Нур
Тегеран, Иран
Седигех Барооткуб
Факультет фундаментальных наук
Университет Боджнорд
Боджнорд, Иран
Никола Фабиано
Институт ядерных наук «Винка» – Национальный институт Республики Сербия
Белградский университет
Белград, Сербия
Стоян Раденович
Факультет машиностроения
Белградский университет
Белград, Сербия

Аннотация. В этой статье мы проверим, остаются ли известные результаты верными, если метрику d заменить на w-расстояние p. Мы показываем, что в некоторых условиях сжатия, где вместо метрики d участвует w-расстояние p, можно предположить симметрию w-расстояния p, и доказательства могут быть короче. Мы говорим о таких результатах, как принцип сжатия Банаха, теорема Каннана, теоремы Бойда–Вонга, Мейра–Килера, Чаттерье, Райха, Харди–Роджерса, Карапинарса и Вардовский и многих других.

Сделав это, мы получим обобщения приведенных выше результатов.

Ключевые слова: фиксированная точка, w-расстояние, p-интерполяционное сокращение типа Каннана, p-сокращение Харди–Роджерса, $(\mathcal{F}, p)$-сокращение.

Journal of Siberian Federal University. Mathematics & Physics 2025, 18(1), 119–129

EDN: WBJQOM
УДК 539.3

Wave Propagation in a Blocky-layered Medium with Thin Interlayers

Evgenii A. Efimov*
Vladimir M. Sadovskii†
Institute of Computational Modelling SB RAS
Krasnoyarsk, Russian Federation

Received 20.08.2024, received in revised form 29.09.2024, accepted 01.11.2024

Abstract. In this paper, a mathematical model of a blocky-layered medium is studied. Deformable elastic blocks and thin elastic and viscoelastic interlayers are considered. Viscoelasticity is taken into account to describe wave attenuation. The wave fields in a medium described by the proposed simplified interlayer model are compared to wave fields which were obtained using the equations of the dynamic elasticity theory for interlayers. The developed computational technology is verified for compatibility with the experimental data.

Keywords: blocky-layered media, thin interlayer, viscoelastic interlayer.

Citation: E.A. Efimov, V.M. Sadovskii, Wave Propagation in a Blocky-layered Medium with Thin Interlayers, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 119–129. EDN: WBJQOM.

Introduction

The concept of a blocky structure of rock masses was proposed by M. A. Sadovskii [1, 2]. According to this concept, the geological medium can be represented as a hierarchical structure consisting of blocks of different scales nested inside each other. The characteristic sizes of blocks may vary from several meters to tens of kilometers. In a medium where the interlayers are more compliant than the blocks, pendulum waves can be observed. Pendulum waves in blocky media is well studied in both theoretical and experimental aspects. When the deformations arise mainly in the interlayers, due to their high compliance, the blocks can be considered as rigid bodies. Discrete periodic models with rigid blocks connected to each other by elastic springs were represented in [3–5]. A similar but more complicated mathematical model that takes into account the elasticity of blocks was considered in [6]. The equations of this model are written relative to the central points of the blocks, and the accelerations of these points depend on the elastic moduli of both the blocks and the interlayers. Wave attenuation in blocky media may occur due to the viscoelasticity of the interlayer material. The behavior of a discrete-periodic medium with elastic blocks and viscoelastic interlayers is quite consistent with the experimental data [6].

A more complicated approach involves dynamic elasticity equations to describe deformations of blocks. Blocky-layered media with sufficiently large number of blocks can be represented as

*efimov@icm.krasn.ru https://orcid.org/0000-0003-1830-6721
†sadov@icm.krasn.ru https://orcid.org/0000-0001-9695-0032

Cosserat continuum. The analysis of wave fields propagating in blocky-layered media and the Cosserat continuum was carried out in [7, 8].

In this paper, we study two-dimensional blocky-layered media with elastic blocks and thin elastic and viscoelastic interlayers. A system of ordinary differential equations is used to describe dynamics of interlayers, while the equations of the dynamic elasticity theory in partial derivatives are used for blocks. A more consistent approach supposes to apply equations of elasticity theory for both blocks and interlayers. However, this method computationally seems to be more difficult, in particular due to different restrictions on the time step in blocks and interlayers. The proposed simplified model of a blocky-layered medium retains thermodynamical compatibility inherent in equations of elasticity theory.

We compare the numerical solutions obtained by the interlayer model described by the equations of elasticity theory and the proposed simplified model. It turns out that in a medium with interlayers and blocks of the same material, non-physical reflections of waves occur near the boundaries of the blocks, which indicates defectiveness of the simplified model. The numerical results for a blocky medium with thin viscoelastic interlayers are in agreement with the experimental data published in the work [6].

1. Mathematical model of a blocky-layered medium

A two-dimensional problem of the dynamics of a blocky-layered medium consisting of rectangular blocks is considered. Motion of each block complies with the system of equations of a homogeneous isotropic elastic medium:

$$\begin{gathered}
\rho\,\frac{\partial v_1}{\partial t} = \frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{12}}{\partial x_2}, \qquad \rho\,\frac{\partial v_2}{\partial t} = \frac{\partial \sigma_{12}}{\partial x_1} + \frac{\partial \sigma_{22}}{\partial x_2},\\
\frac{\partial \sigma_{11}}{\partial t} = (\lambda + 2\mu)\frac{\partial v_1}{\partial x_1} + \lambda\frac{\partial v_2}{\partial x_2}, \qquad \frac{\partial \sigma_{22}}{\partial t} = \lambda\frac{\partial v_1}{\partial x_1} + (\lambda + 2\mu)\frac{\partial v_2}{\partial x_2},\\
\frac{\partial \sigma_{12}}{\partial t} = \mu\left(\frac{\partial v_2}{\partial x_1} + \frac{\partial v_1}{\partial x_2}\right).
\end{gathered} \tag{1}$$

The equations of longitudinal and transverse motions in the elastic interlayer between adjacent blocks in the x_1 direction are written as follows:

$$\begin{gathered}
\rho'\frac{d}{dt}\frac{v_1^+ + v_1^-}{2} = \frac{\sigma_{11}^+ - \sigma_{11}^-}{\delta_1}, \qquad \rho'\frac{d}{dt}\frac{v_2^+ + v_2^-}{2} = \frac{\sigma_{12}^+ - \sigma_{12}^-}{\delta_1},\\
\frac{d}{dt}\frac{\sigma_{11}^+ + \sigma_{11}^-}{2} = (\lambda' + 2\mu')\frac{v_1^+ - v_1^-}{\delta_1}, \qquad \frac{d}{dt}\frac{\sigma_{12}^+ + \sigma_{12}^-}{2} = \mu'\frac{v_2^+ - v_2^-}{\delta_1},
\end{gathered} \tag{2}$$

similarly, along the x_2 axis:

$$\begin{gathered}
\rho'\frac{d}{dt}\frac{v_2^+ + v_2^-}{2} = \frac{\sigma_{22}^+ - \sigma_{22}^-}{\delta_2}, \qquad \rho'\frac{d}{dt}\frac{v_1^+ + v_1^-}{2} = \frac{\sigma_{12}^+ - \sigma_{12}^-}{\delta_2},\\
\frac{d}{dt}\frac{\sigma_{22}^+ + \sigma_{22}^-}{2} = (\lambda' + 2\mu')\frac{v_2^+ - v_2^-}{\delta_2}, \qquad \frac{d}{dt}\frac{\sigma_{12}^+ + \sigma_{12}^-}{2} = \mu'\frac{v_1^+ - v_1^-}{\delta_2}.
\end{gathered} \tag{3}$$

Here v_1, v_2 are components of the displacement velocity vector, σ_{11}, σ_{22}, σ_{12} are components of the stress tensor, $\lambda = \rho(c_p^2 - 2c_s^2)$, $\mu = \rho c_s^2$ are Lame parameters, ρ is density, c_p, c_s are the velocities of longitudinal and transverse elastic waves, respectively, strokes indicate constants for interlayers. The interlayer thickness in both directions assumed to be the same $\delta = \delta_1 = \delta_2$.

Signs «+» and «–» are represent the values on the right and left boundaries of the interlayer, respectively.

The system (1)–(3) is thermodynamically compatible. The law of conservation of energy can be written down as the sum of the kinetic and potential energies of all blocks and interlayers, equal to the integral of flux of the Umov–Poynting vector by time and across the boundary of the block array consisting of $n_1 \times n_2$ blocks [7]:

$$
\begin{aligned}
&\sum_{k_1=1}^{n_1}\sum_{k_2=1}^{n_2}\int_0^{h_1}\int_0^{h_2}\Big(\frac{\rho}{2}\vec{v}^{k_1,k_2}(t,x_1,x_2)^2 + W^{k_1,k_2}(t,x_1,x_2)\Big)dx_1dx_2 + \\
&+\delta_1\sum_{k_1=1}^{n_1-1}\sum_{k_2=1}^{n_2}\int_0^{h_2}\Big(\frac{\rho'}{2}\Big[\frac{\vec{v}^{k_1+1,k_2}(t,0,x_2)+\vec{v}^{k_1,k_2}(t,h_1,x_2)}{2}\Big]^2 + \\
&+\frac{1}{2\rho' {c'}_p^2}\Big[\frac{\sigma_{11}^{k_1+1,k_2}(t,0,x_2)+\sigma_{11}^{k_1,k_2}(t,h_1,x_2)}{2}\Big]^2 + \\
&+\frac{1}{2\rho' {c'}_s^2}\Big[\frac{\sigma_{12}^{k_1+1,k_2}(t,0,x_2)+\sigma_{12}^{k_1,k_2}(t,h_1,x_2)}{2}\Big]^2\Big)\,dx_2 + \\
&+\delta_2\sum_{k_1=1}^{n_1}\sum_{k_2=1}^{n_2-1}\int_0^{h_1}\Big(\frac{\rho'}{2}\Big[\frac{\vec{v}^{k_1,k_2+1}(t,x_1,0)+\vec{v}^{k_1,k_2}(t,x_1,h_2)}{2}\Big]^2 + \\
&+\frac{1}{2\rho' {c'}_p^2}\Big[\frac{\sigma_{22}^{k_1,k_2+1}(t,x_1,0)+\sigma_{22}^{k_1,k_2}(t,x_1,h_2)}{2}\Big]^2 + \\
&+\frac{1}{2\rho' {c'}_s^2}\Big[\frac{\sigma_{12}^{k_1,k_2+1}(t,x_1,0)+\sigma_{12}^{k_1,k_2}(t,x_1,h_2)}{2}\Big]^2\Big)dx_1 = \\
&=\sum_{k_2=1}^{n_2}\int_0^{t}\int_0^{h_2}\Big(p_1^{n_1,\,k_2}(t,h_1,x_2) - p_1^{0,\,k_2}(t,0,x_2)\Big)dx_2dt + \\
&+\sum_{k_1=1}^{n_1}\int_0^{t}\int_0^{h_1}\Big(p_2^{k_1,\,n_2}(t,x_1,h_2) - p_2^{k_1,\,0}(t,x_1,0)\Big)dx_1dt.
\end{aligned} \tag{4}
$$

Here, $\vec{v} = (v_1, v_2)$ is the velocity vector, $p_1 = \sigma_{11}v_1 + \sigma_{12}v_2$, $p_2 = \sigma_{22}v_2 + \sigma_{12}v_1$ are the projections of the power flux vector, W is the elastic potential:

$$W = \frac{(\sigma_{11}+\sigma_{22})^2}{8(\lambda+\mu)} + \frac{(\sigma_{11}-\sigma_{22})^2 + 4\sigma_{12}}{8\mu}. \tag{5}$$

Thermodynamic compatibility guarantees the well-posedness of the initial-boundary value problem with the dissipative boundary conditions, under which the right-hand side of (4) is non-negative.

We consider a boundary value problem for a blocky-layered massif with fixed boundaries ($v_1 = v_2 = 0$ along the boundaries). The numerical solution is calculated in the region $\Omega = [0, L_1] \times [0, L_1]$ with a uniform grid of $N_1 \times N_2$ nodes. At the boundary $x_1 = 0$, at point $x_2 = x_{imp}$ the pressure pulse is $\sigma_{11}(t, x_{imp}) = p(t)$.

The numerical algorithm for solving equations in blocks is based on the two-cyclic splitting method with the respect to spatial coordinates. This method allows to achieve the second order of convergence when splitted one-dimensional problems are solved by finite-difference schemes of at least second order [9]. Godunov scheme with limiting reconstruction of Riemann invariants is used to solve one dimensional problems [10]. The reconstruction procedure provides second-order approximation in monotonic sections of the solution.

Equations in the interlayers are solved using Ivanov dissipation-free scheme [11]. To eliminate artificial scheme dissipation, it is necessary to require that the sum of the values at the upper (indicated by the "hat" symbol) and lower time steps be equal to the sum of the values at the left and right boundaries of the grid cell:

$$v^+ + v^- = \hat{v} + v, \qquad \sigma^+ + \sigma^- = \hat{\sigma} + \sigma.$$

Based on this requirement, at the "predictor" stage of the scheme we obtain the system [13]:

$$\begin{gathered} I^+ = \rho c v^+ + \sigma^+\,, \quad I^- = \rho c v^- - \sigma^-\,, \\ v^+ - v^- = \frac{\delta}{\rho c \delta + \rho' c'^2 \tau}(I^+ - I^- - 2\sigma), \qquad \sigma^+ - \sigma^- = \frac{\rho' \delta}{\rho' \delta + \rho c \tau}(I^+ + I^- - 2\rho c v), \\ v^+ + v^- = \frac{I^+ + I^- - (\sigma^+ - \sigma^-)}{\rho c}, \qquad \sigma^+ + \sigma^- = I^+ - I^- - \rho c (v^+ - v^-). \end{gathered} \tag{6}$$

Here I^+ and I^- are Riemann invariants, calculated on the boundaries of neighboring blocks separated by an interlayer, τ is the time step and h is the space step. On the "corrector" stage we have a system:

$$\hat{v} = v + \frac{\tau}{\rho \delta}(\sigma^+ - \sigma^-), \qquad \hat{\sigma} = \sigma + \frac{\tau \rho c^2}{\delta}(v^+ - v^-). \tag{7}$$

This scheme can be written for independent subsystems for longitudinal and transverse waves propagating with velocities $c = c_p$ and $c = c_s$, respectively. It is necessary to allocate one-dimensional arrays in each direction for interlayers. That is, when solving splitted one-dimensional problems, stresses and strain rates in each interlayer are calculated in only one grid cell.

To account mechanical energy dissipation, we consider viscoelastic interlayers. Viscoelastic interlayers are described by the Poynting–Thomson model, also known as the standard linear solid (SLS) model. The rheological scheme of the model consists of elastic element b_0 connected in series with a parallel connection of a viscous element η and an elastic one b (Fig. 1). An array of SLS-mechanisms connected in parallel is called a generalized standard linear solid (GSLS). This rheological model is widely used in geophysics due to its ability to describe media with a nearly constant quality factor over a certain frequency range. The more mechanisms in the model, the more precisely constant quality factor can be approximated. Denoting s as the stress on elastic element b we can write down the system of equations for the viscoelastic interlayer in the following form:

$$\begin{gathered} \rho' \frac{d}{dt}\frac{v^+ + v^-}{2} = \frac{\sigma^+ - \sigma^-}{\delta_1}, \qquad \frac{1}{b_0}\frac{d}{dt}\frac{\sigma^+ + \sigma^-}{2} = \frac{v^+ - v^-}{\delta_1} - \frac{1}{\eta}\left(\frac{\sigma^+ + \sigma^-}{2} - \frac{s^+ + s^-}{2}\right), \\ \frac{1}{b}\frac{d}{dt}\frac{s^+ + s^-}{2} = \frac{1}{\eta}\left(\frac{\sigma^+ + \sigma^-}{2} - \frac{s^+ + s^-}{2}\right). \end{gathered} \tag{8}$$

For longitudinal waves, the coefficient b is equal to $\lambda + 2\mu$ and for transverse waves is equal to μ, and similarly, b_0 is equal to $\lambda_0 + 2\mu_0$ or μ_0.

This model can be rewritten in terms of the relaxation modulus and relaxation times of stress and strain. Relaxation times can be determined from the known quality factor using the τ-method [12]. Then we can recalculate the elastic moduli and viscosity coefficient.

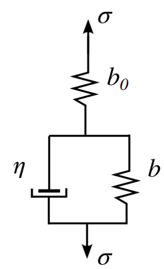


Fig. 1. Rheological scheme of the Poynting–Thomson model

Finite-difference scheme is constructed analogously to (6)–(7), but leads to a more cumbersome form:

$$v^+ - v^- = \frac{1}{\alpha}\big(\beta(I^+ - I^-) - 2\sigma - \frac{2b_0\tau}{2\eta + b\tau}s\big), \qquad \sigma^+ - \sigma^- = \frac{\rho'\delta}{\rho'\delta + \rho c\tau}\big(I^+ + I^- - 2\rho c v\big),$$

$$v^+ + v^- = \frac{I^+ + I^- - (\sigma^+ - \sigma^-)}{\rho c}, \quad \sigma^+ + \sigma^- = I^+ - I^- - \rho c(v^+ - v^-)\,,$$

$$s^+ + s^- = \frac{b\tau}{2\eta + b\tau}\big(I^+ - I^- - \rho c(v^+ - v^-)\big) + \frac{4\eta}{2\eta + b\tau}s\,,$$

$$\hat{v} = v + \tau\frac{\sigma^+ - \sigma^-}{\delta\rho'}, \quad \hat{\sigma} = \sigma + b_0\tau\frac{v^+ - v^-}{\delta} - \frac{b_0\tau}{\eta}\left(\frac{\sigma^+ + \sigma^-}{2} - \frac{s^+ + s^-}{2}\right),$$

$$\hat{s} = s + \frac{b\tau}{\eta}\left(\frac{\sigma^+ + \sigma^-}{2} - \frac{s^+ + s^-}{2}\right),$$

where

$$\alpha = \rho c + \frac{b_0\tau}{\delta} + \frac{b_0\tau\rho c}{2\eta}\left(1 - \frac{b\tau}{2\eta + b\tau}\right), \qquad \beta = 1 + \frac{b_0\tau}{2\eta}\left(1 - \frac{b\tau}{2\eta + b\tau}\right).$$

2. Results of computations

The computations below were performed on a multiprocessor system with cluster architecture. The software package was developed using the MPI library. Each MPI-process performs computations on each block, which consists of smaller blocks. One can specify different interlayer thicknesses for larger and smaller blocks, so that it is possible to simulate wave propagation in hierarchical blocky media.

Simplification of the interlayer model leads to certain inaccuracies. Let us evaluate the behavior of the wave field when propagating near the boundaries of blocks. We consider a medium consisting of four identical rectangular blocks with sides of 12 and 24 m, separated by interlayers with varying thickness. The first case considered concerns a medium with the blocks and interlayers of the same material with properties $\rho = \rho' = 2400$ kg/m^3, $c_p = c'_p = 4500$ m/s, $c_s = c'_s = 2700$ m/s. The load pressure at the upper boundary of the first block at point $x_{imp}=21$ m is $p(t) = p_0 H(t)$, where $H(t)$ is Heaviside function. The Fig. 2 shows snapshots of velocity fields obtained using interlayer model (2)–(3) calculated on uniform grid of $N_1 \times N_2 = 480 \times 960$ nodes (with $h = 0.5$ m). In the medium, where interlayers are modeled by the same equations as blocks, waves propagate like in a homogenous medium. As the thickness of the interlayers increases, partial reflections of waves from the vertical layer become more and more

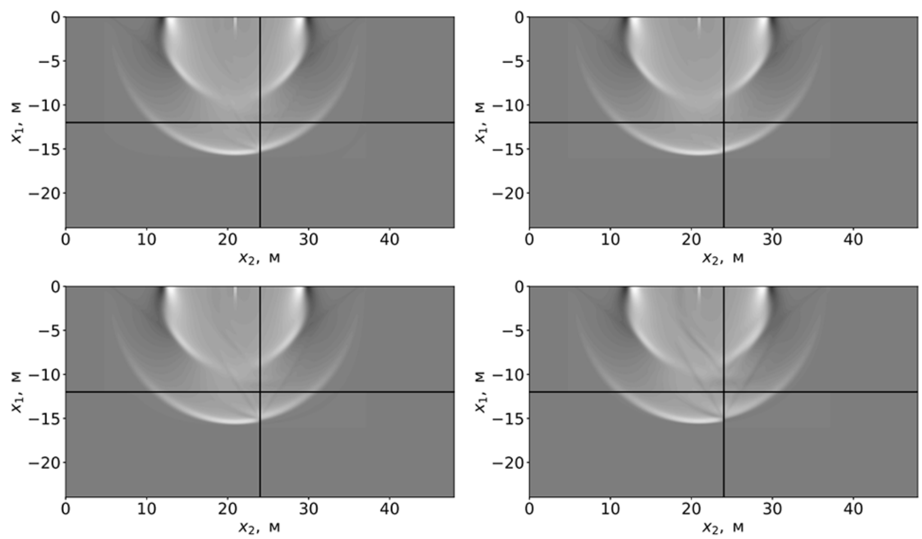


Fig. 2. Snapshots of the velocity v_1 in medium with blocks and interlayers made of the same material obtained using simplified intarlayer model (2)–(3), interlayer thicknesses are $\delta = 0.025$ m (upper left), 0.05 m (upper right), 0.1 m (bottom left), 0.2 m (bottom right)

sufficient. There are almost no reflections from the horizontal interlayer, since the wave passes through it almost perpendicularly.

Let us consider a medium of the same configuration but with a more compliant interlayer material: $\rho' = 2100$ kg/m^3, $c'_p = 2900$ m/s, $c'_s = 1700$ m/s. In this case no visual differences between the snapshots obtained by different interlayer models are observed (Fig. 3). To es-

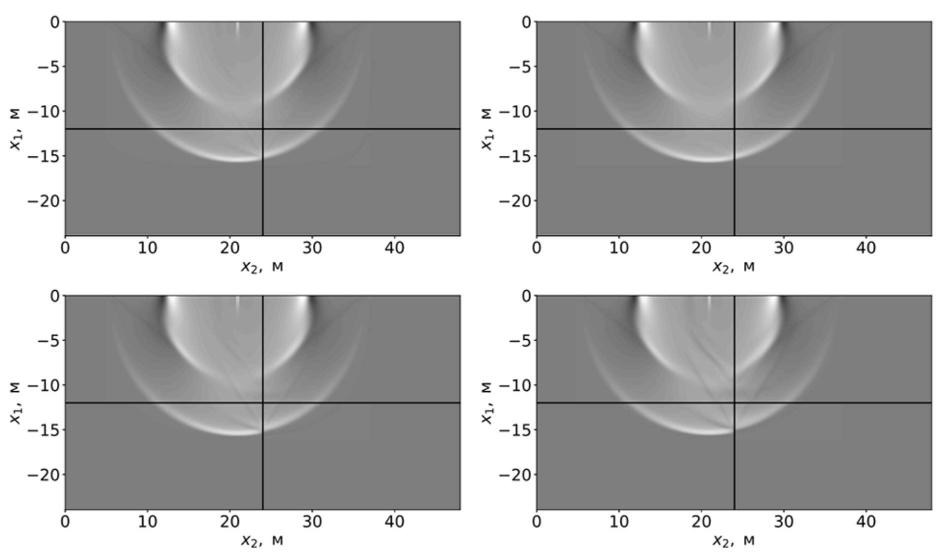


Fig. 3. Snapshots of the velocity v_1 in medium with compliant interlayers $\delta = 0.05$ m (left) and 0.2 m (right) thick for simplified interlayer model (2)–(3) (upper) and for interlayers described by elasticity theory equations (bottom)

timate the error of the numerical solution U obtained with the use of a simplified interlayer model (2)–(3), we compare it to a reference solution U_e calculated for interlayers described by elasticity theory equations. The relative error $err_2 = ||U - U_e||/||U_e||$ of the numerical solution $U = (v_1, v_2, \sigma_{11}, \sigma_{22}, \sigma_{12})$ was calculated using a discrete equivalent of the norm of the space $L_\infty(0,T;\, L_2(\Omega))$:

$$||U|| = \sup_{0<t<T} \sqrt{\iint\limits_{\Omega} \left(\rho \frac{v_1^2 + v_2^2}{2} + W\right) dx_1\, dx_2},$$

where T is the time required for the longitudinal wave to reach the boundary of the computational domain Ω, W is the elastic potential (5). Also we use the norm

$$||U|| = \sup_{0<t<T} \max_{\Omega} |U|$$

to calculate relative error err_C.

Tab. 1 shows the relative errors depending on grid step h for a fixed interlayer thickness. The material of the blocks for all cases has parameters $\rho = 2400$ kg/m^3, $c_p = 4500$ m/s, $c_s = 2700$ m/s, the material of the interlayers varies. Tab. 2 shows the relative errors depending on interlayer

Table 1. The relative error depending on grid step at fixed interlayer thickness $\delta = 0.1$ m

Interlayer material parameters		$\rho' = \rho$, $c'_p = c_p$, $c'_s = c_s$		$\rho' = 2100$ kg/m^3, $c'_p = 2900$ m/s, $c'_s = 1700$ m/s		$\rho' = 1100$ kg/m^3, $c'_p = 1500$ m/s, $c'_s = 800$ m/s	
h, м	δ/h	err_2	err_C	err_2	err_C	err_2	err_C
0.1	1	0.0352	0.272	0.0303	0.123	0.0231	0.0715
0.05	2	0.0483	0.321	0.0286	0.144	0.0258	0.0839
0.025	4	0.0621	0.348	0.0357	0.146	0.0309	0.115
0.0125	8	0.0802	0.350	0.0503	0.153	0.0475	0.131

thickness with a fixed grid. With an increase in the ratio of the interlayer thickness to the grid

Table 2. The relative error depending on interlayer thickness at fixed grid $N_1 \times N_2 = 960 \times 1920$ ($h = 0.025$ m)

Interlayer material parameters		$\rho' = \rho$, $c'_p = c_p$, $c'_s = c_s$		$\rho' = 2100$ kg/m^3, $c'_p = 2900$ m/s, $c'_s = 1700$ m/s		$\rho' = 1100$ kg/m^3, $c'_p = 1500$ m/s, $c'_s = 800$ m/s	
δ, м	δ/h	err_2	err_C	err_2	err_C	err_2	err_C
0.025	1	0.0195	0.154	0.0177	0.0756	0.0117	0.0597
0.05	2	0.0362	0.250	0.0211	0.108	0.0197	0.0704
0.1	4	0.0621	0.348	0.0357	0.146	0.0309	0.115
0.2	8	0.1035	0.414	0.0719	0.161	0.0689	0.237

step δ/h, an increase in error is observed in all cases. It is noticeable that in media with more compliant interlayers the error is slightly lower. Therefore, the model with simplified equations for interlayers can be used to describe blocky media with sufficiently thin and compliant interlayers.

Fig. 4 shows the distribution of the error $|v_1 - v_{1e}|/|v_{1e}|$ in blocky media for layers of different thicknesses on a uniform grid $N_1 \times N_2 = 960 \times 1920$ ($h = 0.025$ m).

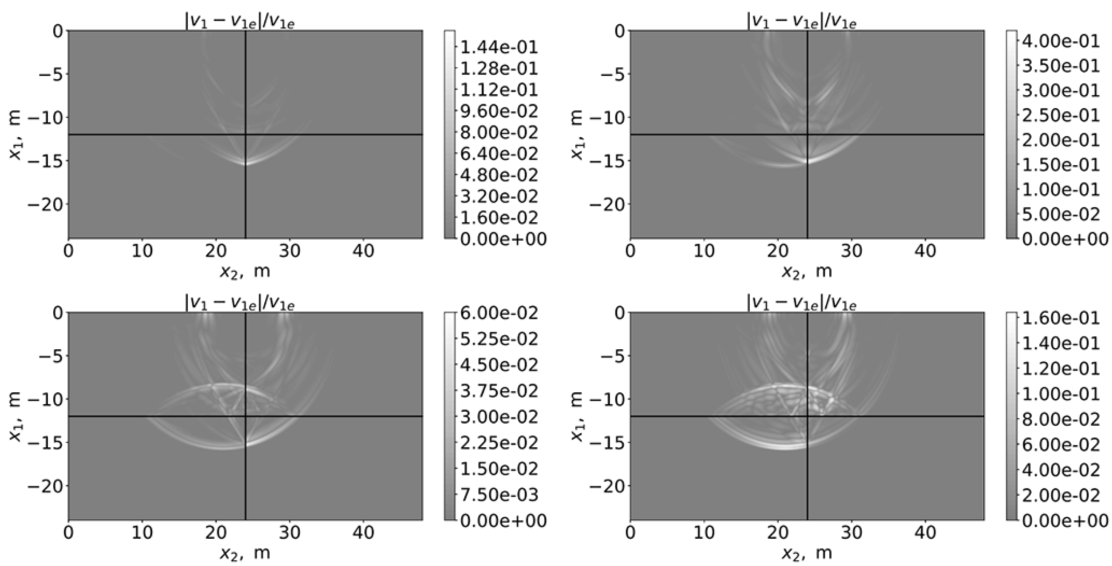


Fig. 4. The relative error in a blocky-layered medium with interlayers $\delta = 0.025$ m (left) and $\delta = 0.2$ m (right) thick, interlayer material with $\rho' = \rho$, $c'_p = c_p$, $c'_s = c_s$ (upper), a more compliant interlayer material $\rho' = 2100$ kg/m^3, $c'_p = 2900$ m/s, $c'_s = 1700$ m/s (bottom)

Verification of the mathematical model and computational technology was carried out according to experimental data published in paper [6]. In the experiments on a biaxial stand, a blocky-layered medium was simulated by an assembly of 36 blocks measuring $89 \times 125 \times 250$ mm, each made of plexiglass ($\rho = 2040$ kg/m^3, $c_p = 2670$ m/s). Blocks were separated by 5 mm thick rubber interlayers with shear moduli in directions x_1 and x_2 equal to $10^7/1.3$ Pa and $1.35 \cdot 10^7/1.3$ Pa, respectively.

It was assumed that the shear moduli of the interlayers correspond to the state of long-term deformation, when both elements of the rheological scheme are deformed (Fig. 1). The Poisson's ratio for all assembly materials was assumed to be 0.3. Fig. 5 shows the diagram of the numerical experiment. The rod striker generated elastic waves in contact with the surface of the block. At point x_{imp}, denoted by the red arrow in Fig. 6, the pulse impact $\sigma_{11}(t, x_{imp}) = p(t)$ with

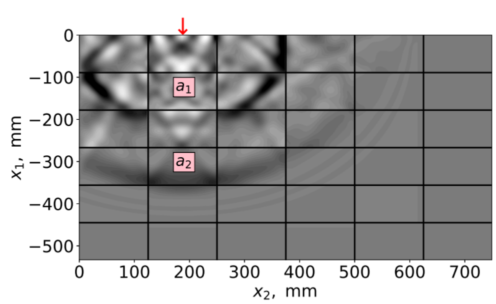


Fig. 5. The numerical experiment diagram. Accelerometers a_1 and a_2 are placed in the central points of the corresponding blocks

duration $T_{imp} = 0.2$ ms has the following form:

$$p(t) = \begin{cases} p_0 \sin(\pi t/T_{imp}), & 0 < t \leqslant T_{imp} \\ 0, & t > T_{imp}. \end{cases}$$

Accelerometers a_1 and a_2 were measuring accelerations $w_i = \partial v_i/\partial t$ for 5 ms in the central points of the corresponding blocks.

Figures 6–9 show the theoretical and experimental results from paper [6] in comparison with the numerical solution of (1) and (8). The experimental dependencies of acceleration on time are denoted by the blue dashed line, the red lines show accelerations calculated using the approach proposed in [6], the green curves correspond to the numerical solution for a medium with elastic blocks and viscoelastic interlayers. The parameters of the SLS were obtained using the τ-method [12] assuming that quality factor Q is nearly constant in the frequency range $[100, 5000]$ Hz. It was assumed that quality factors of the longitudinal and transverse waves are $Q_p = 20$ and $Q_s = 10$, respectively. The lack of data on the material of the interlayers leaves a certain amount of arbitrariness in the choice of parameters of the viscoelastic medium.

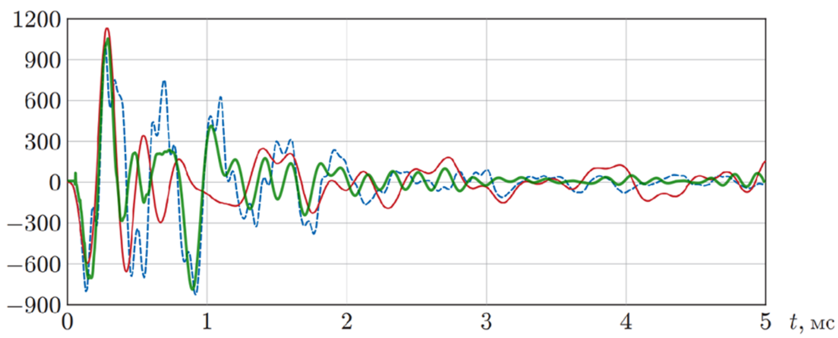


Fig. 6. Waveforms of acceleration w_1, measured in the centre of block a_1

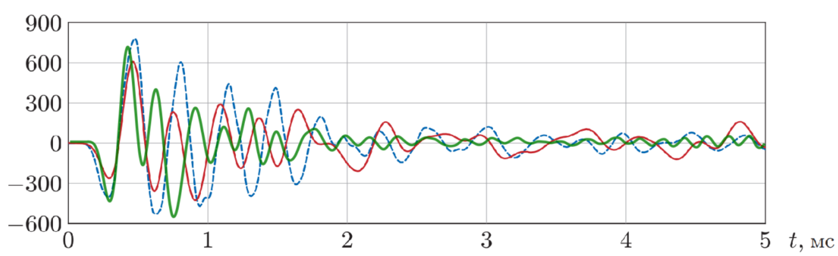


Fig. 7. Waveforms of acceleration w_1, measured in the centre of block a_2

The results of the numerical simulation are in good agreement with experimental data. The calculated acceleration waveform shown in Fig. 6 is almost identical to the experimental measurement. In Fig. 7 one can see the difference in phase, but the qualitative behaviour of the waves remains the same. A more observable difference can be noted in Fig. 8–9 where the experimental high-frequency oscillations with large amplitude were not detected numerically. Most likely, this is due to the fact that accelerations in the experiment were measured on the side surface of the block, while the two-dimensional problem supposes measurements "inside" the thickness of the

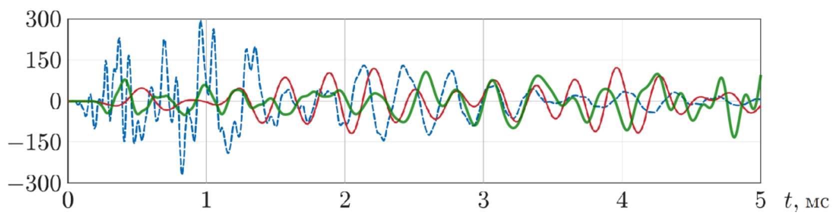


Fig. 8. Waveforms of acceleration w_2, measured in the centre of block a_1

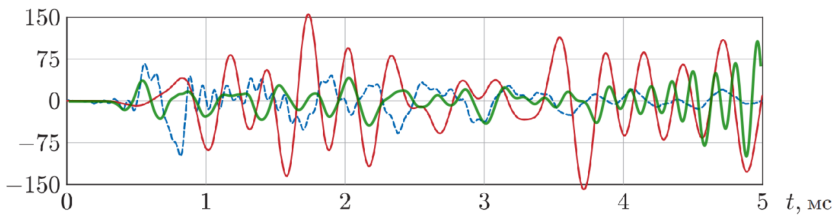


Fig. 9. Waveforms of acceleration w_2, measured in the centre of the block a_2

block. It would be more accurate to apply a three-dimensional model of a blocky-layered medium with the same location of accelerometers as in the real experiment.

Conclusions

The considered simplified interlayer model reliably describes wave processes in blocky-layered media. When blocks and interlayers are made of the same material, non-physical reflections occur and grow as the interlayers get thicker. The solutions for the simplified interlayers model and for the interlayers described by elasticity theory equations are compared. It is observed that the error of the numerical solution obtained by the simplified model increases with increasing ratio of the interlayer thickness to the grid step. The mathematical model was verified on the experimental data published in paper [6]. The presented computations show good agreement with the experimental measurements.

This work is supported by the Krasnoyarsk Mathematical Center and financed by the Ministry of Science and Higher Education of the Russian Federation in the framework of the establishment and development of regional Centers for Mathematics Research and Education (Agreement no. 075-02-2024-1378).

References

[1] M.A.Sadovskii, Natural lumpiness of a rock, *DAN USSR*, **247**(1979), no. 4, 829–831.

[2] M.A.Sadovskii, S.S.Sardarov, Coordination and similarity of the geomotions in connection with natural jointing of rocks, *DAN USSR*, **250**(1980), no. 4, 846–848.

[3] N.I.Aleksandrova, E.N.Sher, A.G.Chernikov, Effect of viscosity of partings in block-hierarchical media on propagation of low-frequency pendulum waves, *Journal of Mining Science*, **44**(2008), no. 3, 225–234. DOI: 10.1007/s10913-008-0012-3

[4] N.I.Aleksandrova, E.N.Sher, Wave propagation in the 2D periodical model of a block-structured medium. Part I: characteristics of waves under impulsive impact, *Journal of Mining Science*, **46**(2010), no. 6, 639–649. DOI: 10.1007/s10913-010-0081-y

[5] N.I.Aleksandrova, The discrete Lamb problem: Elastic lattice waves in a block medium, *Wave Motion*, **51**(2014), no. 5, 30–34. DOI: 10.1016/j.wavemoti.2014.02.002

[6] V.A.Saraikin, A.G.Chernikov, E.N.Sher, Wave propagation in two-dimensional block media with viscoelastic layers (theory and experiment), *Journal of Applied Mechanics and Technical Physics*, **56**(2015), no. 4, 688–697. DOI: 10.1134/S0021894415040161

[7] V.M.Sadovskii, O.V.Sadovskaya, Modeling of elastic waves in a blocky medium based on equations of the Cosserat continuum, *Wave Motion*, **52**(2015), 138–150. DOI: 10.1016/j.wavemoti.2014.09.008

[8] V.M.Sadovskii, O.V.Sadovskaya, M.A.Pokhabova, Modeling of elastic waves in a block medium based on equations of the Cosserat continuum, *Computational continuum mechanics*, **7**(2014), 52–60.

[9] G.I.Marchuk, Splitting methods, Moscow, Nauka, 1988 (in Russian).

[10] A.G.Kulikovskii, N.V.Pogorelov, A.Yu.Semenov, Mathematical aspects of numerical solution of hyperbolic systems, Moscow, Fizmatlit, 2001 (in Russian).

[11] G.V.Ivanov, Yu.M.Volchkov, I.O.Bogulskii, S.A.Anisimov, V.D.Kurguzov, Numerical solution of dynamic elastic-plastic problems of deformable solids, Novosibirsk, Sib. Univ. Izd., 2002 (in Russian).

[12] J.O.Blanch, J.O.Robertsson, W.W.Symes, Modeling of a constant Q: methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique, *Geophysics*, **60**(1995), 176–184.

[13] V.M.Sadovskii, O.V.Sadovskaya, Numerical algorithm based on implicit finite-difference schemes for analysis of dynamic processes in blocky media, *Russian Journal of Numerical Analysis and Mathematical Modelling*, **33**(2018), no. 2, 111–121. DOI: 10.1515/rnam-2018-0010

Распространение волн в блочно-слоистой среде с тонкими прослойками

Евгений А. Ефимов
Владимир М. Садовский
Институт вычислительного моделирования СО РАН
Красноярск, Российская Федерация

Аннотация. Исследуется математическая модель блочно-слоистой среды с тонкими прослойками. Рассматриваются деформируемые упругие блоки и упругие прослойки. Для описания затухания волн учитывается вязкоупругость в прослойках. Проводится численное сравнение упрощённой модели прослоек с прослойками, описываемыми полными уравнениями теории упругости. Результаты численного моделирования сравниваются с экспериментальными данными.

Ключевые слова: блочно-слоистая среда, тонкая прослойка, вязкоупругая прослойка.

EDN: YHUARK
УДК 57.087.31/.37

A Novel Protocol for Cell Nanomechanical Assay combined with Rapid Protein Profiling Via AFM-LSM Pattern Colocalization

Mikhail E. Shmelev*
School of Medicine and Life Sciences, Far Eastern Federal University
Vladivostok, Russian Federation
A.V. Zhirmunsky National Scientific Center of Marine Biology
Far Eastern Branch, Russian Academy of Sciences
Vladivostok, Russian Federation

Anna K. Kravchenko†
School of Medicine and Life Sciences, Far Eastern Federal University
Vladivostok, Russian Federation

Vadim V. Kumeiko‡
School of Medicine and Life Sciences, Far Eastern Federal University
Vladivostok, Russian Federation
A.V. Zhirmunsky National Scientific Center of Marine Biology
Far Eastern Branch, Russian Academy of Sciences
Vladivostok, Russian Federation

Received 10.08.2024, received in revised form 30.09.2024, accepted 23.11.2024

Abstract. The cell mechanical assay has emerged as a powerful approach to studying cellular behavior and protein dynamics. This work presents the novel protocol that combines cell nanomechanical assay with rapid protein expression profiling, enabling comprehensive insight into cellular responses. The new protocol leverages advanced techniques in atomic force microscopy (AFM) to measure the mechanical properties of individual cells, while simultaneously utilizing a laser scanning microscopy for the high-throughput quantification of protein expression levels. This dual-assay method allows researchers to elucidate the relationship between cellular mechanical properties and protein dynamics, uncovering critical insights into cellular physiology and pathophysiology. The effectiveness of the protocol was validated through experiments with cancer cells, showcasing its potential in colocalization of wnt3a ligand molecule and actin cytoskeleton with Young's modulus patterns of the cell. Our findings indicate that this integrated approach not only enhances the accuracy of cellular assessments but also accelerates the understanding of cellular mechanisms at the nanoscale. This protocol holds promise for applications in drug development, diagnostics, and personalized medicine, offering a new lens through which we can explore the intricate interplay between cellular mechanics and protein expression.

Keywords: atomic force microscopy, laser scanning microscopy, image colocalization, glioma, Wnt-signaling.

Citation: M.E. Shmelev, A.K. Kravchenko, V.V. Kumeiko, A Novel Protocol for Cell Nanomechanical Assay combined with Rapid Protein Profiling Via AFM-LSM Pattern Colocalization, J. Sib. Fed. Univ. Math. Phys., 2025, 18(1), 130–143. EDN: YHUARK.

*shmelev.m.e@gmail.com https://orcid.org/0000-0001-7106-1001.
†kravchenko.ak@dvfu.ru
‡vkumeiko@yandex.ru

Introduction

Atomic Force Microscopy (AFM) is a tool often used for studying biological structures and processes at cellular and molecular scales. It offers the ability to image live cells at a resolution much higher than light microscopy allows [1]. AFM is particularly significant in the analysis of cell and biomechanics, where it quantifies mechanical parameters related to cytoskeleton organization [2].

However, quantifying cell mechanics presents its own challenges. The main one is that the cell senses external mechanical stimuli and changes its metabolic profile, which in turn changes its mechanical properties [3]. This is a prerequisite for understanding the role of mechanosensing and the cellular response expressed as changes in cell stiffness and tension. At the same time, measuring the relationship between molecular kinetics and cell mechanics opens up new possibilities for understanding the adaptive mechanobiological mechanisms of cells that have yet to be fully understood [4]. Combining Laser Scanning Microscopy (LSM) with Atomic Force Microscopy (AFM) can significantly enrich data acquisition for cell and biomaterial research.

AFM is a technique that provides high-resolution images, which allows users to view surfaces under any aqueous conditions. It is used to study the structural and mechanical properties of a wide range of biological matters like biomolecules, cells, and tissues. However, AFM studies are typically limited to imaging the surface of the cell membrane [5].

On the other hand, LSM is an optical imaging technique that offers a variety of supplemental datasets. While AFM provides valuable information from the topography mapping of a sample surface, LSM allows the acquisition of molecular marker distributions [6]. Hence, when combined, they can provide comprehensive information on surface structure, mechanical topography and specific molecular distribution.

The combination also permits structural mapping of properties and enables manipulation with molecular precision [6]. This combined mapping of immunohistochemical and topographical properties aids in delivering more comprehensive and detailed data [7].

Previously we proved that mechanical properties of glioma cells depend on the expression of CD44 receptors on the cell surface [8].

We verified this protocol on WNT3a and actin colocalization with nanomechanical properties obtained by AFM.

Wnt ligands are secreted proteins which activate the wnt pathway by interaction with Frizzled protein (Fz) and cytoplasmic low-density lipoproteins LRP 5/6 [9]. Canonical wnt3a pathway involved in all cell and tissue processes including cell division, differentiation, malignization and epithelial-mesenchymal transition [10–12]. Several studies are devoted to anticancer therapy based on targeting on wnt3a protein or its receptor. The main targets are Disheveled (Dvl) protein [13], PORCN, or miRNA assays [14].

It was shown that wnt3a has a significant impact on glioblastoma cell migration and proliferation both in vitro and in vitro. Glioblastoma cells are synthesizing Wnt3a which increases local microglial ARG-1 and STI1 expression, followed by an upregulation of IL-10 mRNA levels, and a decrease in IL1β gene expression. The presence of Wnt3a in microglia-glioblastoma co-cultures increases the formation of cell membrane actin cytoskeleton accompanied by changes in migration capability. *In vivo*, tumors formed from Wnt3a-stimulated glioblastoma cells presented greater microglial infiltration and more aggressive characteristics such as growth rate than untreated tumors [15].

In this study, we propose a fast and effective assay for conducting live cell imaging using the combined strengths of Atomic Force Microscopy (AFM) and Laser Scanning Microscopy (LSM).

1. Materials and methods

1.1. Cell culture assay

Glioma early-passage cell cultures BT32, BT39, BT40, BT52 established from tumor samples described previously were used in this work [8]. The established cell cultures were cultivated until the 80% confluency; then, the cells were removed from the flask surfaces by using 0.25% trypsin solution (Thermo Fisher, Waltham, MA, USA) and reseeded on sterile coverslips laying in cell culture dishes. High-glucose DMEM with 1mM sodium pyruvate and 300 mg/L L-glutamine (Thermo Fisher, Waltham, MA, USA), 10% fetal bovine serum (FBS) (Thermo Fisher, Waltham, MA, USA), penicillin (50 U/ml) and streptomycin (50 ug/ml) (Thermo Fisher, Waltham, MA, USA) was used.

1.2. Atomic force microscopy

We used a Bruker BioScope Resolve microscope (Bruker, USA) for atomic force microscopy (AFM). The probes selected for the experiments were PFQNM-LC-A-Cal and SNL-C (Bruker, USA). Before each experiment, the probe was calibrated. The cantilever spring constant values provided by the manufacturer were cross-checked by performing a thermal noise calibration [16] and the calibration was performed on a solid surface. This calibration was necessary to determine the deflection sensitivity by obtaining multiple force curves on a rigid sample, which formed the basis for subsequent analysis. In addition, we determined the tip radius by reconstructing it, an operation that involved imaging a rough titanium sample (Bruker, USA). Before atomic force microscopy, the cells were fixed with 4% PFA (Sigma, USA) prepared in 1x PBS.

During Atomic Force Microscopy (AFM) scanning, measures were taken to prevent cellular damage and ensure the generation of quality force curves. These steps included limiting the tip velocity to 66 μm/s, setting the peak force tapping frequency at 0.5 kHz, defining the image scan size at 100 μm, and fixing the number of samples per line and the number of lines at 256 each. The initial force curves analysis was carried out utilizing the Derjagin, Muller, Toropov model (DMT-model [17]). This model was crucial in analyzing sample deformation by an amount smaller than the probe's radius. The nanomechanical analysis was carried out using the NanoScope analysis software (Bruker, USA), which was provided with the atomic force microscope. The subsequent data analysis for group classifications (parametric and non-parametric statistics) was carried out using GraphPad Prism 8 software (GraphPad Software, USA).

1.3. Immunocytochemistry (ICC)

Cells were fixed in 4% paraformaldehyde prepared on PBS for 15 min and washed 3 times for 5 min in 0.05% Tween-20 (Helicon, RF) prepared on PBS (PBS-T). The next step was membrane permeabilization by Triton X-100 (Helicon, RF) 0.5% solution prepared on PBS for 5 min at room temperature. Cells were blocked by incubation for 2 h with 3% Bovine Serum Albumin (Sigma–Aldrich, USA) prepared on PBS. Cell labeling was performed by using primary antibodies wnt3a (ServiceBio, China) and Rhodamine–Phalloidin staining (Thermo–Fisher, USA) in supply-recommended titer in PBS for 2 h at room temperature. Alexa Fluor 488 goat anti-rabbit IgG (H+L) (a11034, Thermo–Fisher, USA) secondary antibodies were used. The labeling procedure was performed for 1 h at room temperature. To stain the nucleus, we added DAPI (Sigma–Aldrich, USA) for 10 min with concentration of 300 nM in the final washing step.

1.4. Laser scanning microscopy

Laser Scanning Microscopy (LSM) was conducted using an Olympus FV1200 microscope (Olympus, Japan). We chose a magnification setting of x60 (Olympus UPlanSAPO 60X) and set the scanning resolution to 1600. The coverglasses with fixed and stained cells were transferred on a clean slide and embedded in Mowiol 4-88 solution. In order to yield comparative data on protein expression, we ensured that all images were procured under consistent image acquisition parameters. Additionally, these images were systematically taken in a manner such that they encompassed the fields acquired through Atomic Force Microscopy (AFM).

1.5. Image overlaying and colocalization assay

To effectively overlay the atomic force microscopy (AFM) and laser scanning microscopy (LSM) images, ImageJ software was used. To ensure effective overlay, the color scheme of each AFM image was adjusted to effectively separate invalid zero values as well as the plastic substrate. After that, the AFM image was imported into ImageJ software and manually overlaid onto the LSM image. The Coloc2 plugin was used to calculate the colocalization of AFM and LSM data. The median fluorescence was then colocalized to Young's modulus.

In this study, we employed Pearson's correlation coefficient (PCC) and Manders' colocalization coefficient (MCC) to quantitatively assess colocalization in imaging data. The PCC is calculated using the intensities of the red (Ri) and green (Gi) channels for each pixel (i), along with the mean intensities ($\overline{R}$ and $\overline{G}$) of the respective channels across the entire image. The formula for PCC is presented below:

$$PCC = \frac{\sum_i (R_i - \overline{R}) \times (G_i - \overline{G})}{\sqrt{\sum_i (R_i - \overline{R})^2 \times \sum_i (G_i - \overline{G})^2}} \tag{1}$$

and resulting values can range from +1, indicating a perfect linear relationship between the fluorescence intensities of the two channels, to –1, signifying a perfect inverse relationship Manders' colocalization coefficient (MCC) is a valuable metric for determining the proportion of one protein that colocalizes with another. When analyzing two probes, referred to as R and G, two distinct MCC values are generated: M1, which represents the fraction of R found in compartments that contain G, and M2, which indicates the fraction of G in compartments that contain R. The calculations for these coefficients are as follows:

M1 is calculated as:

$$(M1 = \frac{\sum_i R_{i,\mathrm{colocal}}}{\sum_i R_i}), \tag{2}$$

where $(R_{i,\mathrm{colocal}} = R_i)$ if $(G_i > 0)$ and $(R_{i,\mathrm{colocal}} = 0)$ if $(G_i = 0)$.

M2 is calculated as:

$$(M2 = \frac{\sum_i G_{i,\mathrm{colocal}}}{\sum_i G_i}), \tag{3}$$

where $(G_{i,\mathrm{colocal}} = G_i)$ if $(R_i > 0)$ and $(G_{i,\mathrm{colocal}} = 0)$ if $(R_i = 0)$.

Costes et al. [18] introduced a novel method for automatically determining the threshold value used to identify background levels. This method involves analyzing the range of pixel values that yield a positive Pearson correlation coefficient (PCC). Initially, PCC is calculated for all pixels in the image, and then it is recalculated for pixels corresponding to the next lower red and green intensity values along the regression line. This iterative process continues until the pixel values reach a point where the PCC falls to zero or below. The red and green intensity values at this juncture are established as the threshold values for identifying background levels in each channel. Only pixels with red and green intensity values exceeding their respective thresholds are classified as having colocalized probes. The MCC is then computed as the fraction of total fluorescence within the region of interest that is attributed to these "colocal" pixels [19].

2. Results

2.1. WNT3a intensity does not depend on cell nanomechanical properties

Glioma early-passage cells were seeded on coverslips fixed and atomic force microscopy was done. Following cell labeling and laser scanning microscopy colocalization analysis was performed to investigate the relationship between nanomechanical properties of cells and local expression of WNT3a protein and fibrillar actin in 4 cell cultures. The modified tM1, tM2 coefficients (threshold Manders' coefficients, MCC) and Pearson coefficient (PCC) were calculated, for which the ImageJ software threshold values were used.

The distribution of actin in all studied cell cultures showed significant colocalization with Young's modulus values. For the BT32 cell culture, Manders' colocalization coefficients were 0.817 (tM1) and 0.6 (tM2), indicating the extent of actin fluorescence detected over Young's modulus (tM1) and vice versa (tM2) (Fig. 1A). However, the Pearson correlation coefficient was relatively low at 0.01, reflected by a wide scatter in the scatterplot. This suggests a potential nonlinear relationship between the actin fraction and Young's modulus or a low dependency between these values (Fig. 1B).

Similar patterns were observed in other cell cultures, characterized by high Manders' colocalization coefficients (MCC) and low Pearson correlation coefficients (PCC). For BT39, the tM1, tM2, and PCC values were 0.99, 0.98, and 0.09, respectively; for BT52, the values were 0.98, 0.94, and 0.15; and for BT40, the values were 0.9, 0.68, and 0.04. Colocalization images displayed a predominance of colocalized pixels (gray) over non-colocalized pixels (red for actin and green for nanomechanics) (Fig. 1A). Scatterplots also suggested a complex relationship between Young's modulus and the actin fraction (Fig. 1B).

The distribution of Wnt3a in the studied cell cultures showed considerable heterogeneity in colocalization with Young's modulus values. In the BT39 cell culture, tM1 and tM2 were relatively high at 0.99 and 0.98, respectively, while the Pearson correlation coefficient (PCC) was –0.06. About 97% of the non-zero pixels for Wnt3a overlapped with Young's modulus values (Fig. 1C), though the scatterplot of Wnt3a and Young's modulus intensities was widely dispersed (Fig. 1D).

For BT40 and BT52 cell cultures, tM1 and tM2 values were moderate: 0.81 and 0.49 for BT40, and 0.82 and 0.56 for BT52. Similar to BT39, the PCC values for BT40 and BT52 were close to zero, at –0.05 and –0.06, respectively. The images displayed a significant number of non-colocalized Young's modulus regions (Fig. 2C), and the scatterplots for these cultures also showed a broad distribution (Fig. 1D). In contrast, the BT32 cell culture exhibited weak colocalization between Wnt3a and Young's modulus, with tM1 at 0.07 and tM2 at 0.2.

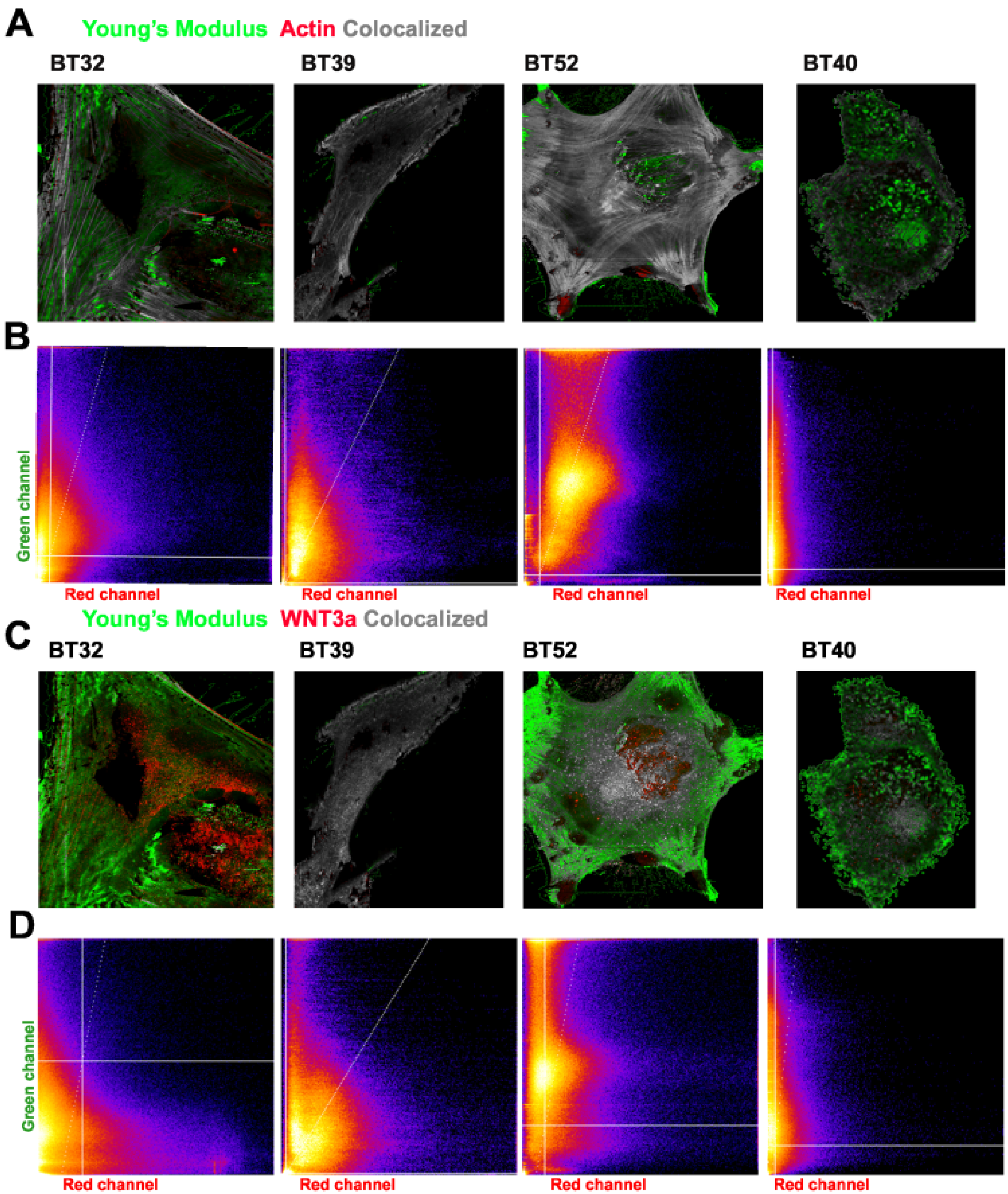


Fig. 1. Actin and WNT3a colocalization with Young's modulus: A. Merged images of actin, Young's modulus and colocalized data; B. Merged images of wnt3a, Young's modulus and colocalized data

The images revealed a large number of non-colocalized patterns for both Young's modulus and Wnt3a, with very few colocalized areas (Fig. 2C), further supported by the scatterplot indicating low colocalization (Fig. 2D).

Probes with both large and small tip radii were tested.

The sharp probe (SNL) demonstrated high indentation values in the cell, which led to "sticking", especially on rough or soft cell areas like the nucleus (Fig. 2A). These probes are suitable for precise analysis of rigid samples like plastic substrate or lyophilized protein sample (Fig. 2B). Probes with a larger tip radius are generally more suitable for AFM imaging of cells (Fig. 2C).

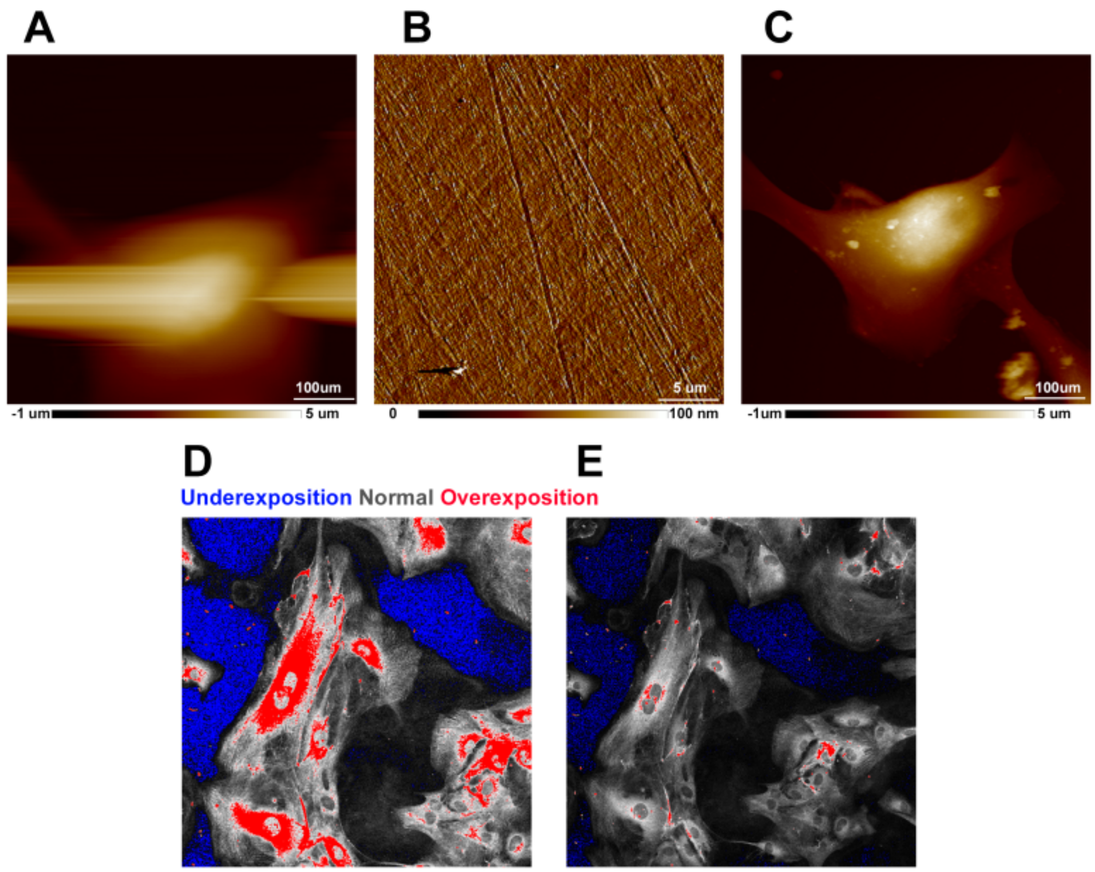


Fig. 2. The influence of the probe shape on AFM visualization: A. Cell visualized by the sharp SNL probe; B. Cell culture plastic visualized by the sharp SNL probe; C. Cell visualized by the probe with high tip radius; D. LSM image in grayscale with inappropriate microscope settings; E. LSM image in grayscale with appropriate microscope settings

Properly adjusting the LSM detector sensitivity and laser power is crucial for colocalization analysis (Fig. 2C,D). Incorrect settings can affect the PCC values due to invalid pixel intensities caused by overexposure or underexposure. Similarly, it's important to adjust the AFM pseudo-color scheme to ensure sufficient contrast for visualizing the cells (Fig. 3A). This includes setting a range for invalid/zero values and substrate data (Fig. 3B), allowing the final AFM image to clearly distinguish invalid data from the substrate (Fig. 3C).

During AFM and LSM scanning, clusters of invalid data may be acquired. Removing this data is essential for accurate colocalization analysis. Instead of manually drawing regions of interest (ROI), it's preferable to use a mask generated from the AFM image (Fig. 4A). Transferring this ROI to the LSM image ensures that any LSM data not covered by AFM is removed (Fig. 4B).

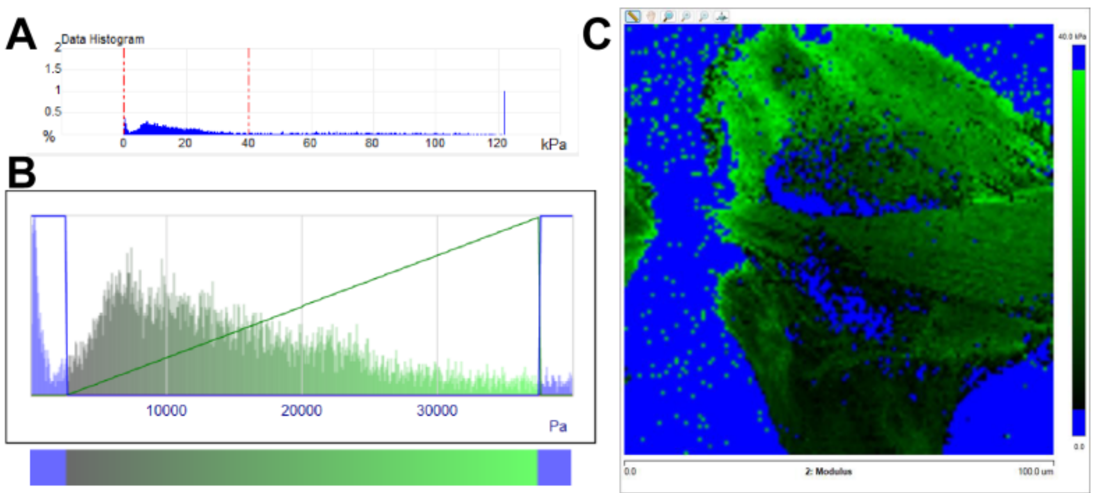


Fig. 3. AFM image color scheme adjustment: A. Data scaling to focus on the cell; B. Editing the color table to visualize the data referred to cell surface; C. The resulting AFM image

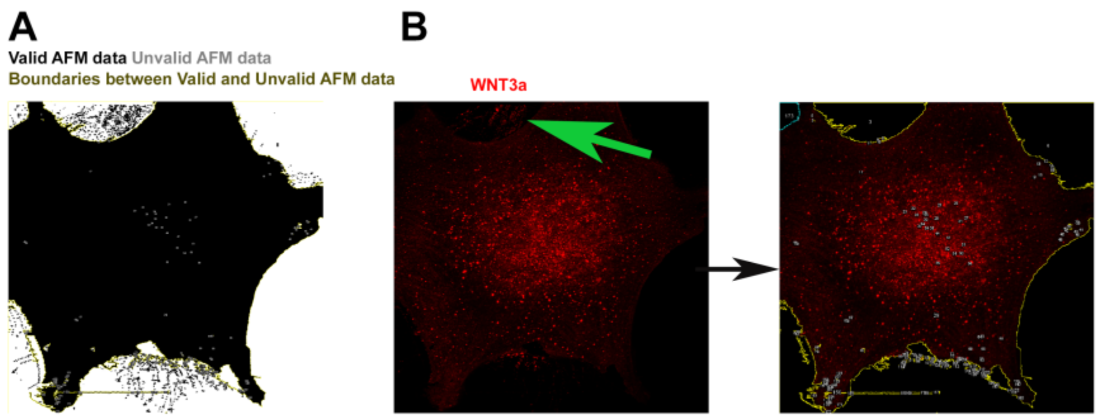


Fig. 4. LSM image editing to exclude the data not represented on AFM slide: A. The mask done on AFM image. White color refers to the surfaces without valid AFM data; B. Initial LSM image that have the same resolution, and the data not represented on AFM image (Green arrow, yellow lines) and cleaned LSM image

2.2. Procedure for AFM-LSM colocalization

1. Passage the cells on sterile coverglasses, or confocal cell culture dishes (Nunc 150680 or similar) for 1–2 days to allow the cells to recover their protein expression and synthesize surface markers, damaged by previous disaggregation.

   (Timing 1–3 d).

2. Propagate the atomic force microscopy:

   a. Remove the media and fix the cells with a 4% buffered PFA solution for 15 minutes.

   b. Carefully rinse the cells 3 times for 5 minutes with the warm PBS. Avoid scratching the cell substrate.

c. Choose an appropriate probe — the sharp and stiff probes are more suitable for small and rigid objects, and not always suitable for the cells (Fig. 2 A,B). The probes with large tip radius are better for the cell visualization, especially when the scan size exceeds 10 um (Fig. 2 C). However, the microscopy's effective resolution usually does not allow to visualize small objects (less than 200 nm), so very high resolution of an AFM image may be useless in order to AFM-LSM colocalization.

d. The critical step is probe calibration. To promote a reliable and valid scanning is required to estimate the deflection sensitivity, spring constant and tip radius/tip half angle (this depends on the model used to calculate the sample Young's modulus).

e. The AFM scanning parameters should be chosen in order to promote high resolution which must not be lower than LSM image and not damage the cell. The optimal resolution for the 100 um AFM image is 256–512 samples/line.

f. It is strongly recommended to perform a Bright field or Phase contrast image of each AFM-examined cell for the effective AFM-LSM overlaying.

(Timing 1–2 d).

3. Perform antibody labeling:

a. For an effective AFM-LSM colocalization the investigated marker must be localized on the cell surface, or very close to it, so membrane permeabilization may not be useful. Moreover, harsh detergents such as Triton X-100 or NP-40 may solubilize cell membrane, membrane-associated antigens and change its mechanical properties. So, use it carefully. The optimal detergent and the permeabilization time should be optimized for each cell line and the investigated protein. One of the most used reagent is Triton X-100 in the concentration of 0.05–0.1% in PBS and the incubation time is 3–5 minutes.

b. Carefully rinse the cells 3 times for 5 minutes with the warm PBS. Avoid scratching the cell substrate.

c. Block the cells using the bovine serum albumin, 3% on PBS for 2 hours on room temperature or overnight on +4. Use the moisture chamber to avoid drying the sample.

d. During the previous step prepare the antibody solution in desired dilution on PBS. Avoid the freeze-thaw cycles of the antibody stock.

e. Carefully rinse the cells 3 times for 5 minutes with the warm PBS. Avoid scratching the cell substrate.

f. Cover the sample with the antibody solution. Incubate 2 hours at room temperature or overnight on +4. Use the moisture chamber to avoid drying the sample.

g. During the previous step prepare the fluorophore-conjugated antibody solution in desired dilution on PBS. Avoid the freeze-thaw cycles of the antibody stock and light exposure to prevent photobleaching.

h. Carefully rinse the cells 3 times for 5 minutes with the warm PBS. Avoid scratching the cell substrate.

i. Cover the sample with the antibody solution. Incubate 2 hours at room temperature or overnight on +4 away from light.

j. Carefully rinse the cells 3 times for 5 minutes with the warm PBS. Avoid scratching the cell substrate.
k. Stain the nucleus with DAPI or PI and actin cytoskeleton with Phalloidin-fluorophore conjugate if necessary, away from light.
l. If you used coverslips as a cell substrate, transfer them on clean slides, with an addition of 15–20 ul of mounting medium. Place coverslip over the slide using tweezers. If using an aqueous mounting medium, seal with limonene or nail polish.
(Timing 1 d).

4. Examine the samples using Laser scanning microscopy. For cell visualization combined with AFM, objectives with a magnification of more than 60x are most preferred.
   a. When doing microscopy, adjust the laser power and detector sensitivity in order to avoid overexposed and underexposed pixels (Fig. 2D,E). Use the same image acquisition setting to acquire comparable intensity data.
   b. If possible use the same picture resolution (pixels/um) as the AFM resolution.
   (Timing 1d).
5. Set up the AFM Young's Modulus Image Color Scheme.
   a. Adjust the image data scale to show the cell surface in detail. The data scale should range from zero to the hardest part of the cell surface (Figure 3A).
   b. Set up the color scheme. The color plot should include 3 regions: zero and invalid data points; the bulk of the cell surface; and the substrate region. The bulk of the color plot should be in a single color channel — red, green, or blue — and should not overlap with zero or substrate data points, which should be in a different single color (Fig. 3B,C).
   c. Export the resulting image as a high-resolution, uncompressed .tif file.
   (Timing 20–30 min).
6. Overlay the AFM and LSM image using ImageJ or another raster graphics editor.
   a. Modify LSM image in order to effective overlay the AFM image — flip picture if necessary, rotate and resize it.
   b. Crop the LSM to AFM picture size and match the cell structures on both images.
   c. Import the AFM image to ImageJ, separate the channels by Image → Color → Make Composite command, then split the channels by Image → Color → Split Channels command. Save the obtained images in .tif.
   d. Import the prepared LSM image to ImageJ and do the same manipulations as described in the previous step. Save the obtained images in .tif.
   (Timing 2–4 h).
7. Perform the image colocalization analysis using ImageJ.
   a. Create a negative image mask based on the AFM image obtained on 8c step by using Image → Adjust → Threshold command. Use the values between 0 and 1 (Fig. 4A). This step is needed to remove all the data on the LSM image which is not covered by AFM.

b. Create the composite Region of Interest by Analyze $\rightarrow$ Analyze particles command. The values of the fields "Size" and "Circularity" must be "0-Infinity" and "0.00-1.00", respectively, parameters "Add to manager", "Include holes", "Overlay", "Composite ROIs" must be tagged on. Other parameters are out of interest in this analysis. As a result a bulk of regions of interest will be obtained. (Fig. 4B).

c. Transfer the ROIs obtained on the previous step to each stack of LSM images obtained on the 8d step and delete the data in these ROIs.

d. Use Coloc2 or ColocalisationThreshold plugin to perform pixel-wise colocalization analysis.

(Timing 1 h).

3. Discussion

It was shown that β-actin and other cytoskeleton components have a significant impact on cell rheology. However, not only cytoskeleton itself modulates cell mechanics, it always bounded to various proteins , such as myosin and tropomyosin, which could modulate cell mechanics via contractile forces [20], or could change actin conformation [21]. The weak Pearson correlation between cell surface nanomechanics and actin cytoskeleton localisation may be explained by localisation of actin fibers deep inside the cell or altered contractility of actin-myosin compounds and therefore, non-linear dependency between Young's modulus and actin expression [22]. However, high MCC values indicate that actin cytoskeleton provides a significant impact on cell mechanics.

Wnt3a is a target gene of regulatory RNA such as miR-491 and miR-491 that mediates epithelial-mesenchymal transition (EMT). Additionally, miR-491 regulated the proliferation through the Wnt/β-Catenin pathway by targeting Wnt3a [11]. Wnt3a downregulation leads to Wnt-signaling alteration and increased sensitivity to temozolomide in vitro [23]. It was shown that Wnt3a receptor — LRP6 protein is associated with lipid rafts [24] and signal transduction modulated by Wnt3a is hypothesized to modulate local membrane hardness and could be an effective nanomechanical marker for EMT. However, the results, which show varying levels of colocalization across different cell cultures, cast doubt on this thesis.

The method used for this investigation could be used for the various implementations connected to the AFM and LSM studied and used for the direct proof of physical relations between studying protein and membrane stiff structures such as lipid rafts or cytoskeleton local complexes.

AFM is rarely used in combination with LSM. Several papers including high-resolution AFM-LSM scanning [25], actin cytoskeleton scanning [26], cell receptor visualization [27] and SEM-LSM scanning [28]. However, these studies did not perform a quantitative assessment of data from one area obtained by different methods. Moreover, hardware optical laser scanning AFM systems are not widely used and are expensive. Our method will make it possible to effectively use data obtained in various ways for multivariate analysis of biological systems.

Conclusion

Comparing data from different microscopy techniques can be challenging. Our ImageJ-based method offers a fast, reliable, and free way to analyze cell surfaces captured by various imaging methods, both optical and non-optical. This approach enhances the versatility of the data and enables a quantitative evaluation of specific features of interest.

This research was funded by the Russian Science Foundation (project no. 20-15-00378n).

References

[1] M.Radmacher, Studying the Mechanics of Cellular Processes by Atomic Force Microscopy, in Methods in Cell Biology, Academic Press (Cell Mechanics), 2007, 347–372. DOI: 10.1016/S0091-679X(07)83015-9

[2] G.Zhou, et al., Cells nanomechanics by atomic force microscopy: focus on interactions at nanoscale, *Advances in Physics: X*, **6**(2021), no. 1, 1866668. DOI: 10.1080/23746149.2020.1866668

[3] P.Roca-Cusachs, V.Conte, X.Trepat, Quantifying forces in cell biology, *Nature Cell Biology*, **19**(2017), no. 7, 742–751. DOI: 10.1038/ncb3564

[4] M.Skamrahl, et al., Simultaneous Quantification of the Interplay Between Molecular Turnover and Cell Mechanics by AFM-FRAP, *Small (Weinheim an Der Bergstrasse, Germany)*, **15**(2019), no. 40, e1902202. DOI: 10.1002/smll.201902202

[5] M.Cai, et al., Cell membrane sample preparation method of combined AFM and dSTORM analysis, *Biophysics Reports*, **8**(2022), no. 4, 183–192. DOI: 10.52601/bpr.2022.220004

[6] V.M.Farniev, et al., Nanomechanical and Morphological AFM Mapping of Normal Tissues and Tumors on Live Brain Slices Using Specially Designed Embedding Matrix and Laser-Shaped Cantilevers, *Biomedicines*, **10**(2022), no. 7, 1742. DOI: 10.3390/biomedicines10071742

[7] M.Krieg, et al., Atomic force microscopy-based mechanobiology, *Nature Reviews Physics*, **1**(2019), no. 1, 41–57. DOI: 10.1038/s42254-018-0001-7

[8] M.E.Shmelev, et al., Nanomechanical Signatures in Glioma Cells Depend on CD44 Distribution in IDH1 Wild-Type but Not in IDH1R132H Mutant Early-Passage Cultures, *International Journal of Molecular Sciences*, **24**(2023), no. 4, 4056. DOI: 10.3390/ijms24044056

[9] M.Pashirzad, et al., Role of Wnt3a in the pathogenesis of cancer, current status and prospective, *Molecular Biology Reports*, **46**(2019), no. 5, 5609–5616. DOI: 10.1007/s11033-019-04895-4

[10] F.Lu, et al., miR-497/Wnt3a/c-jun feedback loop regulates growth and epithelial-to-mesenchymal transition phenotype in glioma cells, *International Journal of Biological Macromolecules*, **120**(2018), no. Pt A, 985–991. DOI: 10.1016/j.ijbiomac.2018.08.176

[11] Y.Meng, F.-R.Shang, Y.-L.Zhu, MiR-491 functions as a tumor suppressor through Wnt3a/β–catenin signaling in the development of glioma, *European Review for Medical and Pharmacological Sciences*, **23**(2019), no. 24, 10899–10907. DOI: 10.26355/eurrev_201912_19793

[12] S.Patra, et al., Dysregulation of histone deacetylases in carcinogenesis and tumor progression: a possible link to apoptosis and autophagy, *Cellular and molecular life sciences: CMLS*, **46**(2019), no. 5, 3263–3282. DOI: 10.1007/s00018-019-03098-1

[13] J.Shan, et al., Identification of a specific inhibitor of the dishevelled PDZ domain, *Biochemistry*, **44**(2005), no. 47, 15495–15503. DOI: 10.1021/bi0512602

[14] Y.Wang, et al., hsa-miR-216a-3p regulates cell proliferation in oral cancer via the Wnt3a/β-catenin pathway, *Molecular Medicine Reports*, **27**(2023), no. 6, 128. DOI: 10.3892/mmr.2023.13015

[15] D.Matias, et al., GBM-Derived Wnt3a Induces M2-Like Phenotype in Microglial Cells Through Wnt/β-Catenin Signaling, *Molecular Neurobiology*, **56**(2019), no. 2, 1517–1530. DOI: 10.1007/s12035-018-1150-5

[16] N.Mullin, J.K.Hobbs, A non-contact, thermal noise based method for the calibration of lateral deflection sensitivity in atomic force microscopy, *Review of Scientific Instruments*, **85**(2014), no. 11, 113703. DOI: 10.1063/1.4901221

[17] B.V.Derjaguin, V.M.Muller, Yu.P.Toporov, Effect of contact deformations on the adhesion of particles, *Journal of Colloid and Interface Science*, **53**(1975), no. 2, 314–326. DOI: 10.1016/0021-9797(75)90018-1

[18] S.V.Costes, et al., Automatic and quantitative measurement of protein-protein colocalization in live cells, *Biophysical journal*, **86**(2004), no. 6, 3993–4003. DOI: 10.1529/biophysj.103.038422

[19] K.W.Dunn, M.M.Kamocka, J.H.McDonald, A practical guide to evaluating colocalization in biological microscopy, *American Journal of Physiology-Cell Physiology*, **300**(2011), no. 4, 723–742. DOI: 10.1152/ajpcell.00462.2010

[20] I.Jalilian, et al., Cell elasticity is regulated by the tropomyosin isoform composition of the actin cytoskeleton, *PloS One*, **10**(2015), no. 5, e0126214. DOI: 10.1371/journal.pone.0126214

[21] L.Simone, et al., AQP4 Aggregation State Is a Determinant for Glioma Cell Fate, *Cancer Research*, **79**(2019), no. 9, 2182–2194. DOI: 10.1158/0008-5472.CAN-18-2015

[22] J.H.McDonald, K.W.Dunn, Statistical tests for measures of colocalization in biological microscopy, *Journal of Microscopy*, **252**(2013), no. 3, 295–302. DOI: 10.1111/jmi.12093

[23] N.Kaur, et al., Wnt3a mediated activation of Wnt/β–catenin signaling promotes tumor progression in glioblastoma, *Molecular and Cellular Neurosciences*, **54**(2013), 44–57.

[24] G.Riitano, et al., LRP6 mediated signal transduction pathway triggered by tissue plasminogen activator acts through lipid rafts in neuroblastoma cells, *Journal of Cell Communication and Signaling*, **14**(2020), no. 3, 315–323. DOI: 10.1016/j.mcn.2013.01.001

[25] Y.Chen, et al., Spectral analysis of irregular roughness artifacts measured by atomic force microscopy and laser scanning microscopy, *Microscopy and Microanalysis: The Official Journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, **20**(2014), no. 6, 1682–1691. DOI: 10.1017/S1431927614013385

[26] K.Meller, C.Theiss, Atomic force microscopy and confocal laser scanning microscopy on the cytoskeleton of permeabilised and embedded cells, *Ultramicroscopy*, **106**(2006), no. 4-5 320–325. DOI: 10.1016/j.ultramic.2005.10.003

[27] A.V.Moskalenko, et al., Single protein molecule mapping with magnetic atomic force microscopy, *Biophysical Journal*, **98**(2010) no. 3, 478–487. DOI: 10.1016/j.bpj.2009.10.021

[28] K.Szafranska, et al., From fixed-dried to wet-fixed to live — comparative super-resolution microscopy of liver sinusoidal endothelial cell fenestrations, *Nanophotonics*, **11**(2022), no. 10, 2253–2270. DOI: 10.1515/nanoph-2021-0818

Новый протокол для анализа клеточной наномеханики в сочетании с быстрым белковым профилированием с помощью колокализации паттернов АСМ-ЛСМ

Михаил Е. Шмелев
Факультет медицины и естественных наук Дальневосточного федерального университета
Национальный научный центр морской биологии имени А.В. Жирмунского ДВО РАН
Владивосток, Российская Федерация
Анна К. Кравченко
Факультет медицины и естественных наук Дальневосточного федерального университета
Владивосток, Российская Федерация
Вадим В. Кумейко
Факультет медицины и естественных наук Дальневосточного федерального университета
Национальный научный центр морской биологии имени А.В. Жирмунского ДВО РАН
Владивосток, Российская Федерация

Аннотация. Атомно-силовая микроскопия стала одним из ключевых методов изучения клеток и белков. В этой работе представлен новый протокол, который сочетает в себе наномеханическое исследование клеток с быстрым профилированием экспрессии белков, что позволяет получить новый источник данных для фундаментальных и прикладных исследований в области клеточной биологии. Новый протокол основан на методах атомно-силовой микроскопии (АСМ) для измерения механических свойств отдельных клеток и лазерной сканирующей микроскопии (ЛСМ) для высокопроизводительного количественного анализа уровня экспрессии белков. Такой подход позволяет оценить взаимосвязь между механическими свойствами клеток и динамикой белков, раскрывая важные аспекты физиологии и патофизиологии клеток. Эффективность протокола была подтверждена экспериментами с раковыми клетками, демонстрируя его потенциал в колокализации молекулы лиганда wnt3a и актина цитоскелета с картинами модуля Юнга клетки. Разработанный подход найдет свое применение в разработке лекарств, диагностике злокачественных опухолей и персонализированной медицине, предлагая новый взгляд на сложное взаимодействие между механикой клеток и локальной экспрессией белков.

Ключевые слова: атомно-силовая микроскопия, лазерная сканирующая микроскопия, колокализация, глиомы, wnt-сигналинг.